

Vocal-tract model with two directions: Static design for a dummy head and dynamic design for a speaking machine

Takayuki Arai

Sophia Univ. (Japan)
arai@sophia.ac.jp

Abstract

Physical models of the human vocal tract have been developed for many purposes, including education in acoustics and phonetics, speech and language pathology, and speech science/technology. To cover different application areas, variations are needed. When organizing the types of vocal-tract models developed for such a wide range, one of the major axes is static vs. dynamic aspects. In this study, we demonstrate two models that are the current extremities of the two directions of this axis. The first model is an ultimate static one in which the vocal-tract configuration is one vowel. This model is embedded in a dummy head so that it can be applied to produce exactly the same vowel repeatedly or for a very long duration with a realistic directional radiation pattern. The second model is a recently developed dynamic one with several blocks simulating articulatory movements. The cam mechanisms help to change the vocal-tract configurations in real time.

Index Terms: physical models of the human vocal tract, static model, dummy head, dynamic model, speaking machine

1. Introduction

As described in [1–3], we have developed several different types of physical models of the human vocal tract so far. Because of the variations, they cover a wide range of areas and are utilized for many different purposes. For example, we have shown that straight tube models, which are a simplified version of the human vocal tract, have advantages for demonstrating acoustic theories thanks to their simplicity [3]. Bent models, such as the one shown in Fig. 1 [4], are also useful because the vocal tract is bent at about 90 degrees in the middle of its length, which is more realistic to the human vocal tract. Those who are studying speech science or phonetics can easily imagine where the vocal tract is located in the head with this model.

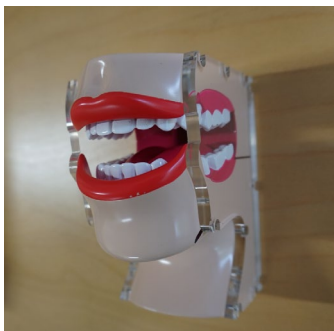


Figure 1: Anatomical-type bent model [4].

Models like the one in Fig. 1 are static and their configuration never changes. In contrast, there is another type of model in which the shape changes temporally. Models of this latter type, which we call dynamic models, have advantages for showing dynamics in speech production. In the previous Show and Tell session in 2024, we presented a preliminary version of a block-type dynamic model [5].

Thus, the axis of static vs. dynamic aspects of vocal-tract models is considered to be one of the major axes. In this study, we introduce two of the cutting-edge extremities of the static-dynamic axis among different types of vocal-tract models.

2. Two types of models

2.1. Static model embedded in a dummy head

Figure 2 shows the dummy head used for this study, where the bent vocal-tract model in Fig. 1 is embedded in the dummy head. For the neck part, a reed-type sound source, SS-R30 [6], is connected to the hole at the bottom of the vocal tract. Because the vocal-tract configuration of the bent model is designed for vowel /a/, the dummy head in Fig. 2 also produces the same vowel.

2.2. Block-type dynamic models

Figures 3 and 4 show two of the block-type dynamic models utilized for this study: VTM-UT30-D6 and VTM-UT30-D9. These models are newer versions of VTM-UT45-D6, which we proposed in 2024 [5]. The VTM-UT30-D6/D9 models have smaller dimensions, with the vocal-tract length/width of 135 mm/30 mm instead of 170 mm/45 mm, and the block depth of 20 mm instead of 25 mm. The cam mechanisms set underneath the model are also improved and now use 3D-printed discs, so that we can easily switch one phrase to another.



Figure 2: Dummy head with the bent model (right).

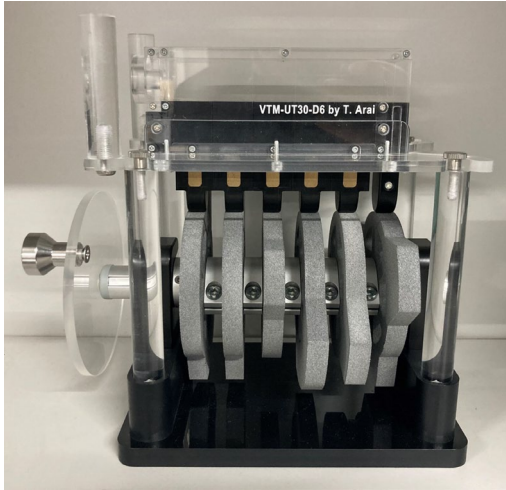


Figure 3: VTM-UT30-D6 model with the cam mechanisms set underneath.

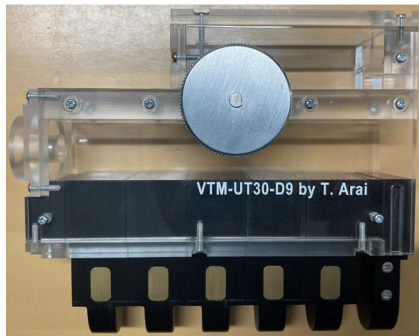


Figure 4: VTM-UT30-D9 model with a nasal cavity placed on top of the main tract.

3. Experiments

3.1. Dummy head with the bent model

One of the applications of this dummy head with the bent model is to produce exactly the same vowel repeatedly or for a very long duration with a realistic directional radiation pattern. Therefore, we tested it to produce vowel /a/ for more than one minute, which is not possible for a human speaker. To do this, an air compressor was used. Figure 5 shows a spectrogram of the output sound from the dummy head that has directional radiation characteristics. As we can see, steady-state patterns were observed. Furthermore, the directional radiation patterns measured from this dummy head should show more realistic results than the ones obtained from a dummy head with an embedded loudspeaker. Such measurements can be utilized to design radiation patterns of human speakers or singers in a virtual-reality environment.

3.2. VTM-UT30-D6/D9 models

With the VTM-UT30-D6 model, different phrases can be produced by changing different sets of rotating discs. Figure 6 shows spectrograms of two phrases: “How are you?” in English and “Arigatō” in Japanese. The advantage of the VTM-UT30-

D9 model is that it can produce not only oral but also nasal sounds. Figure 7 shows a spectrogram of vowel /a/ as the degree of nasopharyngeal coupling increased in time. A pole-zero pair due to nasality can be observed towards the end of the utterance.

4. Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Number 24K06423 and Faculty of Science and Technology Proposal-based Research Fund, Sophia University.

5. References

- [1] T. Arai, “Education in acoustics and speech science using vocal-tract models,” *J. Acoust. Soc. Am.*, vol. 131, no. 3, pp. 2444–2454, 2012.
- [2] T. Arai, “Vocal-tract models and their applications in education for intuitive understanding of speech production,” *Acoust. Sci. Technol.*, vol. 37, no. 4, pp. 148–156, 2016.
- [3] T. Arai, “Education in basic acoustics for acoustic phonetics and speech science,” *J. Acoust. Soc. Am.*, vol. 152, no. 5, pp. 2746–2757, 2022.
- [4] T. Arai, “Vocal-tract model with static articulators: Lips, teeth, tongue, and more,” *Proc. INTERSPEECH*, 4028–4029, 2017.
- [5] T. Arai, R. Suzuki, C. Earp, S. Tsuji, and K. Ochi, “Production of phrases by mechanical models of the human vocal tract,” *Proc. INTERSPEECH*, 987–988, 2024.
- [6] T. Arai, “Sound sources used in speech production research with physical models of the human vocal tract,” *Proc. International Workshop on the History of Speech Communication Research (HSCR)*, 79–84, 2019.

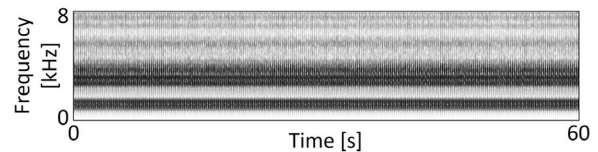


Figure 5: Spectrogram of the output sound from the dummy head.

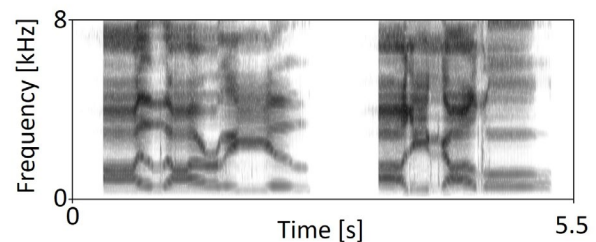


Figure 6: Spectrograms of two phrases: “How are you?” in English (left) and “Arigato” in Japanese (right).

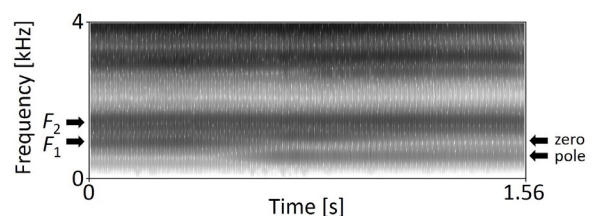


Figure 7: Spectrogram of vowel /a/ with the gradually increasing degree of nasopharyngeal coupling.