Prototype of a Vocal-Tract Model for Vowel Production Designed for Education in Speech Science

Takayuki Arai, Nobuyuki Usuki, and Yuji Murahara

Dept. of Electrical and Electronics Eng.,
Sophia University, Tokyo, JAPAN
arai@sophia.ac.jp

ABSTRACT

We present a manipulative tool for use in education in speech science. The tool consists of several, square, plastic disks each of which has holes of various diameters. Combined, the holes in 10–17 disks simulate the vocal tract by creating an acoustic tube. Students may study the effect the shape of an acoustic tube has on acoustic output by reordering the disks. After demonstrating the disks in an actual classroom, results show that the tool helped students grasp the relationship between vocal tract configuration and acoustic output. This suggests that students are better able to understand abstractions regarding the acoustics of speech if they have access to a 3-dimensional model such as ours.

1. Introduction

Successfully designed educational tools help students understand phenomena in the various fields of study they illustrate. As educators have learned more about cognitive processes and have applied that knowledge in the classroom, they have augmented traditional teaching methods such as straight lecture with more hands-on projects and lesson plans.

As societies become more technologically advanced, there is a growing trend toward computer-based educational tools. One of the benefits of computer aided learning is that lessons may be tailored to individual learning styles. Further, computer aided learning software can more successfully incorporate different learning styles into a given lesson than is possible with most traditional paper-based textbooks. However, a significant disadvantage of computer-aided tools is that real world conditions are inherently unpredictable, and it is difficult to design exercises without a specific knowledge of the context of their use.

The use of manipulatives in the classroom avoids some of the problems inherent in computer-aided learning, because learning conditions are more predictable. In addition, manipulatives are hands-on, which facilitates the learning process for tactile learners. Manipulatives are particularly suitable for subjects involving abstract concepts.

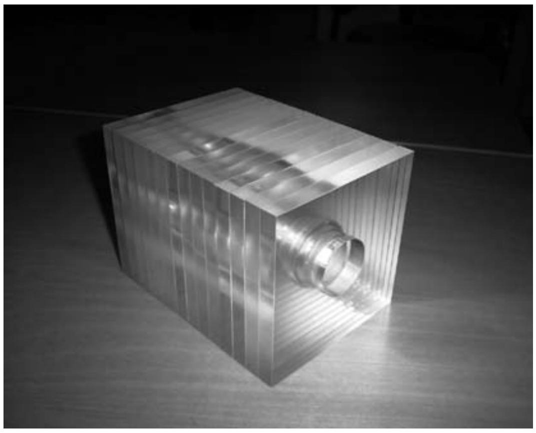

Figure 1: Prototype of vocal-tract model for vowel /a/.

In this paper, we present a manipulative tool for use in education in speech science. The tool is intended to help students grasp the relationship between the configuration of the vocal tract and acoustic output. This is an important topic for speech science students, and one that is difficult to grasp because the vocal tract is largely unseen except perhaps by X-ray or MR images. Our thinking is that students will be better able to gain an understanding of the abstract concepts involved in the acoustics of speech if they have access to a 3-dimensional model such as ours.

The tool we developed consists of several, square, plastic disks each of which has holes of various diameters. Once combined, a set of 10-17 disks simulates the vocal tract by creating an acoustic tube where the holes line up. By manipulating the disks into various configurations students can study the effect the shape of an acoustic tube has on acoustic output. In Section 2, we describe this prototype of vocal tracts in detail and show their acoustic properties. In Section 3, we describe results after demonstrating them in an actual classroom.

Table 1: Diameter (d) and area (A) for Japanese five vowels.

Index# (from mouth)	/a/ d [mm]	/a/ A [mm^2]	/i/ d [mm]	/i/ A [mm^2]	/u/ d [mm]	/u/ A [mm^2]	/e/ d [mm]	/e/ A [mm^2]	/o/ d [mm]	/o/ A [mm^2]
(1)	32	804	24	452	14	154	24	452	14	154
(2)	26	531	14	154	14	154	18	254	22	380
(3)	30	707	6	28	18	254	20	314	26	531
(4)	34	908	8	50	20	314	16	201	34	908
(5)	38	1134	8	50	22	380	16	201	34	908
(6)	34	908	14	154	22	380	16	201	38	1134
(7)	30	707	18	254	16	201	20	314	34	908
(8)	26	531	26	531	12	113	22	380	28	616
(9)	20	314	30	707	16	201	26	531	22	380
(10)	16	201	30	707	24	452	28	616	16	201
(11)	14	153	30	707	26	531	30	707	14	154
(12)	12	113	30	707	26	531	30	707	16	201
(13)	20	314	30	707	26	531	30	707	20	314
(14)	34	908	30	707	26	531	30	707	26	531

2. Prototypes of Vocal Tracts

Several prototypes of the human vocal tract have been developed. An example of an early vocal-tract model is the mechanical speech synthesizer by Von Kempelem [1]. In this model, vibrating reeds simulate a voiced sound source. Vowel-like sounds are produced with a leather tube, the shape of which is altered to approximate the characteristics of each vowel. Some physical models of vocal tract prototypes have been implemented in laboratory environments for research purposes (e.g., [2]). Museums are another source of vocal tract prototypes (e.g., [3]). However, the latter are not suitable for use in the classroom because they are designed primarily for exhibition.

In this paper we propose a simple prototype of vocal tract models that may be used in the classroom. The model consists of a set of plastic plates, each with a hole in the center. When placed side-by-side the holes in the plates form a tube whose cross-sectional area changes in a step-wise fashion. Each plastic plate is 100mm $\times$ 100mm $\times$ 10mm. The diameters of the holes vary based on measurements by Chiba and Kajiyama [2] for five Japanese vowels. For comparison, we also made several plates having holes of the same diameter to form a uniform tube approximating the central reduced vowel, schwa.

The acoustic tube formed by the disk holes has an opening at each end, corresponding to the two ends of the vocal tube, the mouth and glottis. When a sound originates at the glottis end of the tube a vowel-like sound is emitted from the other end. Fig. 1 shows a prototype by setting up disks next each other.

For the sound source we used an electrical larynx of the type often used by patients whose vocal folds have been removed due to cancer of the larynx. A horn driver can also be used to feed an arbitrary source signal. Figs. 2-7 shows FFT spectra with LPC-based spectral envelopes of six output signals. The spectra were produced by six different sets of vocal-tract models, each having an electrical larynx sound source.

3. Lab Experiments

We used the vocal tract models in a speech science classroom during a lecture in order to investigate the effects our prototypes have in a classroom setting. The majority of the listening audience were members of the Phonetics Lab. as well as students in the department of Electrical and Electronics Engineering. After reviewing acoustic theories such as source-filter theory, acoustic theory of a uniform tube, and perturbation theory, the prototypes were used to demonstrate the theories and to conduct a hands-on experiment.

Because the sound source by itself has no formants, it was channeled through a uniform tube to produce them. The same sound source was also channeled through the different models to produce vowels of varying qualities. The plastic disks are transparent so that the acoustic tube formed by the adjacent holes is visible to the observer. Likewise, the location of the constriction in the tube is visible from the outside of the model. This design allowed the audience to associate the quality of the vowel heard with the location of physical constriction on the model. Furthermore, by channeling sound sources with different frequencies through the tube, the audience observed that the pitch of the outcoming sound is determined by the fundamental frequency of the input signal. These relationships observed between point of constriction and vowel quality as well as between fundamental frequency and pitch are key to understanding source-filter theory.

Perturbation theory [2,4] predicts a shift in formant frequency resulting from constriction(s) in an acoustic tube. The effect of perturbation on formant frequency is governed by whether the constriction is located at a node (region of volume velocity maxima) or an antinode (region of volume velocity minima). The general relationship is as follows [5]:

- A local constriction of the tube near a volume velocity maximum lowers the formant frequency.
- A local constriction of the tube near a volume velocity minimum raises the formant frequency.

By changing the order of the disks to simulate constriction at a node or antinode, the shift in formant(s) can be heard, and it can be seen in a spectral analysis of the outcoming sounds. This exercise helped the audience to understand how formant frequencies are affected by a change in constriction point in the vocal tract, which relates directly to an understanding of how vowels change depending on the location(s) of constriction in the vocal tract.

In general the audience's understanding of vowel production increased after playing with the prototype. The experimenters received a good deal of positive feedback from participants. We conclude form our experience that when used in a classroom environment, our model is particularly effective for increasing student understanding of the acoustic theory of speech production.

4. Summary

We presented a manipulative tool for use in education in speech science. The tool helped students grasp the relationship between the configuration of the vocal tract and acoustic output. Specifically, through hands-on experimentation observers were exposed to three key acoustic concepts which led to a better understanding of 1) the association between vowel quality and location of constriction in the vocal tract 2) the fundamental frequency of the input as a determiner of the pitch of the output and 3) the prediction of the direction of formant frequency change depending on the location of the point of constriction in the tube.

The transparent design of the tool was key to its effectiveness, because the shape of the acoustic tube was visible from the outside of the model. In addition, this prototype has an advantage over models in which tube shape change is effected by squeezing a rubber tube. Specifically, our model allows replicability of precise acoustic tube configurations and, thus, corresponding acoustic outputs.

The lab experiments showed that the prototype of vocal-tract models is effective in a classroom. This is probably true not only for students of speech processing or speech science/phonetics, but also for those without training in these fields. In the future, we would like more feedback from our audiences to discover possible improvements.

5. Acknowledgments

We would like to thank the members of the Phonetics Lab. at Sophia University who participated in the lecture and gave us useful comments. We would also like to express our grateful thanks to Terri Lander of the Oregon Graduate Institute of Science and Technology on her helpful comments.

6. References

[1] B. Gold and N. Morgan, *Speech and Audio Signal Processing*, John Wiley & Sons, 2000.

[2] T. Chiba and M. Kajiyama, *The Vowel: Its Nature and Structure*, Phonetic society of Japan, 1958.

[3] Hamamatsu Science Museum, http://www.kagakukan.city.hamamatsu.shizuoka.jp/tenji/index.html

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

[5] R. D. Kent and C. Read, *The Acoustic Analysis of Speech*, Singular Publishing Group, Inc., San Diego, 1992.

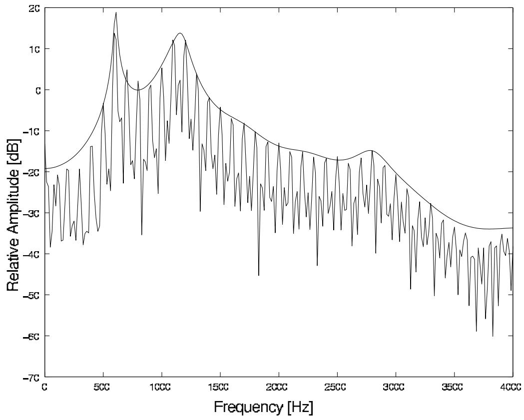

Figure 2: The spectra of vowel /a/ produced by the prototype.

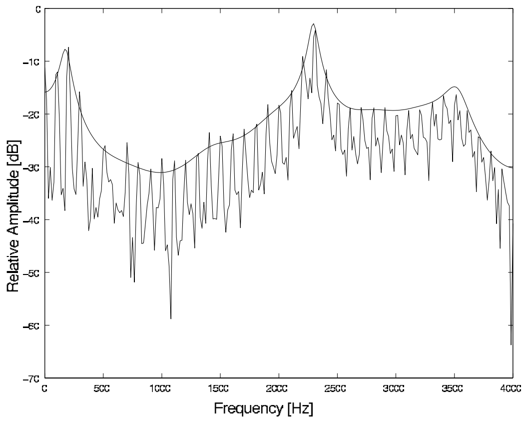

Figure 3: The spectra of vowel /i/ produced by the prototype.

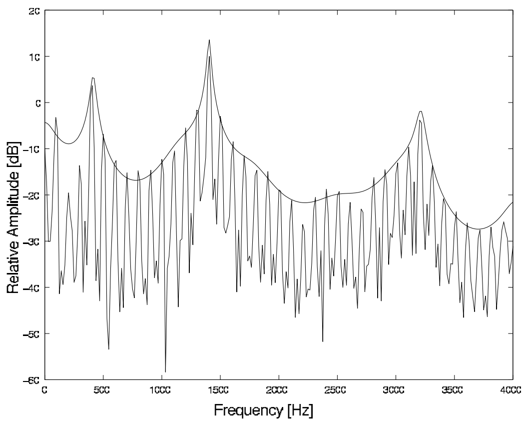

Figure 4: The spectra of vowel /u/ produced by the prototype.

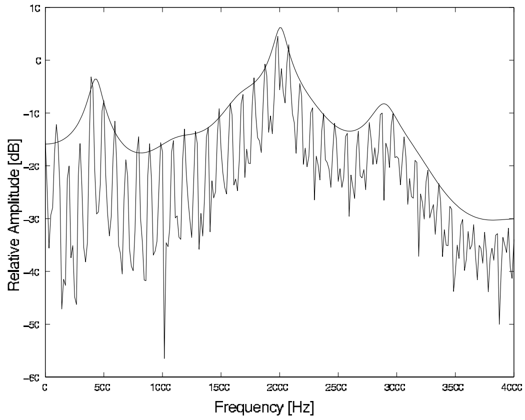

Figure 5: The spectra of vowel /e/ produced by the prototype.

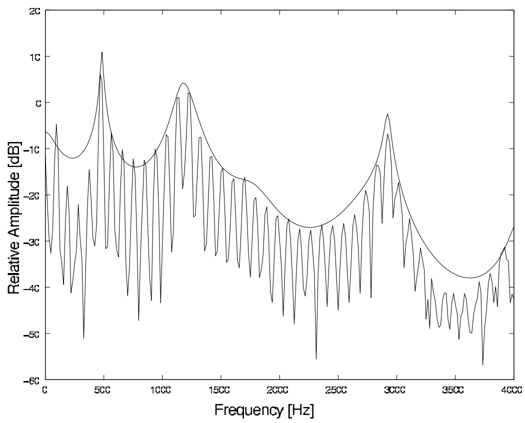

Figure 6: The spectra of vowel /o/ produced by the prototype.

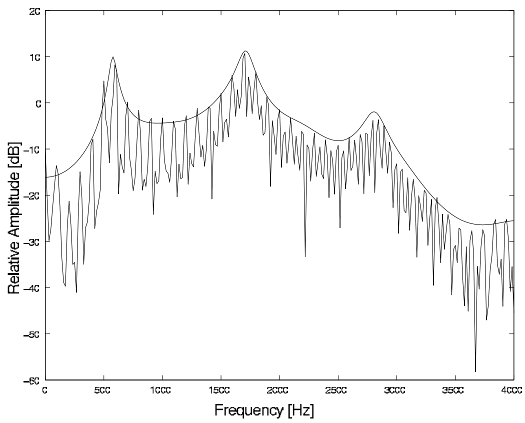

Figure 7: The spectra of vowel 'schwa' produced by the prototype.