Education in acoustics of nasalized vowels using physical models of the human vocal tract with nasal cavity

○Takayuki Arai (Sophia Univ.) and Matias Budhiantho (Satya Wacana Christian Univ.)

1 Introduction

Teaching the acoustics of speech is important not only for students engaged in speech technology but also those in speech science and phonetics. For intuitive education in the acoustics of speech production, we have developed and use different types of physical models of the human vocal tract [1-3]. Some of these models have a nasal cavity attached to the main vocal tract so that the nasalization of vowels can be demonstrated.

Previous studies (e.g., [4]) have presented spectrographic representations of nasalized vowels produced by a physical model, demonstrating spectral change in time when vowels /a/ and /i/ are gradually nasalized. In the case of /a/, the first formant (F1) becomes lower when the vowel is nasalized, whereas the second formant (F2) stays more or less the same. This formant change is mainly focused in details in the present study.

2 Measurement

In this section, we present our research objectives, the experimental setup, and the results of our measurements.

2.1 Objectives

Our laboratory has developed a set of vocal-tract models to simulate the effect of coupling with the nasal cavity. This set consists of one model of the nasal cavity only and one head-shaped model for the vowel /a/ with a moveable velum that can be used to simulate the effect of nasal coupling. In the current experiment, we measured a) the response of only the nasal cavity and b) the response of the head-shaped /a/ model with various degrees of velopharyngeal opening.

2.2 Experimental setup

We measured the models using a multiple swept-sine signal from a WinMLS 2004 PC measurement system. The signal was fed to the driver unit (TOA, TU-750) of a horn speaker through a power amplifier (Onkyo, MA-500U). Acoustic responses from the models were captured using the microphone attached to a sound level meter (RION, NL-32). The microphone was positioned on an axis approximately 200 mm away at the height of the nostril of the models. While this distance roughly approximates the closest distance between person-to-person conversation, the acoustic radiation of the mouth and nostril can be regarded as radiation from a simple source [5]. The A-weighted sound pressure level was set to approximately 70 dB, which is more than 40 dB above the room ambient noise.

We used a USB external audio interface (Roland, UA-25EX) to both supply the excitation signal from the PC and to take the signal response from the microphone to the PC. The sampling frequency of 48 kHz and a 16-bit quantization were used in the signal processing of the measurements.

Figure 1: Measurement setup for nasal cavity model.

Figure 2: Measurement setup for head-shaped /a/ model.

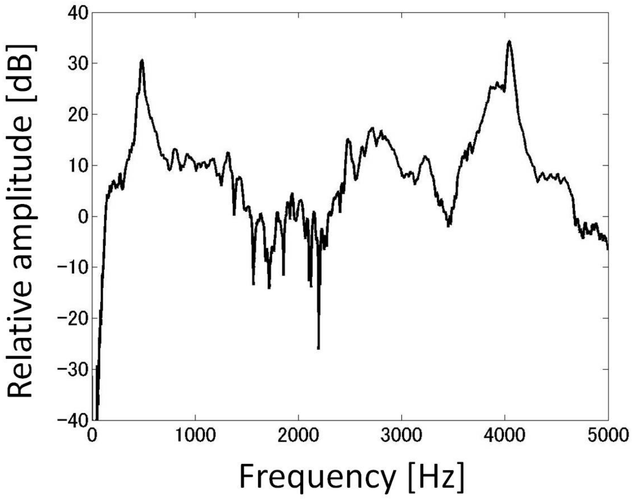

Figure 3: Frequency response of nasal cavity model.

The responses of the head-shaped /a/ model were measured from a closed velopharyngeal opening set to 0° through a maximum opening of 40° in 10° increments.

Figure 1 shows the experimental setup for the nasal cavity and Fig. 2 shows that for the head-shaped /a/ model.

2.3 Results

Figure 3 shows the frequency response of the nasal cavity model derived from the measured impulse response; the resonance frequency of the nasal cavity model was approximately 482 Hz. Figure 4 shows the frequency responses of the head-shaped /a/ model as the velopharyngeal opening is increased.

The frequency responses in Fig. 4 as the degree of the velopharyngeal opening increases reveal that

1) the F1 frequency increases but its amplitude decreases,
2) the peak frequency around 500 Hz increases and its amplitude also increases, and
3) an anti-resonance dip starts to appear.

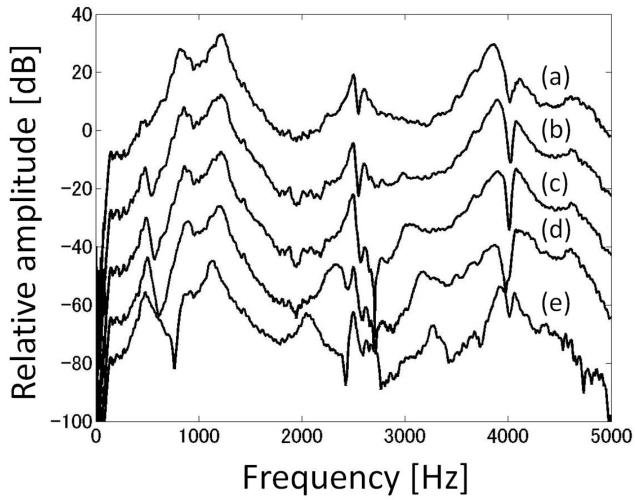

Figure 4: Frequency responses of head-shaped /a/ model with increasing velopharyngeal opening from 0° to 40° in 10° increments (plotted by shifting 20 dB apart).

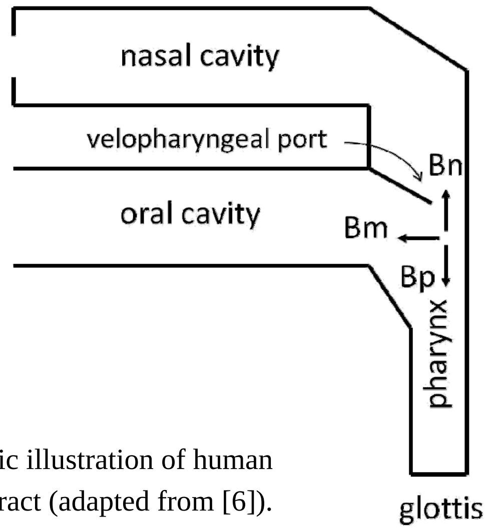

Figure 5: Schematic illustration of human vocal tract (adapted from [6]).

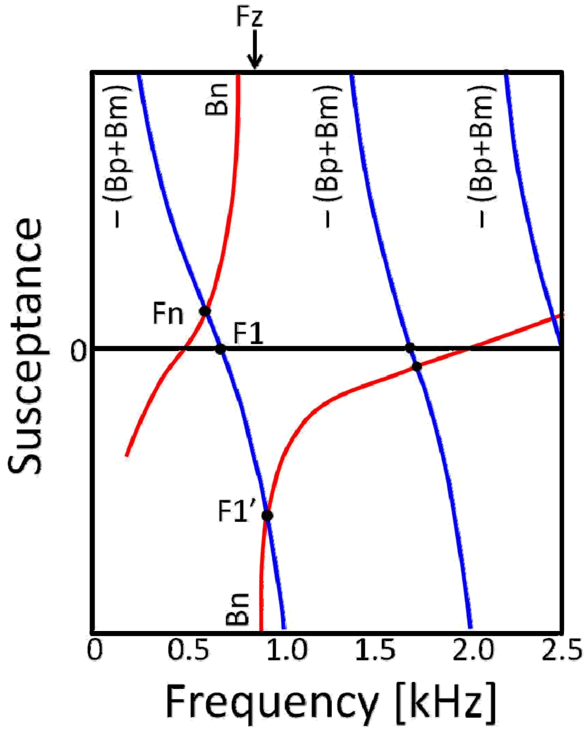

Figure 6: Susceptance curves as a function of frequency (adapted from [5]).

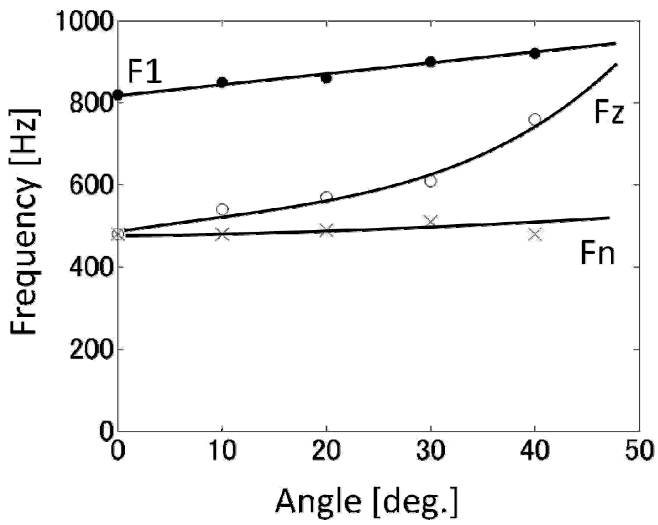

Figure 7: F1, Fn, and Fz shifts as the degree of the velopharyngeal port opens.

3 Discussion

In a 1960 study [6], Fujimura modeled a vocal tract with an electric circuit, paying particular focus to susceptances (as shown in Fig. 5), where Bn was susceptance looking into the nasal cavity, Bm was susceptance looking into the oral cavity, and Bp was susceptance looking into the pharyngeal cavity. The pole and zero frequencies can be determined by the Bn curves and the – (Bp + Bm) curves, as shown in Fig. 6. In this section, we discuss the spectral change observed in the F1 region with the velopharyngeal coupling.

The transfer function of the pharynx and the oral cavity have poles where Bp + Bm = 0. These poles correspond to "formants": F1, F2, and so forth. The transfer function of the nasal cavity has poles where Bn = 0. In the previous section, the first resonance peak of the nasal cavity model was approximately 500 Hz. The entire transfer function has poles where Bn + Bm + Bp = 0. The locations can be found where the Bn curves and – (Bp + Bm) curves intersect. We call these the nasal poles (Fn). In contrast, the transfer function of the pharynx and the oral cavity have zeros when Bn goes on to infinity. Likewise, the transfer function of the pharynx and the nasal cavity have zeros when Bm or – (Bp + Bm) go on to infinity. We call these nasal zeros (Fz).

The frequency shifts of F1, Fn, and Fz for vowel /a/ are plotted in Fig. 7 based on the experimental results in Section 2. The smoothed curves are also plotted schematically in the same figure. The starting frequencies of the curves of F1, Fn, and Fz were approximately 820, 480, and 480 Hz, respectively, when there was no velopharyngeal coupling. Then, as the velopharyngeal port opened, F1 and Fn gradually increased while Fz rapidly increased. This resulted in the original F1 peak diminishing due to the Fz and the Fn peak becoming dominant, taking the place of F1 [4,5,7]. More specifically, the dramatic change in these curves occurs from 30° to 40°. This means that the perceived nasality of vowel /a/ becomes clearer when the velopharyngeal port is greater than a certain area; however, when the area is less than that, the nasality of vowel /a/ is less perceived.

4 Conclusion

When we teach acoustics in speech science and phonetics, an intuitive explanation is needed. Because the theory behind it is rather complicated, we need to combine effective demonstrations using physical models with the nasal cavity and conceptualized diagrams to explain how the spectrum is affected by nasalization. In the future, we will continue to fine-tune explanations of the spectral change for even better understanding.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Number 24501063.

References

[1] T. Arai, "Education system in acoustics of speech production using physical models of the human vocal tract," *Acoust. Sci. & Tech.*, 28(3), 190-201, 2007.

[2] T. Arai, "Education in acoustics and speech science using vocal-tract models," *J. Acoust. Soc. Am.*, 131(3), 2444-2454, 2012.

[3] T. Arai, "Mechanical models of the human vocal tract," *Acoustics Today*, 9(4), 25-30, 2013.

[4] T. Arai, "Acoustic analysis of formant shifts in nasalized vowels," *The Phonetician*, 104-105, 2012-I-II, 39–50, 2012.

[5] K. N. Stevens, *Acoustic Phonetics*, Cambridge, MA, MIT Press, 1998.

[6] O. Fujimura, "Spectra of nasalized vowels," *Research Laboratory of Electronics Quarterly Progress Report*, MIT, 58, 214–218, 1960.

[7] S. Maeda, "Acoustics of vowel nasalization and articulatory shifts in French nasal vowels," in *Phonetics and Phonology: Nasals, Nasalization, and the Velum*, 5, M. K. Huffman and R. A. Krakow, eds., San Diego, CA, Academic Press, 147–167, 1993.