Simple Physical Models of the Vocal Tract for Education in Speech Science

Takayuki Arai

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

Abstract

In the speech-related field, physical models of the vocal tract are effective tools for education in acoustics. Arai's cylinder-type models are based on Chiba and Kajiyama's measurement of vocal-tract shapes. The models quickly and effectively demonstrate vowel production. In this study, we developed physical models with simplified shapes as educational tools to illustrate how vocal-tract shape accounts for differences among vowels. As a result, the five Japanese vowels were produced by tube-connected models, where several uniform tubes with different cross-sectional areas and lengths are connected as Fant's and Arai's three-tube models.

Index Terms: speech science, vocal-tract model, education in acoustics, vowel production, acoustic tube

1. Introduction

1.1. Background

We have proposed several different physical models of the human vocal tract and they are being used for education in speech science. We initially proposed "cylinder-type" and "plate-type" models [1,2] based on the measurement of vocal-tract shapes done by Chiba and Kajiyama (1941) [3]. Subsequently, we developed the lung models and head-shaped models [2,4] with which we explain the mechanism of human speech production. Recently, we proposed a vocal-tract model with a gel-type flexible tongue [5] as well as a "sliding three-

tube model [6]." We have confirmed that these physical models are useful as educational tools when covering basic concepts in speech production, such as, source-filter theory as well as the relationship between vocal-tract shape and vowel quality [3]. The combination of the lung models and the head-shaped models is essential for teaching the mechanism of human speech production from breathing to phonation and articulation, because it enables learners to see the physical configuration of the vocal tract in the context of its location in the body [2,4]. Furthermore, the flexible-tongue model can be installed within the head-shaped model, so that learners can change the location of the tongue and hear the resulting acoustic sounds [5]. Other head-shaped models have a "movable" soft palate. In other words, one-type of model has a rotating valve as the soft palate, so that the coupling between the oral cavity and the nasal cavity is achieved with different degrees [2,4]. The other type of model has a "flexible" soft palate and pharyngeal wall made of a gel-type material, so that we can move the location of the soft palate to achieve different patterns of velopharyngeal closure simulating cleft-palate patients who present hypernasality. One can use our research to select the optimal model to use for education depending on the pedagogical target. (e.g., [2]).

1.2. Two versions of the "cylinder-type" models

The "cylinder-type model", one of our early models, is effective for demonstrating the source-filter theory and the relationship between vocal-tract shape (filter) and vowel

Figure 1: *The earliest version of Arai's "cylinder-type" models. The internal vocal-tract shape is sculpted from a transparent acrylic bar [1,2]. The vowels are /i/, /e/, /a/, /o/, and /u/ from left to right (the glottis is at the bottom, the lip is at the top).*

Figure 2: *The second generation of Arai's "cylinder-type" models. The internal shapes are exactly the same as the earlier models in Fig. 1, but the thickness of the wall of these models is constant (2.5 mm). The models are made of a material called "chemical wood" in this figure. The vowels are /i/, /e/, /a/, /o/, and /u/ from left to right (the glottis end is the bottom, the lip end is the top).*

quality. The round-bottle shape of this model is based on Chiba and Kajiyama's measurement of vocal-tract shape [1], and each independent model corresponds to one of the five Japanese vowels. Once one inputs a sound source from the glottis end, a vowel is produced; this type of model is very useful because a learner can hear the vowel quality while observing vocal-tract shape at the same time. Fig. 1 shows a set of the cylinder-type models made of acrylic resin. Fig. 2 shows a set of the second generation of cylinder-type models, which have exactly the same internal shapes as the earlier models but whose walls maintain a constant thickness (2.5 mm).

Because the thickness is constant and the vocal-tract shapes are reflected to the outer shape, the material does not have to be transparent. Figure 2's models are made of a material called "chemical wood." In these models, we can feel the shape by touching the models, which is important for hands-on demonstrations for everyone, including visually-impaired people.

1.3. The sliding three-tube model

With "the sliding three-tube (S3T) model," we can produce different vowel sounds by sliding an inner cylinder within the outer cylinder. In this model, the inner cylinder forms a constriction within the vocal tract, as shown in Fig. 3. Because of this simple structure, the rough estimation of the first few resonance frequencies can be easily computed [7] by decomposing the whole model into three parts:

a quarter-length resonator with the first resonance frequency of $\dfrac{c}{4\ell_3}$,

a half-length resonator with the first resonance frequency of $\dfrac{c}{2\ell_1}$, and

a Helmholtz resonator with the resonance frequency of

$$\frac{c}{2\pi}\sqrt{\frac{A_2}{A\ell_1\ell_2}}\ .$$

Therefore, this S3T model is not only useful for demonstrating the source-filter theory and the relationship between vocal-tract shape and vowel quality, but also for explaining the basics of the acoustic theory of vowel production for undergraduate and graduate students. Further, the simple structure of this model makes it possible for a science workshop where children make their own vocal-tract models (see details in [8]).

1.4. Goal

In this study, we developed vocal-tract models with simplified vocal-tract shapes to highlight the aspects of vocal-tract shape which account for differences among vowels. In the following section, we introduce the "tube-connected" models.

2. Tube-connected vocal-tract models

2.1. Design

The shapes of the cylinder-type models [1,2] were based on Chiba and Kajiyama [3]. In [3], they measured actual human vocal-tract shapes and simplified the shapes with the first-order approximation. The plate-type models [1,2] were
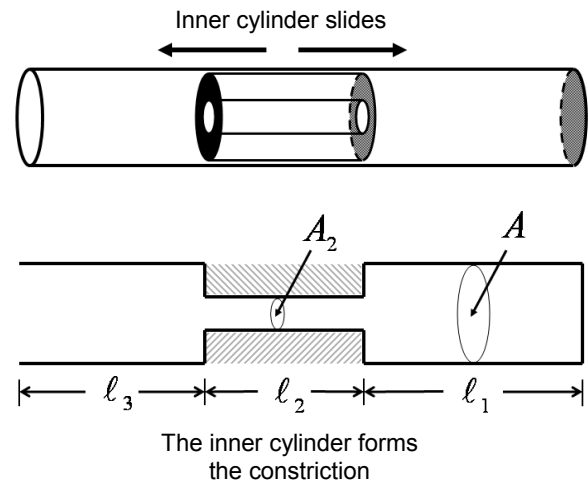

Figure 3: *Schematic figure of the "sliding three-tube model" (adapted from [6]). The inner cylinder (constriction) slides back and forth inside the outer cylinder and the shape of the vocal tract varies, simulating different vowels.*

achieved by their step-wise approximation, but the models can still produce intelligible vowel sounds. The question which arose next was how far we can simplify the shapes while maintaining intelligibility. One of the answers is Fant's three-tube model [9]. In this model, a shape of a vowel can be approximated with three connected uniform tubes with different diameters and lengths. Arai's S3T model [6] is based on this concept. In this study, we developed the tube-connected (TC) models as shown in Fig. 4.

As shown in Fig. 5, this newly proposed TC model was designed by combining the advantages of the cylinder-type and the S3T models. Because we want to use the models as educational tools, we did not want to compromise vowel quality or intelligibility. The simplicity came from the S3T, but we implemented an additional narrowing at the larynx for all vowels and lip rounding for the vowels /u/ and /o/ as in the cylinder-type models.
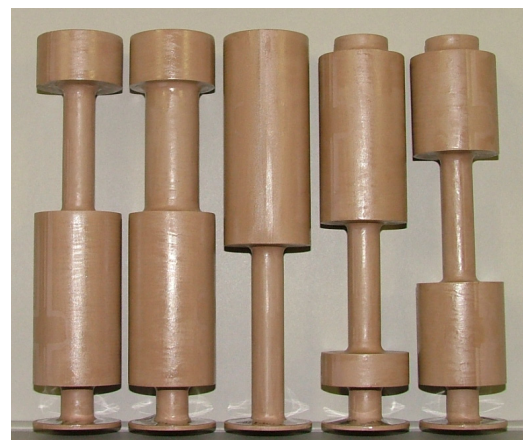

Figure 4: *Tube-connected models. The vowels are /i/, /e/, /a/, /o/, and /u/ from left to right (the glottis end is the bottom, the lip end is the top). The material is "chemical wood."*
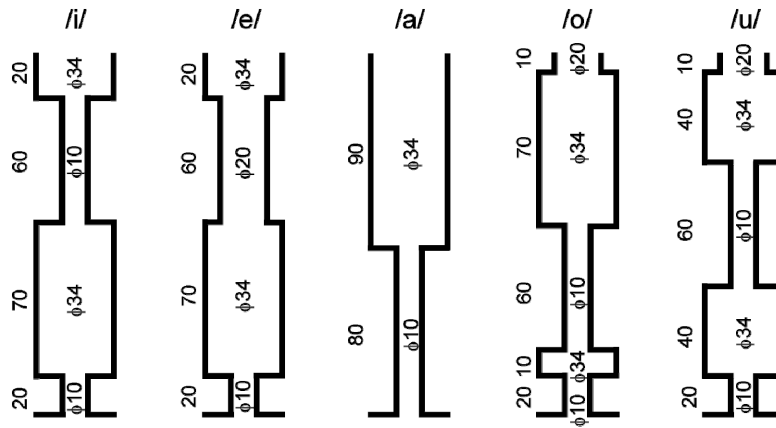
Figure 5: *Dimensions of the tube-connected models. The numbers inside and outside a tube is the hole diameter and the length of that section, respectively. All units are in mm.*

As you can see in Fig. 5, the TC models are based on the S3T model, the outer cylinder of which has a hole diameter of 34 mm. There is a narrow constriction at the larynx with a diameter of 10 mm and a length of 20 mm. For all vowels of the TC models, the length of the inner cylinder for the tongue constriction is 60 mm. As the constriction slides from the larynx end to the lip end, the output vowel changes /a/ > /o/ > /u/ > /i/. In the case of the vowels /u/ and /o/, lip rounding was implemented. The vowel /e/ was achieved when the constriction was positioned as for vowel /i/, but to a wider degree. For pedagogical reasons, the diameter and length of each section of the tubes were selected from a small set of numbers (this also helps to reduce the manufacturing costs).

2.2. Acoustic analysis

Figure 6 shows spectrograms of sounds produced by the TC models. The sounds were digitally recorded using a PCM recorder (SONY, PCM-D1) and its internal microphone; the lip end of the TC models was placed approximately 20 cm from the microphone (48 kHz sampling and 16 bit quantization). An electrolarynx (SECOM, EL-X0010) was used as a sound source. As you see from Fig. 6, the formants for each vowel are clearly visible. The first formants of the high vowels /i/ and /u/ were weak due to frequency characteristics of the electrolarynx. Table 1 shows the first two formant frequencies (F1 and F2) for each vowel estimated by Wavesurfer. This table also shows the F1 and F2 of the cylinder-type models [1] and average F1 and F2 frequencies of the male Japanese speakers [10].

From Table 1, we can confirm that the F1 and F2 frequencies are close to the averages of actual speakers for each vowel. F1 and F2 are better for the vowel /o/. In fact, /o/ was shown to be much more intelligible by an informal listening test. Because we implemented the lip rounding for the vowel /u/, F1 and F2 were lower, and the sound produced came closer to rounded /u/ in English.

3. Discussion

The newly proposed TC models can just as quickly and effectively demonstrate vowel production as the cylinder-type models [1,2]. Especially, they are useful when covering basic concepts in speech production, such as, source-filter theory as well as tangible information about the relationship between vocal-tract shape and vowel quality. At the same time, the TC

models have the simplicity showcased in the S3T models, so that their resonance frequencies are easily estimated by simple approximation. In addition, we can demonstrate the resultant resonance (formant) frequencies well match to the ones estimated from the theory.

The major differences between the TC models and the S3T models are a narrow constriction at the larynx for all vowels
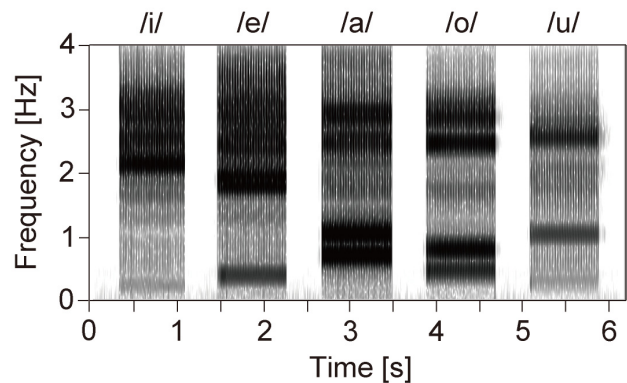

Figure 6: *Spectrogram of sounds produced by the tube-connected models.*

Table 1: *Comparison of F1 and F2 frequencies among the TC models, the cylinder-type models [1], and the averages of native male speakers of Japanese [10].*

	TC models		Cylinder-type models [2]		Avg. of male speakers [10]	
	F1 [Hz]	F2 [Hz]	F1 [Hz]	F2 [Hz]	F1 [Hz]	F2 [Hz]
/i/	239	2157	230	1992	263	2263
/e/	398	1835	405	1868	475	1738
/a/	717	1035	654	1121	775	1163
/o/	481	800	436	1054	550	838
/u/	318	1039	499	1276	363	1300

and lip rounding for vowels /u/ and /o/. We investigated the effects of laryngeal constriction and lip rounding on vowel quality. As a result, the narrow constriction at the larynx improves, in particular, the quality of vowel /i/. As for lip rounding, vowels /u/ and /o/ were investigated with and without lip rounding, and the vowels with lip rounding were more intelligible especially for vowel /o/.

The earliest version of Arai's cylinder-type models was implemented by sculpting the internal vocal-tract shape from an acrylic bar [1,2]. Therefore, the acrylic resin needed to be transparent in order to show the internal shape. In contrast, for the TC models shown in Fig. 4, the thickness is constant and the vocal-tract shapes are reflected in the outer shape, so transparency is not required as the second generation cylinder-type models shown in Fig. 2. Thus, the TC models can be made of non-transparent materials, such as, chemical wood as in the models of Fig. 4. Because the internal shape is reflected to the outer shape, we can feel the shape by touching the models. This is an important aspect useful in hands-on demonstrations, especially for visually-impaired people who can learn about vocal tract shape via tactile impression.

We also compared the output vowel sounds between different models made of different materials. Figure 7 shows the LPC spectral envelopes of vowel /a/ produced by the TC models made of chemical wood, acrylic resin, and wood. From Fig. 7, although wood shows the lowest Q factor, we can see that the differences in frequency are small. This is important, because now when we teach we know that what materials the models are made of is less crucial in terms of vowel quality.

In the present study, we focused on the development of vocal-tract models with simplified vocal-tract shapes as educational tools in order to teach what it is about vocal-tract shape that accounts for the differences among vowels. The TC models are one of the answers to that question. These models are suitable for a variety of audiences, from a science workshop for children through a basic acoustics course at the graduate level. We can compare the output sounds from the vocal-tract models with a precise shape measured by MRI [11], the cylinder-type models, and the TC models by feeding a single sound source to discuss the similarities/differences among three different types of models. Because everyone understands what speech and voice are, even children are interested by the phenomena in front of them without the need to explain any difficult acoustic theory. This is one of the advantages of using physical models as opposed to merely
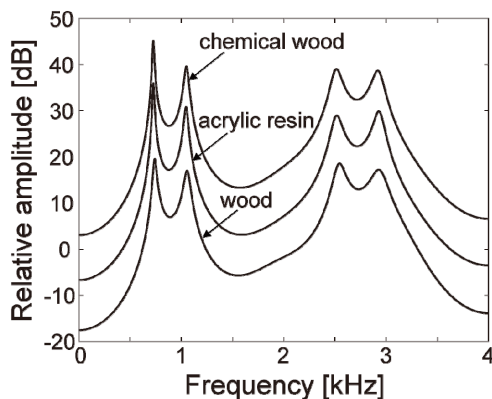
textbooks for education.

It is important to select tools that are consistent with ones educational purpose. Simplification is sometimes helpful in order to focus on the essential, and gloss over the irrelevant. The same principle applies for selecting a vocal-tract model. Which simplification you choose, should be determined by your educational objectives. The S3T and TC models are simple and easy to use to introduce acoustic theory. However, although a straight tube is simple, if we wish to teach the dynamic movement of the tongue, a bent tube would be more appropriate because the tongue movement is even "simpler" in a bent tube than in a model with a straight tube. Thus, which simplification is preferable depends upon your educational objectives.

4. Conclusions

In this study, we proposed tube-connected models with simple vocal-tract shapes. These models can be used as an educational tool for acoustics in speech science just as the cylinder-type models are used. Because of their simple structure, we can teach essential concepts, such as, how the location of tongue constriction and the degree of constriction effect vowel quality. Because which educational tools we use will depend on our educational objectives, we can see that the TC models definitely add an important choice to our palate of educational models for acoustics in speech science. We look forward to continuing our development of new educational tools in the future, to specifically address additional pedagogical targets.

5. Acknowledgements

6. References

[1] Arai, T., "The replication of Chiba and Kajiyama's mechanical models of the human vocal cavity," *J. Phonetic Soc. Jpn.*, 5(2):31-38, 2001.

[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] Chiba, T. and Kajiyama, M., *The Vowel: Its Nature and Structure*, Tokyo-Kaiseikan Pub. Co., Ltd., Tokyo, 1941.

[4] Arai, T., "Lung model and head-shaped model with visible vocal tract as educational tools in acoustics," *Acoust. Sci. & Tech.*, 27(2):111-113, 2006.

[5] Arai, T., "Gel-type tongue for a physical model of the human vocal tract as an educational tool in acoustics of speech production," *Acoust. Sci. & Tech.*, 29(2):188-190, 2008.

[6] Arai, T., "Sliding three-tube model as a simple educational tool for vowel production," *Acoust. Sci. & Tech.*, 27(6):384-388, 2006.

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

[8] Arai, T., "Science workshop with sliding vocal-tract model," *Proc. of Interspeech*, 2827-2830, 2008.

[9] Fant, G., *Theory of Speech Production*, Mouton, The Hague, Netherlands, 1960.

[10] Kasuya, H., Suzuki, H. and Kido, K., "Changes in pitch and first three formant frequencies of five Japanese vowels with age and sex of speakers," *J. Acoust. Soc. Jpn.*, 24(6):355-364, 1968.

[11] Honda, K., Takemoto, H., Kitamura, T., Fujita, S. and Takano, S., "Exploring human speech production mechanisms by MRI," *IEICE Trans. on Information and Systems*, E87-D(5):1050-1058, 2004.

Figure 7: *Spectral envelopes of vowel /a/ produced by the TC models made of chemical wood, acrylic resin, and wood (plotted by shifting 10 dB apart).*