INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Acoustic analysis of detailed three-dimensional shape of the human nasal cavity and paranasal sinuses

Tatsuya Kitamura[1], *Hironori Takemoto*[2], *Hisanori Makinae*[3], *Tetsutaro Yamaguchi*[4], *Kotaro Maki*[4]

[1]Konan University, Japan
[2]Chiba Institute of Technology, Japan
[3]National Research Institute of Police Science, Japan
[4]Showa University, Japan

t-kitamu@konan-u.ac.jp

Abstract

The nasal and paranasal cavities have a labyrinthine shape and their acoustic properties affect speech sounds. In this study, we explored the transfer function of the nasal and paranasal cavities, as well as the contribution of each paranasal cavity, using acoustical and numerical methods. A physical model of the nasal and paranasal cavities was formed using data from a high-resolution 3D X-ray CT and a 3D printer. The data was acquired from a female subject during silent nasal breathing. The transfer function of the physical model was then measured by introducing a white noise signal at the glottis and measuring its acoustic response at a point 20 mm away from the nostrils. We also calculated the transfer function of the 3D model using a finite-difference time-domain or FDTD method. The results showed that the gross shape and the frequency of peaks and dips of the measured and calculated transfer functions were similar, suggesting that both methods used in this study were reliable. The results of FDTD simulations evaluating the paranasal sinuses individually suggested that they contribute not only to spectral dips but also to peaks, which is contrary to the traditional theories regarding the production of speech sounds.

Index Terms: X-ray CT data, 3D printer, Physical model, Acoustic measurement, Finite-difference time-domain method

1. Introduction

The nasal cavity is divided into a right and left tract by the nasal septum. Its anterior end, also known as the nostril, opens to the face, and its posterior end connects to the oropharynx through the choana and the nasopharynx. The paranasal sinuses include the bilateral maxillary sinuses, frontal sinuses, sphenoidal sinuses, and ethmoidal sinuses; all are connected to, and interact acoustically with, the nasal cavity. The shape of the nasal cavity and paranasal sinuses is complex and asymmetric, and differs from person to person. Because the shape changes very little while speaking, its acoustic properties could provide speaker-specific information. However, little attention has been paid to the acoustic contributions of the nasal and paranasal cavities to speech sounds. The present study explores the acoustics of a detailed three-dimensional (3D) shape of these cavities using measurements obtained from the physical model and finite-difference time-domain (FDTD) acoustic simulations.

The morphology of the nasal cavity and paranasal sinuses and their acoustic properties have been investigated via multiple approaches, including advanced imaging techniques, acoustic measurement techniques, and numerical methods. Dang *et al.* [1] conducted morphological and acoustical analyses of the nasal and paranasal cavities using a magnetic resonance imaging (MRI) technique. The study revealed individual differences in size and shape and the acoustic effects of these differences using a one-dimensional (1D) acoustic propagation model. The investigators reported that the right and left nasal tracts interacted acoustically with each other, and the paranasal sinuses acted as side branches that generated spectral dips in the lower frequencies. However, because the nasal and paranasal cavities are thin-walled labyrinthine structures, the resolution of the MRI data was insufficient, and the 1D model was found to be inadequate in calculating the acoustic effects. Matsuzaki *et al.* [2] compared the transfer functions calculated using 1D and 3D models and showed that the differences were significant when the models included the paranasal sinuses.

The unique acoustics created by individual variations in the shape of the nasal and paranasal cavities contribute to speaker characteristics in speech sounds. Amino and Arai [3] evaluated which consonants contributed most to speaker identification and concluded that nasal consonants were the most identifiable, regardless of whether the listeners were familiar with the speakers. Amino *et al.* [4] examined the relationship between perceptual speaker similarity and nasality in speech sounds. The investigators revealed that listeners could identify speaker characteristics in the nasal output of speech sounds even when the speech did not contain nasal consonants. Dang *et al.* [5] demonstrated that sound can radiate through the velum during the production of non-nasal sounds indicating that the acoustic properties of the nasal and paranasal cavities affect overall speech sounds. It is therefore indispensable to explore the acoustic properties derived from the complex shapes of the cavities in order to refine the acoustic theory of speech production.

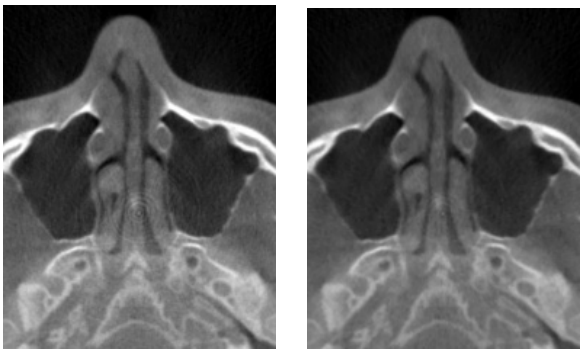
In this paper, we aim to (1) create a 3D physical model of the nasal and paranasal cavities from high-resolution X-ray CT images and measure its transfer function acoustically and (2) calculate the transfer function of the cavities using an FDTD method. The transfer function obtained by acoustic measurement provides a benchmark for evaluating numerical analysis methods. The FDTD simulation evaluates the acoustical status of the cavities and the individual acoustical contribution of each cavity.

2. Materials and methods

2.1. Three-dimensional X-ray CT data

This study was conducted using volumetric X-ray CT data of the head and neck regions of an 18-year-old Japanese female patient with no abnormalities of the nasal and paranasal cavities. The data were obtained using a CB MercuRay dentomaxillofacial cone-beam X-ray CT system (Hitachi Ltd., Tokyo, Japan)

http://dx.doi.org/10.21437/Interspeech.2017-107



(a) (b)

Figure 1: *The original (a) and median filtered (b) X-ray CT images.*

Figure 2: *Physical nasal and paranasal tract model of a female subject during silent nasal breathing.*

at the dental hospital of Showa University for the purpose of clinical examination. The data were acquired in the transverse plane during quiet nasal breathing using an X-ray tube voltage of 100 kVp, 0.377 mm $\times$ 0.377 mm pixel size, 512 $\times$ 512 pixel image size, 0.377 mm slice thickness, and 512 slices. Use of the X-ray CT data was approved by the institutional review board of Showa University.

2.2. Physical nasal and paranasal tract model

The 3D shape of the facial surface and the supraglottic airway including the nasal cavity, paranasal sinuses, and pharyngeal and laryngeal cavities was reconstructed from the X-ray CT data. The 3D shape was then replicated precisely, creating a physical model using a 3D printer. Because the central area of the X-ray CT image was contaminated by concentric noises, and the brightness varied both within an image and between images, air and tissue could not be accurately separated by simple thresholding. Therefore, the concentric noises were restrained by a fifth order median filter for radial directions on the 200 pixel $\times$ 200 pixel center area of the images. The original and median filtered images are shown in Fig. 1 (a) and (b), respectively.

The two regions were then segmented using a supervised machine learning method. The Trainable Weka Segmentation of Fiji (an extended version of ImageJ, NIMH, Bethesda, MD, USA) was used to binarize the image data and extract the air regions both inside and outside the body. The plugin provides machine learning based segmentation algorithms by calling functions of Weka (Waikato Environment for Knowledge Analysis) on Fiji. For segmentation, regions that belonged to either the air or body tissue classes, were selected on several images by GUI operations; those regions were then used as training data sets for a machine learning task that computed a classifier that segmented the two classes. A random forest classifier was used in this study. The performance of the classifier could be improved by adding labeled training data and the trained classifier could be utilized for segmentation of other images. By introducing the machine learning method, we improved the reproducibility of the segmentation and reduced the amount of work that would otherwise have been required to trace all the air regions manually.

The air regions were determined not only in the transverse plane but also in the coronal plane by the above procedure. The vocal tract regions were then determined as a logical disjunction of these air regions. If the vocal tract regions were extracted from the images of only one dimension, narrow pathways of the nasal and paranasal sinuses were not traced correctly and interconnections between the sinuses were not preserved. Therefore, the air regions were extracted in the two planes and the results were merged into one vocal tract shape.

The air regions inside and outside the body were reconstructed as 3D volume data; these data were converted into an STL file format, which represented 3D CAD model triangulation. Next, the surface of the 3D model was smoothed. These processes were performed using a 3D image editor, Mimics (Materialise, Leuven, Belgium). A 2-mm wall was then extruded outside the 3D model by an STL editor, 3-matic (Materialise, Leuven, Belgium). The wall under the glottis was removed to allow input of sound. Lastly, a physical nasal cavity and paranasal sinus model (Fig. 2), referred to here as the physical model, was formed with 0.254 mm layer thickness using a uPrint SE 3D printer (Stratasys Ltd., Eden Prairie, MN, USA). The printer built the physical model from the thermoplastic polymer, acrylonitrile butadiene styrene. The support structures were composed of an alkali-soluble material and were removed with an alkaline solution. The physical model was created in six pieces to allow confirmation that all the support structures had been dissolved; the pieces were then glued together to make the whole physical model.

2.3. Acoustic measurement method

Figure 3 shows a diagram of the setup for measurement of the acoustic characteristics of the physical model. The measurement was performed in a soundproof room while keeping the temperature at 20 °C. A 10 mm thick aluminum cap with a 1.2 mm diameter vertical hole and a 30 cm $\times$ 30 cm plane baffle made of a SOH-2 sound-absorbing material (SONEX, Inc., Seattle, WA, USA) was attached to the outlet of a P-300 mid-range horn driver unit (UNI-PEX Co., Ltd., Osaka, Japan). Then, the physical model was mounted on the cap with the center of the glottis located immediately above the vertical hole in the cap. The dimensions of the vertical hole were determined so as to allow acoustic decoupling of the physical model from the horn driver unit and to simulate a closed-glottis status during speaking [6][7].

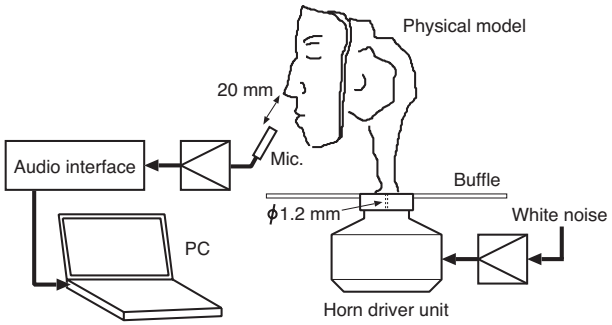


Figure 3: *Diagram of measurement setup.*

A white-noise signal was introduced into the physical model at the glottis from the horn driver unit via an RSDA302U amplifier (Rasteme systems Co., Ltd., Niigata, Japan). An acoustic response from the physical model was recorded at a point 20 mm away from the nostrils, for a period of 5 seconds, at a sampling rate of 20 kHz. Sixteen-bit resolution was achieved using a 2669 microphone (Bruel and Kjaer, Naerum, Denmark), 2690 amplifier (Bruel and Kjaer, Naerum, Denmark), UA-5 USB audio interface (Roland Co., Ltd., Shizuoka, Japan), and a PC. The output signal of the microphone was high-pass filtered at 20 Hz by the amplifier to remove the DC offset of the acoustic signal. The transfer function of the physical model was then obtained using a short-time Fourier transform. The log spectra of the recorded sound were calculated using a 2048-point Hanning window and a 2048-point frame period, and then averaged with respect to the frames.

Prior to measurement, the transfer function of the whole setup was compensated by carrying out inverse filtering. The acoustic response of the measurement system without the physical model present was first measured for 5 seconds using the method described above. The coefficients of a p-th order linear predictor, a_n, were next calculated for the acoustic response. The order p was set to 50. The transfer function of the inverse filter was obtained using the formula

$$H^{-1}(z) \quad = \quad \sum_{n=0}^{p} a_n z^{-n}. \qquad (1)$$

The white-noise signal was then filtered with the inverse filter. Inverse filtering was repeated by inputting the resultant signal, which resulted in a relatively flat transfer function.

2.4. Numerical analysis

The transfer function of the 3D shape of the supraglottic airway beginning at the glottis was calculated using an FDTD method. Details of the method were described by Takemoto *et al.* [8]. The STL data used for the 3D modeling were converted back to voxel data to create a computer model for FDTD simulations. The computer model involved radiation space anterior to the facial surface. The simulation constants were as follows: air density $\rho = 1.17$ kg/m^3, speed of sound c = 346.7 m/s, time step of calculation $\Delta t = 5.0 \times 10^{-7}$, grid width = 1.0 mm, and sound absorption coefficient on the surface of the body tissue = 0.008. A point source of the volume velocity was located immediately above the glottis with the following Gaussian pulse

$$g(t) \quad = \quad e^{-\left\{\frac{\Delta t n - T}{0.29T}\right\}^2} \quad (\mathrm{m}^3/\mathrm{s}) \qquad (2)$$

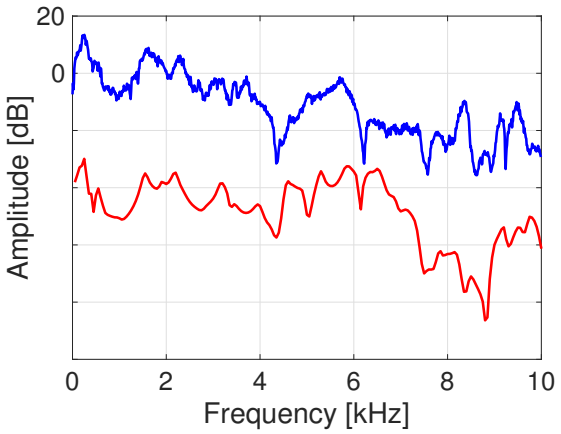


Figure 4: *Transfer function of the nasal and paranasal physical model (blue) and transfer function obtained by an FDTD simulation (red).*

where $T = 0.646/f_0$ and f_0 = 10 kHz. Sound propagation was calculated for a 20 ms period and the observation point was placed 20 mm away from the anterior nares.

The transfer functions without the paranasal sinuses were also calculated to examine the acoustic contribution of each sinus. The paranasal sinuses were eliminated in order of descending volume – the left maxillary sinus (LMS), right maxillary sinus (RMS), sphenoidal sinus (SS), and frontal sinus (FS) – then the transfer functions were calculated.

3. Results and discussion

Figure 4 depicts the transfer functions obtained from the acoustic measurement and the acoustic simulation. The general shape and the frequency of each spectral peak and dip of the transfer functions were found to be similar, suggesting that both of the methods used in the present study were reliable. However, the similarity of the transfer functions was less than that observed by Takemoto *et al.* [8]. The investigators in that study demonstrated that the transfer functions of the vocal tract (without the nasal cavity and paranasal sinuses) calculated by the FDTD method corresponded extremely well with the acoustically-measured transfer functions of vocal tract physical models.

There are a couple of possible reasons for the greater discrepancy between the transfer functions obtained by the two methods in this study. The first is related to wall vibration. The paranasal sinuses are joined complexly, sharing thin walls, and thus could interact acoustically with each other via the walls. However, the FDTD method used in both the current study and the one by Takemoto *et al.* [8] ignored sound transmission through the walls. It is possible that the effects of sound transmission caused the difference between the transfer functions obtained from the acoustic measurement and the acoustic simulation. The second reason is the difference in the accuracy of the physical models. The physical models were formed by industrial stereolithographs in Takemoto *et al.* [8], while the physical model was created by a consumer 3D printer in the present study.

Figure 5 illustrates the transfer functions of the vocal tract: without LMS, without LMS and RMS, without LMS, RMS, and SS, and without LMS, RMS, SS, and FS. The transfer functions

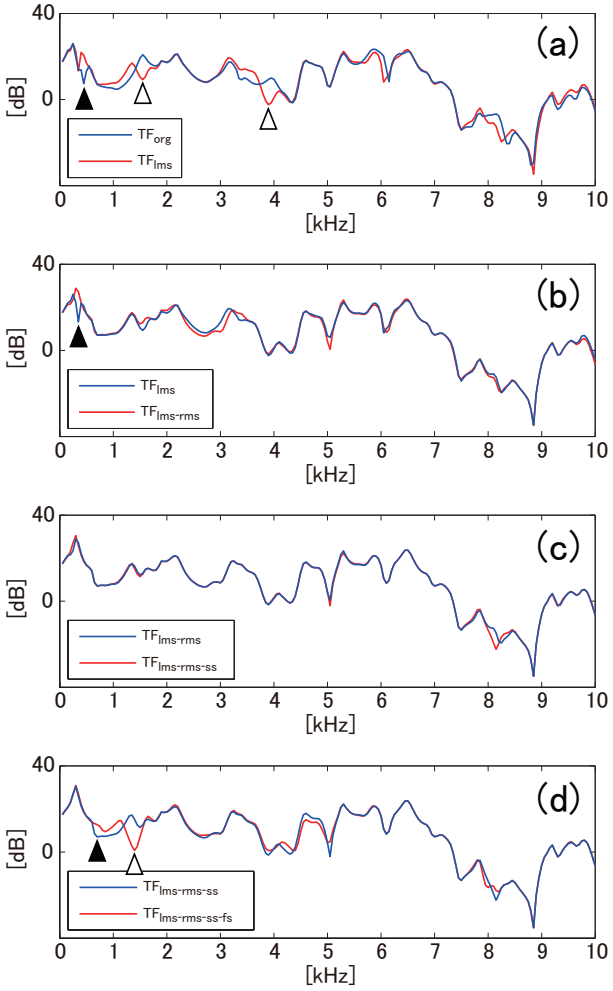


Figure 5: *Effects of elimination of the paranasal sinuses on the transfer function of the vocal tract. The transfer functions* $\mathrm{TF_{org}}$ *and* $\mathrm{TF_{lms}}$ *(a),* $\mathrm{TF_{lms}}$ *and* $\mathrm{TF_{lms-rms}}$ *(b),* $\mathrm{TF_{lms-rms}}$ *and* $\mathrm{TF_{lms-rms-ss}}$ *(c), and* $\mathrm{TF_{lms-rms-ss}}$ *and* $\mathrm{TF_{lms-rms-ss-fs}}$ *(d). Open and filled triangles indicate dips that appeared and disappeared by eliminating the sinuses, respectively.*

are referred to as $\mathrm{TF_{org}}$, $\mathrm{TF_{lms}}$, $\mathrm{TF_{lms-rms}}$, $\mathrm{TF_{lms-rms-ss}}$, and $\mathrm{TF_{lms-rms-ss-fs}}$. Filled triangles in the figure indicate spectral dips that disappeared when the paranasal sinuses were eliminated; the LMS, RMS, and SS contributed to the spectral dips at 0.45 kHz (Fig. 5a), 0.35 kHz (Fig. 5b), and 0.7 kHz (Fig. 5d), respectively, showing agreement with previous studies of the acoustic effects of the paranasal sinuses [1][9]. The three spectral peaks may have been caused by the first resonance of each paranasal sinus. Theoretically, other dips could be generated by higher resonance of the sinuses, but no other dip disappeared in the frequency range up to 10 kHz as a result of elimination of the paranasal sinuses. As shown in Fig. 5c, the SS did not contribute to any spectral dips; this is likely because one of the two pathways from the SS to the nasal cavity was very narrow and the other was disconnected.

The elimination of the paranasal sinuses from the nasal tract did not only remove the spectral dips as mentioned above but also removed peaks and generated peak-dip (pole-zero) pairs. These peak-dip pairs are indicated by open triangles in Fig. 5. By eliminating the LMS, the spectral peaks at 1.55 kHz and 3.95 kHz disappeared and peak-dip pairs appeared at 1.50 kHz and 3.80 kHz as shown in Fig. 5a. In addition, by eliminating the FS, the spectral peak at 1.35 kHz disappeared and the peak-dip pair appeared at 1.25 kHz as shown in Fig 5d. These results suggest that the paranasal sinuses contribute to the generation of peaks. In previous studies and according to the conventional theory of speech sound production [1][10][11], the acoustical function of the paranasal sinuses has been considered to be as side branches. The results of this study, that the paranasal sinuses contributed to generation of the peaks, suggest that the paranasal sinuses could act not only as side branches but also as resonators.

4. Conclusions

In this paper, we measured the acoustic properties of a physical nasal and paranasal model reconstructed from data from a high-resolution volumetric X-ray CT of a female subject. We also calculated the transfer function of the cavities using the FDTD method. We showed that the acoustic characteristics determined by the two methods nearly matched, demonstrating that the numerical simulation method provides an accurate analysis of the complex nasal and paranasal cavity shape. Combination of 3D printing and numerical simulation technologies should continue to provide reliable information regarding acoustics and speech science.

Although additional subjects are needed to confirm our results, numerical analyses suggest that the paranasal sinuses contribute not only to anti-resonances but also to resonances on the transfer function. This means that the paranasal sinuses may not only act as side branches of the nasal tract, which is contrary to common belief. Research into the acoustic properties of detailed nasal and paranasal cavities may give us a new perspective on the acoustic theory of speech sound production.

5. Acknowledgements

This study was supported by JSPS KAKENHI (Nos. 15K00263, 25280026, and 25280066). The authors thank Prof. Kimihiro Yamanaka from Konan University for his help with the use of the 3D printer.

6. References

[1] J. Dang, K. Honda, and H. Suzuki, "Morphological and acoustical analysis of the nasal and the paranasal cavities," *J. Acoust. Soc. Am.*, vol. 96, no. 4, pp. 2088–2100, 1994.

[2] H. Matsuzaki, A. Serrurier, P. Badin, and K. Motoki, "One-dimensional and three-dimensional propagation analysis of acoustic characteristics of Japanese and French vowel /a/ with nasal coupling," *Acoust. Sci. & Tech.*, vol. 35, no. 1, pp. 35–41, 2014.

[3] K. Amino and T. Arai, "Effects of linguistic contents on perceptual speaker identification: Comparison of familiar and unknown speaker identifications," *Acoust. Sci. & Tech.*, vol. 30, no. 2, pp. 89–99, 2007.

[4] K. Amino, H. Makinae, and T. Kitamura, "Nasality in speech and its contribution to speaker individuality," *Proc. Interspeech2014*, pp. 1688–1692, 2014.

[5] J. Dnag, W. Jianguo, K. Honda, and T. Nakai, "A study on transvelar coupling for non-nasalized sounds," *J. Acoust. Soc. Am.*, vol. 139, no. 1, pp. 441–454, 2016.

[6] T. Kitamura, H. Takemoto, S. Adachi, and K. Honda, "Transfer functions of solid vocal-tract models constructed from ATR MRI

database of Japanese vowel production," *Acoust. Sci. & Tech.*, vol. 30, no. 4, pp. 288–296, 2009.

[7] K. Honda, T. Kitamura, H. Takemoto, S. Adachi, P. Mokhtari, S. Takano, Y. Nota, H. Hirata, I. Fujimoto, Y. Shimada, S. Masaki, S. Fujita, and J. Dang, "Visualization of hypopharyngeal cavities and vocal-tract acoustic modeling," *Computer Methods in Biomechanics & Biomedical Engineering*, vol. 13, no. 4, pp. 443–453, 2009.

[8] H. Takemoto, P. Mokhtari, and T. Kitamura, "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method," *J. Acoust. Soc. Am.*, vol. 128, no. 6, pp. 3724–3738, 2010.

[9] J. Dang and K. Honda, "Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation," *J. Acoust. Soc. Am.*, vol. 100, no. 5, pp. 3374–3383, 1996.

[10] K. N. Stevens, *Acoustic Phonetics*. Cambridge, London: The MIT Press, 2000.

[11] T. Pruthi, C. Y. Espy-Wilson, and B. Story, "Simulation and analysis of nasalized vowels based on magnetic resonance imaging data," *J. Acoust. Soc. Am.*, vol. 121, no. 6, pp. 3858–3873, 2007.