INTERSPEECH 2021: Show & Tell Contribution
30 August – 3 September, 2021, Brno, Czechia

Downsizing of vocal-tract models to line up variations and reduce manufacturing costs

Takayuki Arai

Sophia University, Tokyo, Japan
arai@sophia.ac.jp

Abstract

Demonstrating vowel production with physical models of the human vocal tract is a part of intuitive education in speech science. The adult male vocal tract was most often used as a model in the past because of the limited availability of physical models, but discussions on different vocal tract sizes were ongoing. Therefore, we focused on downsizing the vocal-tract models in this study, especially the straight models. We reduced the cross-sectional area function for the sliding three-tube model (including the total length) to female adult and child sizes. Furthermore, we created fixed straight models of similar dimensions for the five Japanese vowels. We found that the intelligibility of each model was preserved as long as the ratios of the cross-sectional areas were maintained even if the cross-sections were less than the average human sizes. This indicates that we can reduce the cost of manufacturing the models, as cost is typically a barrier when the models are used for pedagogical purposes.

Index Terms: physical models of the human vocal tract, straight model, vowel production, education in speech science

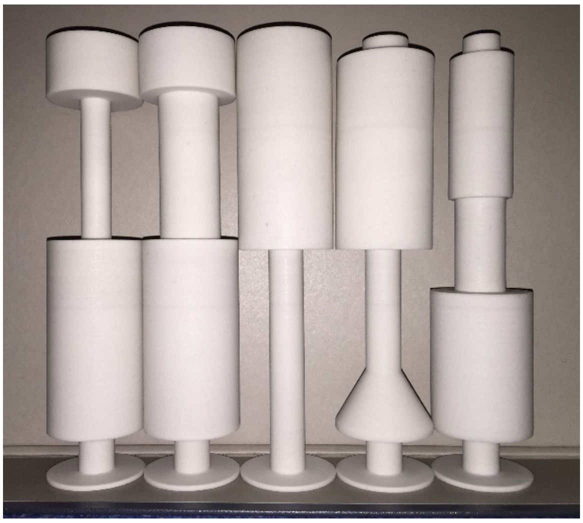

Figure 1: *Set of VTM-T20 models (/i/, /e/, /a/, /o/, and /u/ from the left). They are designed based on VTM-N20 models originally from Chiba and Kajiyama's measurements and simplifications [4]. Further simplifications were applied to achieve the VTM-T20 designs for pedagogical purposes.*

1. Introduction

We developed a series of physical models of the human vocal tract, including the VTM-N20 and VTM-T20 straight models (e.g., [1–3]). They successfully demonstrate the source and the filter of speech production and the relationship between vocal-tract configurations and vowel qualities. Figure 1 shows a set of VTM-T20 models. In this figure, five tubes correspond to the five Japanese vowels: /i/, /e/, /a/, /o/, and /u/ (from left to right). The VTM-T20 models were designed based on VTM-N20 models, originally from Chiba and Kajiyama's measurements and simplifications [4]. We applied further simplifications to the designs of the VTM-N20 models to configure the VTM-T20 models for pedagogical purposes. Simpler models make theories on the source and the filter of speech production and the relationship between the configurations and vowel qualities easier to understand. Therefore, we created the simple designs for VTM-T20 in Fig. 1.

When we teach speech science, we must discuss size and speaker normalization. Even if vocal tract sizes differ between adult male, adult female, and children, we are able to recognize the same vowels. To teach about this, we use sliding three-tube models, or VTM-S20 models, with different tube lengths [2]. An example of a VTM-S20 model is shown in Fig. 2: a simple combination of the outer straight tube and the inner slider. When we position the inner slider in different locations inside the outer tube, we can produce different vowel qualities. Changing the diameter of the inner slider also produces different qualities, such as vowels /i/ vs. /e/.

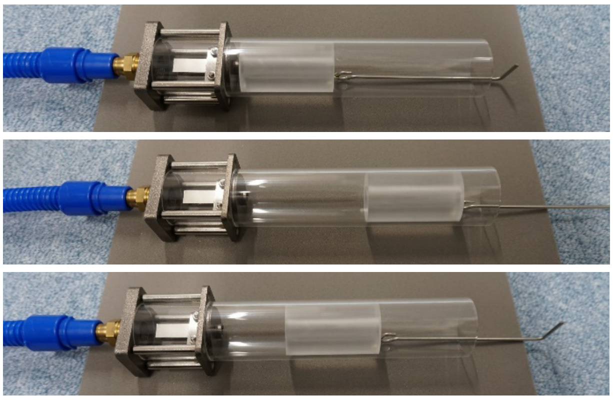

Figure 2: *VTM-S20 model (/a/, /i/, and /u/ from the top). The locations of the inner slider inside the outer tube differ for different vowels.*

VTM-S20 models with different tube lengths can demonstrate different sizes and produce a similar set of vowels (a fundamental frequency that corresponds to the tube length must be chosen.)

Thus, we can teach several topics in speech science with the aforementioned models depending on what we want to emphasize. We ask the following questions in this study:

1) Are we able to reduce the diameters of the VTM-S20 model?
2) Are we able to design a smaller set of VTM-T20 models?

2. VTM-S24

Figure 3 shows our VTM-S24 model. The diameter of the outer tube is 24 mm, and the length of the outer tube is 145 mm. The diameter of the inner slider is 18 mm, and the length of the inner slider is 40 mm. The vowel quality of the output sounds is as intelligible as that with the VTM-S20.

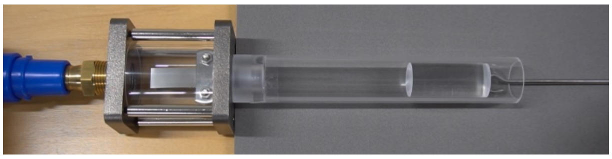

Figure 3: *VTM-S24 model. Instead of the 40-mm diameter and 170-mm or 190-mm length used for VTM-S20 model, the diameter is 24 mm, and the length of the outer tube is 145 mm.*

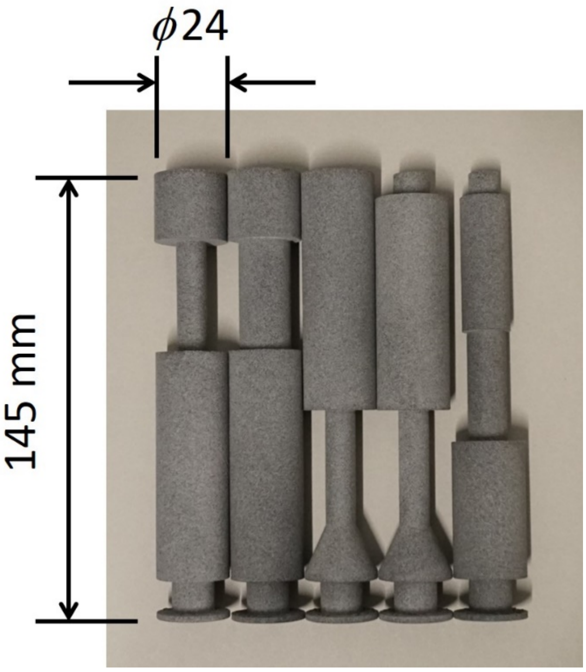

Figure 4: *Set of VTM-T24 models (/i/, /e/, /a/, /o/, and /u/ from the left). The maximum diameter of each tube is 24 mm, and the length of each tube is 145 mm.*

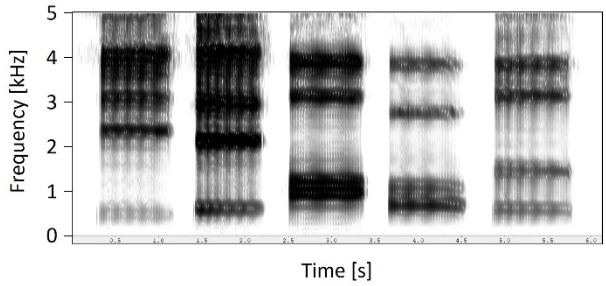

Figure 5: *Spectrogram of output sounds produced by each VTM-T24 model (/i/, /e/, /a/, /o/, and /u/ from the left).*

3. VTM-T24

Figure 4 shows a set of our VTM-T24 models. The dimensions of our VTM-T24 models were based on the VTM-S24 model. Therefore, the maximum diameter of each tube is 24 mm, and the length of each tube is 145 mm. Although the diameters of the VTM-T24 models are less than the average size of human vocal tracts, the output sounds from the models are as intelligible as the VTM-T24 models. Figure 5 shows a sound spectrogram of output sounds from the models: /i/, /e/, /a/, /o/, and /u/. As you can see in this figure, the formants are clearly shown in the proper frequency ranges.

4. Discussion and Conclusion

In this study, we reduced the dimensions of VTM-S20 and VTM-T20 models and designed VTM-S24 and VTM-T24 models. We determined that the quality of vowels was maintained between the VTM-S20/T20 and the VTM-S24/T24 models. When using the physical models in educational situations, we had to consider manufacturing costs. Therefore, we asked the following question:

"How much cheaper are we able to make the VTM-T24 than the VTM-T20?"

We estimated the reduction rate of the total manufacturing costs between the VTM-T24 and VTM-T20 models to be 48%.

5. Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Numbers 18K02988, 20K03074, and 21K02889.

6. References

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

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

[3] Arai, T., "Vocal-tract models and their applications in education for intuitive understanding of speech production," *Acoust. Sci. Tech.*, 37 (4), 148–156, 2016.

[4] Chiba, T. and Kajiyama, M., *The Vowel: Its Nature and Structure*, Tokyo-Kaiseikan, Tokyo, 1941–42.