Two Extensions of Umeda and Teranishi's Physical Models of the Human Vocal Tract

Takayuki Arai

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

Abstract

The physical model designed by Umeda and Teranishi simulating an arbitrary shape of the human vocal tract was a straight tube with a set of plastic plates inserted from one side. Although this model has the advantage that users can configure any shape of the vocal tract, manually manipulating several plates simultaneously is difficult. In this study, we present two models extending Umeda and Teranishi's work to overcome this disadvantage. The first model has a straight tube similar to the original Umeda and Teranishi's model, but the weight of the plates enables them to return to resting position automatically. The second model has a bent tube with the oral and pharyngeal cavities connected at 90 degrees. This feature simulates the actual human vocal tract. The plates move back to their original positions by means of spring coils. In both cases, the plates' automatic return movement facilitates manual manipulation as compared to the original Umeda and Teranishi's model.

Index Terms: physical model, vocal-tract model, Umeda and Teranishi's model, vowel production, education in speech science

1. Introduction

Umeda and Teranishi [1] designed a physical model of the human vocal tract as shown in Fig. 1. A set of plastic plates (bars) are inserted from one side of the model to simulate an arbitrary shape of the vocal tract. This figure also shows that another set of plates are inserted, which change the shape of the nasal cavity.

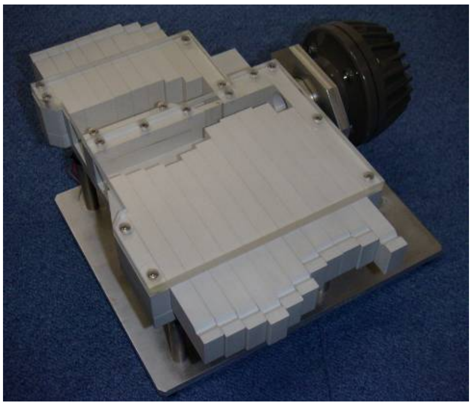

Figure 1: *Umeda and Teranishi's model. A set of plastic plates are inserted from the side.*

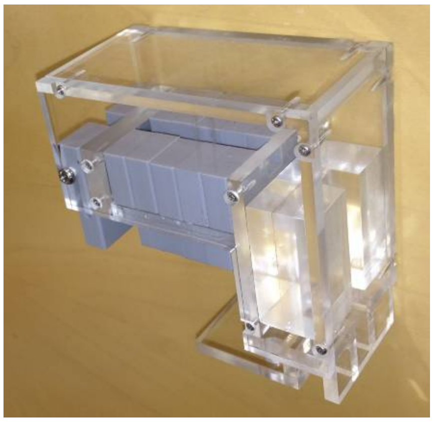

Figure 2: *Front-cavity model for simulating bunched English /r/ [4].*

One of the advantages of Umeda and Teranishi's model (U&T model) is that users can manually manipulate the model to achieve an arbitrary vocal-tract shape, compared to other static models [2,3]. We reported its usefulness for education in speech science. However, one disadvantage of the U&T model is that it is difficult to simultaneously manipulate multiple plates manually. Therefore, in this study, we designed two extensions of the U&T model based on the front-cavity model originally introduced in [4].

2. Designs

Figure 2 shows the front-cavity model in [4], in which the model was used to simulate bunched English /r/. In this model, a set of plastic plates are inserted in the front-cavity, and the plates can move back to their original position by their own weight. Therefore, the first extended U&T model proposed in this study is straight, like the original U&T model, but the set of plates return to resting position automatically. The second extended model is the bent model like the front-cavity model in [4], but the set of plates return to resting position by means of spring coils.

Figure 3 shows the first model, in which the 15 plates are lined up next each other. In this model, there are slight gaps between two adjacent plates, so that the plates are in the lowest position when at rest. Users can manipulate the plates manually by pushing up the plates from the bottom part of the model. The plates move back to the resting position by their own weights "automatically"; this was not possible with the U&T model.

Copyright © 2015 ISCA

Figure 3: *The first extended model based on Umeda and Teranishi's model.*

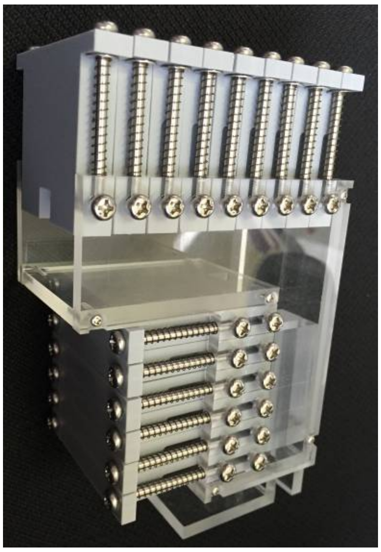

Figure 4: *The second extended model based on Umeda and Teranishi's model.*

Figure 4 shows the second model. This model is bent, and there are 9 plates lined up in the oral cavity and 6 plates lined up in the pharyngeal cavity. In this model, each plate stays at the resting position due to the spring coil. A user can manipulate the plates by pushing the plates manually, and the plates return to resting position by the restoration force of the springs.

3. Discussion

We tested the two extended models in Figures 3 and 4 to produce certain target sounds using manual manipulation. First, vowels, especially the five Japanese vowels, can be produced relatively easily with both of the two models. Second, vowel sequences and diphthongs can also be produced with both of the two models. This was not achieved with the original U&T model, because the quick movement required of multiple plates was extremely difficulty to achieve manually. In the case of our computer-controlled version of the U&T model in [5], we are able to produce dynamic motion of the vocal-tract shape. However, with the original model, fast movements are not possible. Figure 5 shows the spectrogram of the produced vowel sequence of /iaia/ from the first model. As shown in this figure, smooth transitions were observed.

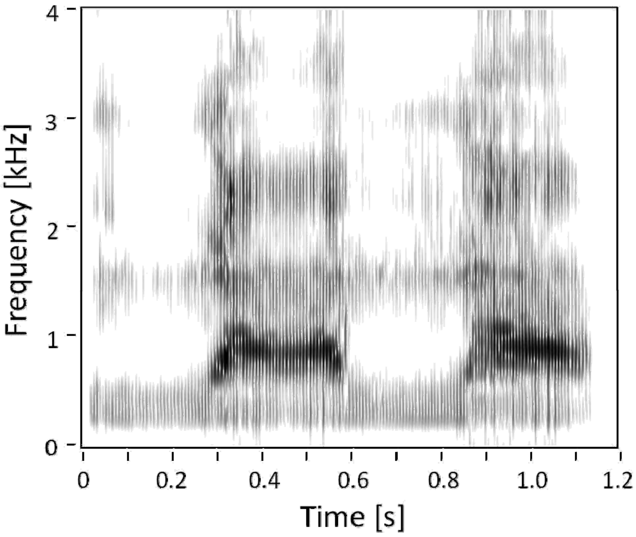

Figure 5: *The spectrogram of the vowel sequence /iaia/ produced by the first extended model.*

Some consonants were also tested. We were able to produce approximants relatively successfully, such as, /w/, /j/, /r/, and /l/. The latter two sounds were also achieved with the front-cavity model in [4]. However, with this model, the back-cavity was fixed and there were limitations with respect to the sounds that could be produced. With the two proposed models, however, all dimensions of the vocal tract can be changed, and any combination of approximants and vowels can be produced.

4. Conclusions

In this study, we designed and tested two versions of the extended models based on the U&T model. With both models, the vocal-tract configuration returned to a uniform shape "automatically," allowing users to manually manipulate the models more easily as compared with the original model. The quality of sounds, however, was sometimes not perfect, and we will make further improvements in the future.

5. Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Numbers 24501063 and 15K00930. I would also like to thank Terri Lander for her support.

6. References

[1] Umeda, N. and Teranishi, R., “Phonemic feature and vocal feature: Synthesis of speech sounds, using an acoustic model of vocal tract,” *J. Acoust. Soc. Jpn., 22(4)*, 195-203 1966.

[2] 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.

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

[4] Arai, T., “Retroflex and bunched English /r/ with physical models of the human vocal tract,” *Proc. of INTERSPEECH*, 706-710, 2014.

[5] Arai, T., “Mechanical vocal-tract models for speech dynamics,” *Proc. of INTERSPEECH*, 1025-1028, 2010.