INTERSPEECH 2017: Show & Tell Contribution
August 20–24, 2017, Stockholm, Sweden

Vocal-tract Model with Static Articulators: Lips, Teeth, Tongue, and More

Takayuki Arai

Department of Information and Communication Sciences
Sophia University, Tokyo, Japan
arai@sophia.ac.jp

Abstract

Our physical models of the human vocal tract successfully demonstrate theories such as the source-filter theory of speech production, mechanisms such as the relationship between vocal-tract configuration and vowel quality, and phenomena such as formant frequency estimation. Earlier models took one of two directions: either simplification, showing only a few target themes, or diversification, simulating human articulation more broadly. In this study, we have designed a static, hybrid model. Each model of this type produces one vowel. However, the model also simulates the human articulators more broadly, including the lips, teeth, and tongue. The sagittal block is enclosed with transparent plates so that the inside of the vocal tract is visible from the outside. We also colored the articulators to make them more easily identified. In testing, we confirmed that the vocal-tract models can produce the target vowel. These models have great potential, with applications not only in acoustics and phonetics education, but also pronunciation training in language learning and speech therapy in the clinical setting.

Index Terms: physical model, vocal-tract model, vowel production, articulators, education in acoustic phonetics

1. Introduction

So far, we have developed a series of physical models of the human vocal tract (e.g., [1-3]), and they successfully demonstrate various theories, phenomena and mechanisms in acoustic phonetics. For example, to illustrate the source-filter theory of speech production, we can combine a target vocal-tract model with a sound source to produce a vowel sound. We can also relate vocal-tract configuration to vowel quality, and we can estimate formant frequencies. Some of our earlier models were simple and demonstrated only few target themes. The sliding three-tube model [2] was developed along this line. Other models simulated human articulation in various ways, such as the flexible-tongue model [3]. In this study, we have designed a hybrid model which produces a target vowel and illustrates human articulation in a more broad sense.

2. Designs

The proposed hybrid model is static, and each model of this type produces only one vowel. As well as making a target vowel, the model simulates the way the human articulators produce sound with the lips, teeth, and tongue. Figure 1 shows the design of the proposed model from the side view, having the same gross dimensions. The main part (the colored portion in this figure) has left and right sagittal surfaces, and its width (the depth in the figure) is 60 mm.

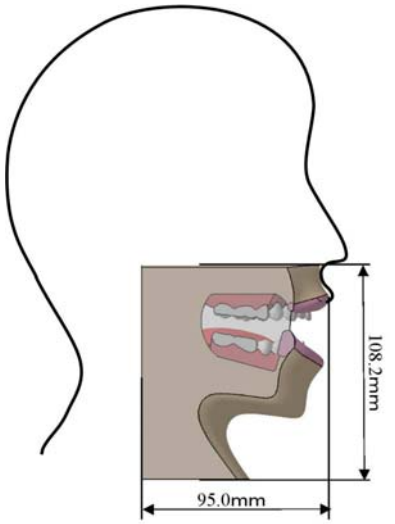

Figure 1: *Design and gross dimensions of the proposed model of the human vocal tract.*

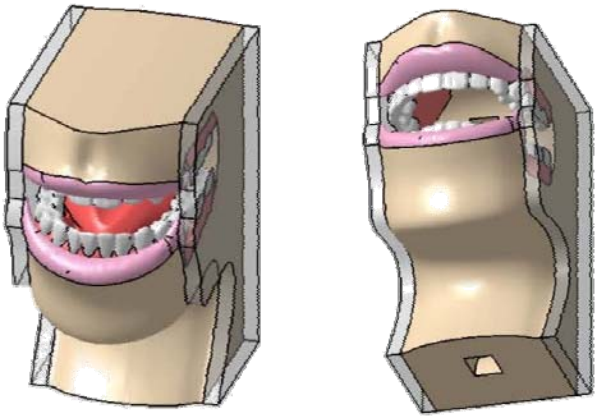

Figure 2: *The proposed model viewed from different angles*

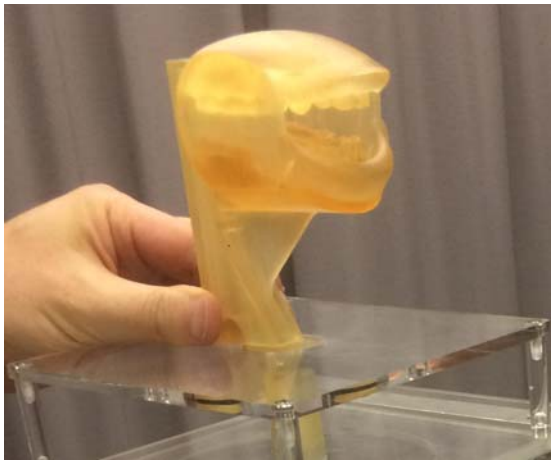

Figure 3: *An acoustically equivalent model of the proposed model for vowel /a/.*

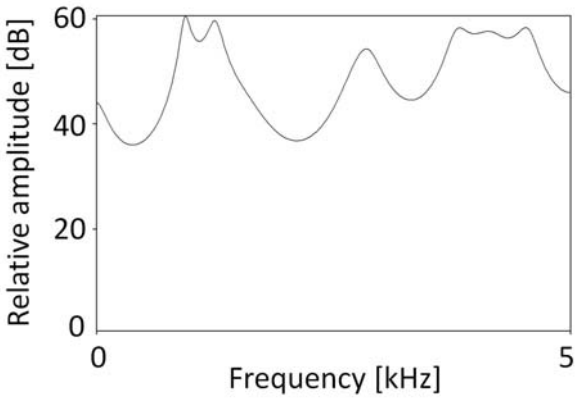

Figure 4: *Spectral envelope based on the LPC of an output sound recorded through the acoustically equivalent 3-D model of the proposed model for vowel /a/.*

Figure 2 shows the main part viewed from two different angles. As you can see in Figure 2, on the outer side of both sagittal surfaces, transparent plates are attached to cover the holes in the sagittal surfaces. Because the cover plates are transparent, the inside of the oral cavity is visible from the outside. We also added coloring for each articulator for easier identification.

The fundamental vocal-tract configuration of the model in Figure 1 was taken from the head-shaped model for the vowel /a/ in our previous studies (e.g., [1]).

3. Acoustic analysis

To confirm that the proposed model can produce the target vowel /a/, we first created an acoustically equivalent model with 3-dimensional printing. Figure 3 shows the 3-D printed version of the model. We used a reed-type sound source as an input sound. The output signals from the model were recorded digitally with a digital audio recorder (Marantz, PRM661MK II) with a microphone (Sony, ECM-MS957). The original 48 kHz sampling frequency for the recordings was retained for the perceptual evaluation but downsampled to 10 kHz for the acoustic analysis.

The acoustic analysis was done by computing spectral envelopes based on linear-predictive coding, or LPC. Figure 4 shows the resulting spectral envelope by Praat [4]. The average frequencies of the first and second formants (F1, F2) were 940 Hz and 1266 Hz, respectively.

4. Conclusions

In this study, we have designed a new, hybrid model which incorporates both directions taken by our previous models: simplification of production and simulation of human articulation. The proposed hybrid model is static and only produces one vowel per model. However, the model simulates the human articulators, including the lips, teeth, and tongue. Finally, we confirmed that the vocal-tract model can produce the target vowel. This model has great potential for applications in education in acoustics and phonetics, as well as pronunciation training in language learning and speech therapy in the clinical domain, because it is more intuitive than some of our previous models.

5. Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Numbers 15K00930.

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] Boersma, P., "Praat, a system for doing phonetics by computer," *Glot International*, 5:9/10, 341-345, 2001.