

# Education in Acoustics using Physical Models of the Human Vocal Tract

*Takayuki Arai*

Department of Electrical and Electronics Engineering  
Sophia University, Japan  
arai@sophia.ac.jp

## Abstract

It could be said that Acoustics forms a bridge between speech production and speech perception, and as such it lies at the intersection of several speech communication related fields. The field of Speech and Hearing Sciences, for example, deals with speech production and speech perception. Speech Pathology also relates crucially to Acoustics. In addition, the linguistic fields of Phonetics and Phonology have points of intersection in common with Acoustics, particularly Phonetics, which comprises three important subfields: Articulatory Phonetics, Acoustic Phonetics and Auditory Phonetics (related to Psycho-acoustics). Speech technology, including automatic speech recognition, speech synthesis and speech coding, is overlaid as an application of these fields.

Because acoustics is related to so many fields, such classes are generally comprised of students from a wide variety of backgrounds. At Sophia University, the author is teaching acoustics not only to Engineering students but also to students majoring in fields such as Linguistics, Psychology, and Speech Pathology. We believe that an education in acoustics is important not only for college-level students, but also for high-school or potentially even elementary-school students. Therefore, we are motivated to develop intuitive and effective methods for educating students of different ages and from varied backgrounds. Toward this aim, we proposed using physical models of the human vocal tract as educational tools, and we verified their usefulness in the classroom. We have also started exhibiting our models at a science museum targeting for 10-12 year-old children. Furthermore, we identified other areas where the models were needed, for example, as educational tools for speech pathologists and patients with speech disorders and hearing impairment. Finally, we have developed new models for language learning, such as models for the consonants /r/ and /l/, so that second language learners appreciate them, as well.

## 1. Introduction

These days, a variety of tools for teaching Acoustics are widely available including textbooks, recordings,

databases, physical tools, and computer-based tools. People are paying more and more attention to education in Acoustics (e.g., [1,2]); we have reported our attempts at several conferences [3-9].

Important aspects of physical and computer-based tools are as follows. For physical tools, 1) they are real, enabling students to understand the material intuitively; 2) we can show students phenomena in real-time; and 3) students can use them for hands-on experiments. For computer-based tools, 1) we can deal with complex and/or high-speed phenomena virtually; 2) we can visualize unseen phenomena; 3) multi-media environments are useful (e.g., images / video, audio and texts can be displayed); and 4) we can effectively use the World Wide Web.

As I have proposed in [10], the integration of physical and computer-based tools yields great benefit. On a computer, we can simulate a vowel sound from the area function of the human vocal tract [11]. From the same data, we can also design a physical model of the vocal tract that produces a vowel sound. By comparing the two signals, students are able to determine whether a physical phenomenon is supported by a theory.

Thus, both physical and computer-based tools are advantageous for teaching Acoustics. In this paper, we will review the aspects of our proposed physical models of the human vocal tract. At the same time, we will introduce some other physical models of the human vocal tract and discuss their features.

## 2. Physical models of the human vocal tract

Physical tools are, in general, useful for intuitive learning, and their demonstration is usually done in real time. Some can be used for hands-on experiments. We recently developed physical models of the human vocal tract, which provide extremely important real-time examples for intuitive learning. Designing such models is not new. In the 18th century, Kratzenstein and von Kempelen proposed mechanical models for vowel and consonant production [12-15]. In the 19th century, Faber demonstrated his "Talking Head" [16]. More recently, Riesz designed a mechanical talker [13]. Chiba and

Kajiyama (1941), on the other hand, made physical models based on their measurement to show that the shape of the vocal tract determines the quality of vowel sounds [17]. Umeda and Teranishi (1966) also made a physical model with several sliding strips, so that an arbitrary configuration of the vocal tract can be implemented [18]. Some museums have physical models of the human vocal tract (e.g., Exploratorium in San Francisco [19]).

### 2.1. Umeda and Teranishi's model of the human vocal tract

Umeda and Teranishi (1966) made a simple device that simulates a human vocal tract acoustically [18]. The cross-sectional areas of this model may be changed by moving 10-mm (or 15-mm) thick plastic strips, closely inserted from one side, as shown in Fig. 1. The model has a nasal branch as well.

Various vowels and other sustained sounds are produced according to the configuration of the model. Glottal sounds are sent into one end (glottis) of the model and emitted from the other end (mouth). The driving unit of a horn speaker was used as a sound source. By using the model, Umeda and Teranishi investigated phonemic and vocal features of speech.

We used a replica of their model as an educational tool (Fig. 1). We input several sound sources into the model: an impulse train, triangular glottal pulses, white noise, an LPC residual signal and an acoustic signal recorded just outside a speaker's larynx during phonation. For sustained vowels, the strips are properly positioned prior to making sounds. For transient sounds, we can slide the strips along the time while a sound source is fed. Although it was sometimes difficult to

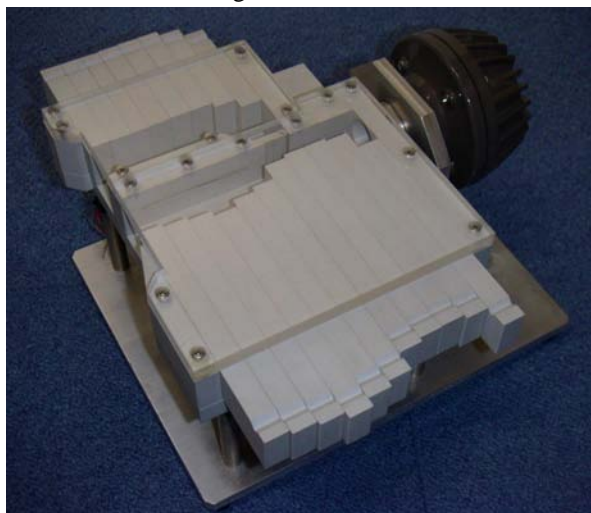


Figure 1: Umeda and Teranishi's model of the human vocal tract. In this figure, each plastic strip is 15 mm wide; the 11 strips are used to form an oral cavity, the total length of which is 165mm. This model has a nasal cavity (the top part). The velopharyngeal coupling is controlled by a rotating a valve.

make perfect diphthongs by sliding them by hand, we were able to produce several degrees of time-varying sounds.

### 2.2. Arai's models of the human vocal tract

Arai (2001) replicated Chiba and Kajiyama's models and showed that they are extremely effective in the class [3]. They proposed a simple but powerful demonstration of vowel production using the physical models and sound sources, such as an electrolarynx ([3-9]; see also [20]). It was confirmed that the models help students to intuitively understand acoustic theories, especially source filter theory [21] and perturbation theory [22].

Fig. 2 shows two types of physical models of the human vocal tract: the cylinder model (on the top) and the plate model (on the bottom). The two models are made of acrylic resin because it is both transparent and easy to sculpt. For the cylinder model, the cavity forms a round bottle-shape, based on the measurements by Chiba and Kajiyama [17]. For the plate model, each plate has a hole in the center so that when placed side-by-side the holes form an acoustic tube, the cross-sectional area of which changes in a step-wise fashion. When a sound source is connected to one end of either of the models, a vowel-like sound is emitted from the other end. We typically use an electrolarynx and a whistle-type artificial larynx (as shown in Fig. 2) for sound sources.



Figure 2: Arai's model of the human vocal tract (top: cylinder model; bottom: plate model with a whistle-type artificial larynx). They are available through NTT Advanced Technology Co. Please see the following URL for more information:

<http://www.sp4win.com/eng-vtm/eng-vtm.htm>

### 3. Effectiveness as educational tools

We believe mechanical models of the human vocal tract are useful for educating students of various backgrounds and ages (e.g., [5]). Physical models such as ours are especially suitable for non-technical students, because they are more intuitive. Also, we believe that education in acoustics is important not only for college-level students, but even high-school or potentially even elementary school students, for whom such hands-on models are even more important.

We have already used Arai's models [3] in the classroom for students majoring in interdisciplinary fields such as Speech and Hearing Sciences, Engineering, Linguistics, Psychology, and Speech Pathology. We confirmed that they are powerful tools for education. Furthermore, we had a chance to teach Acoustics to high-school students by using Arai's models [9]. Finally, we also used Umeda and Teranishi's model in Acoustics class for college and graduate students. In this section, we would like to discuss their usefulness through teaching in real classes.

We confirmed that physical models, when used in a classroom environment, are particularly effective for increasing student understanding of the theories of speech production because of the following points:

- 1) Because of the model's transparency, the location of the constriction is visible to the naked eye, as is the overall shape of the cavity. This design helps observers associate the quality of a vowel with the location of constriction on the model.
- 2) Source-filter theory is also easily taught with the models. By feeding several different sound sources into a model, we can teach that pitch is determined by the fundamental frequency of the sound source. At the same time, students are able to see that the quality of a vowel is determined by the shape of the resonator, independent of the sound source. Finally, we are able to show students that harmonic structure is independent from the resonances of the acoustic tube.
- 3) By changing the order of the plates or sliding the strips to simulate constrictions at nodes and antinodes, students are able to hear the effects of formants shifting position. Additionally, we can provide spectral analyses of the output sounds, so students are able to *see* how the frequencies of the formants changed, as well. Being able to hear and see the effects of formant shift helps learners understand how vowels change depending on the location of constriction(s) in the vocal tract.
- 4) Measurements taken from the models are reproducible, so students can go back to an arbitrary measurement and get the same result, which helps them to test their hypotheses as they learn these concepts.
- 5) Using the models along with computer simulation software makes it possible to compare a measured spectrum with one derived from theoretical computation, something useful for advanced students.
- 6) The cylinder-type model is particularly effective when we need to make a quick demonstration of vowel production because the model is "sound-source-ready," that is, there is no set-up time.
- 7) The plate type model is more appropriate for hands-on laboratory experiments because students can change the shape of the model. Although it requires more set-up time to line up the plates, students can work with them as if playing a game.
- 8) By using Umeda and Teranishi's model, time-varying sounds can be produced by sliding the strips to change the configuration of the vocal tract. This enables us to produce diphthongs and transient sounds, which often sound more human like. This model is also valuable when we measure the sensitivity of sound quality to the amount of change in the configuration.

Recently, Arai's models were also used in Acoustics class at the Massachusetts Institute of Technology (M.I.T.). The students showed a lot of interest in seeing and hearing real models of the vowels with an excitation source. They were mainly interested in the different shapes for the vowels, and the capability to simulate a variety of other shapes [23].

### 4. New attempts

As mentioned earlier, teaching Acoustics is also important for younger generations. As an attempt, we helped to develop an exhibition at the Science Museum "Ru-Ku-Ru" in Shizuoka City, Japan. This museum has just opened under the concepts of "watch, listen and touch." Arai's cylinder and plate-type models of the human vocal tract are installed in an exhibition of that museum (originally designed for 10-12 year-old children). For the cylinder-type model, children are able to push the bellows connected to a sound source to produce vowels through the models. The cylinders are mounted on head-shaped plates for different vowels. The plate-type model is set up for a hands-on experiment. Children can line up the plates in the order indicated to make Japanese vowels. The models are well suited to the focus of this museum.

Another attempt that we are developing is an application of our vocal tract model to language learning and speech training. For those who have speech disorders and/or hearing impairment, it is sometimes very useful if we can show them the ideal shape of the vocal tract and/or produce sounds through the model. For those who are learning a new language and have difficulty pronouncing sounds that do not exist



Figure 3: New physical models for consonants /r/ (right) and /l/ (left).

in his/her language, the models might be useful for the same reasons. Figure 3 shows a newly developed models for consonants /r/ and /l/ (designs are based on Stevens, 1998 [22]). Besides this model, we have developed models for stridents, nasalized vowels [6], etc.

## 5. Conclusions

First, we reviewed previously proposed physical models of the human vocal tract as educational tools and their usefulness in the classroom. We also discussed a new attempt to use them at a museum for younger generations. We further identified other areas where the models were needed, for example, as educational tools for speech pathologists and patients with speech disorders and hearing impairments. Furthermore, second language learners appreciate them, as well.

## 6. Acknowledgements

I would like to thank all of the people who helped me in various ways, especially Terri Lander, Janet Slifka, Kenneth N. Stevens, Noriko Umeda, Eri Maeda, Hirokazu Sato, Naoki Ishii, Thomas D. Rossing and Koichi Yoshihisa.

## 7. References

- [1] [http://www.dcs.shef.ac.uk/spandh/ed\\_arena/](http://www.dcs.shef.ac.uk/spandh/ed_arena/)
- [2] <http://www.ling.umu.se/arena/>
- [3] Arai, T., "The replication of Chiba and Kajiyama's mechanical models of the human vocal cavity," *Journal of the Phonetic Society of Japan*, 5(2):31-38, 2001.
- [4] Arai, T., Usuki, N. and Murahara, Y., "Prototype of a vocal-tract model for vowel production designed for education in speech science," *Proc. of Eurospeech*, 4:2791-2794, Aalborg, 2001.
- [5] Arai, T., "An effective method for education in acoustics and speech science: Integrating textbooks, computer simulation and physical models," *Proc. of the Forum Acusticum Sevilla*, 2002.
- [6] Saika, N., Maeda, E., Usuki, N., Arai, T. and Murahara, Y., "Developing mechanical models of the human vocal tract for education in speech science," *Proc. of the Forum Acusticum Sevilla*, 2002.
- [7] Maeda, E., Usuki, N., Arai, T. and Murahara, Y., "The importance of physical models of the human vocal tract for education in acoustics in the digital era," *Proc. of China-Japan Joint Conference on Acoustics*, 163-166, Nanjing, 2002.
- [8] Arai, T., Maeda, E., Saika, N. and Murahara, Y., "Physical models of the human vocal tract as tools for education in acoustics," *Proc. of the First Pan-American/Iberian Meeting on Acoustics*, Cancun, 2002.
- [9] Maeda, E., Arai, T., Saika, N. and Murahara, Y., "Lab experiment using physical models of the human vocal tract for high-school students," *Proc. of the First Pan-American/Iberian Meeting on Acoustics*, Cancun, 2002.
- [10] Arai, T., "Physical and computer-based tools for teaching Phonetics," *Proc. of ICPhS*, 1:305-308, Barcelona, 2003.
- [11] Maeda, S., "A digital simulation method of the vocal-tract system," *Speech Communication*, 1:199-229, 1982.
- [12] Dudley, H. and Tarnoczy, T. H., "The speaking machine of Wolfgang von Kempelen," *J. Acoust. Soc. Am.*, 22:151-166, 1950.
- [13] Flanagan, J. L., *Speech Analysis, Synthesis and Perception*, Springer-Verlag, 1965.
- [14] Gold, B. and Morgan, N., *Speech and Audio Signal Processing*, John Wiley & Sons, 2000.
- [15] <http://www.haskins.yale.edu/haskins/HEADS/SIMULACRA/kratzenstein.html>
- [16] Lindsay, D., "Talking Head," *American Heritage of Invention & Technology*, 13, 1997.
- [17] Chiba, T. and Kajiyama, M., *The Vowel: Its Nature and Structure*, Tokyo-Kaiseikan Pub. Co., Ltd., Tokyo, 1941.
- [18] 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 (in Japanese).
- [19] [http://www.exploratorium.edu/exhibits/vocal\\_vowels/](http://www.exploratorium.edu/exhibits/vocal_vowels/)
- [20] Lander, T. and Arai, T., "Using Arai's vocal tract models for education in Phonetics," *Proc. of ICPhS*, 1:317-320, Barcelona, 2003.
- [21] Fant, G., *Acoustic Theory of Speech Production*, Mouton, The Hague, 1960.
- [22] Stevens, K. N., *Acoustic Phonetics*, The MIT Press, Cambridge, 1998.
- [23] Stevens, K. N., personal communication.