ACOUSTICAL LETTER

Sliding three-tube model as a simple educational tool for vowel production

Takayuki Arai*

Department of Electrical and Electronics Engineering, Sophia University,
7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan

(*Received 8 May 2006, Accepted for publication 7 July 2006*)

Keywords: Education in acoustics, Vocal tract model, Vowel production, Speech science, Acoustic phonetics
PACS number: 43.70.Bk, 43.10.Sv [doi:10.1250/ast.27.384]

1. Introduction

A series of physical models of the human vocal tract for education in acoustics and speech science has been proposed from our group [1–8], and we successfully showed the effectiveness of hands-on activities for an intuitive understanding of the mechanisms and phenomena. Arai [1] replicated Chiba and Kajiyama's physical models of the human vocal tract on the basis of their measurement [9]. This Arai's model [1] consists of the cylinder-type and plate-type vocal-tract models; they are simple but offer a powerful demonstration of vowel production with a sound source such as an electrolarynx or a whistle-type artificial larynx. A driver unit of a horn speaker can also be used as a transducer to produce an arbitrary sound source. One can feed signals to the driver unit not only from an oscillator, but also from a computer using a digital/analog converter and an amplifier, so that any arbitrary signal can be a source signal.

We have recently showed additional physical models of the human vocal tract that are useful for education. We have shown Umeda and Teranishi's model [10] with several sound sources fed through a driver unit in pedagogical situations [11]. In this model, one can change the cross-sectional areas of their model by moving 10-mm (or 15-mm) thick plastic strips, closely inserted from one side. In Arai [8], we further extended our previously proposed physical models of the human vocal tract to the lung models and head-shaped models. The head-shaped models can produce vowel sounds and provide a visual demonstration of how the vocal tract is positioned in the head. The lung models with the whistle-type artificial larynx give a visual demonstration of the human respiratory system and phonation.

In one extended version of the head-shaped model [8], the tongue could be manipulated by hand. Therefore, many different vowels can be produced with the model by changing the position of the tongue. In Umeda and Teranishi's model [10], the shape of the vocal tract can also be changed by moving a set of the thick plastic strips. None of the models, however, had a simple way to change the vocal tract shape in order to produce different vowel sounds. In other words, we needed high degrees of freedom to control the vocal tract shape in the previous models. In this study, we, design a three-tube model with a simple mechanism to produce several different vowels.

2. Sliding three-tube model

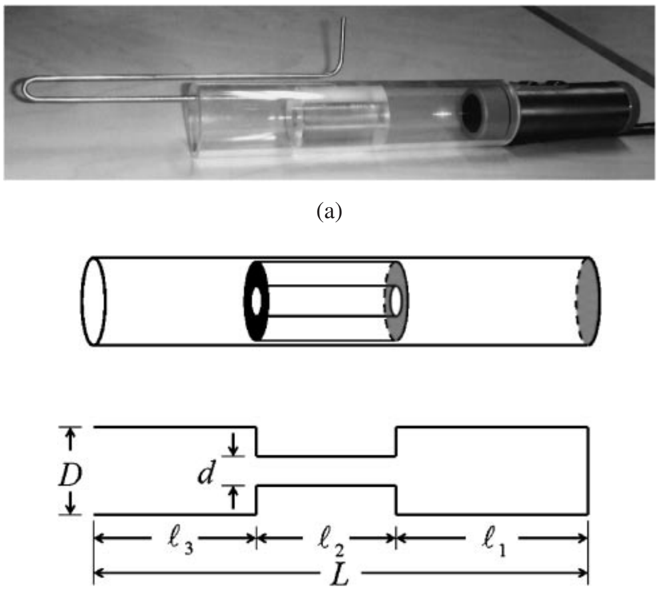
Fant [12] simulated vocal-tract resonances with a three-tube resonator model using an electrical circuit. In this simulation, he drew nomograms as a function of the position of a constriction. We propose "the sliding three-tube (S3T) model" as an implementation of a physical model which varies the constriction position in a three-tube resonator.

Figure 1 shows the S3T model with a schematic view of its mid-saggital cross-section. This model is an idealized system of coupled resonators and can be viewed as a tube having a uniform area function with a single narrow constriction. As you can see from this figure, this S3T model consists of two parts: the outer and inner cylinders. The outer cylinder, whose length is L, is a uniform tube with constant diameter D (the cross-sectional area $A = (\pi/4)D^2$). The inner cylinder has much shorter length (l_2) and its diameter d is less than D (the cross-sectional area $A_2 = (\pi/4)d^2$). The first (back) and third (front) tubes are separated by a narrow tube. Because the inner cylinder slides inside the outer cylinder, the lengths of the back and the front tubes, or l_1 and l_3, vary from 0 to $L - l_2$ under the condition that the overall length is constant, i.e., $l_1 + l_2 + l_3 = L$.

To measure the resonance frequencies of the S3T model, we recorded the output signals from the model. For the recordings, the parameters are set to the values listed in Table 1. Both the inner and the outer cylinders were acrylic resin, and the thickness of the outer cylinder was 3 mm. In this case, a driver unit (TOA TU-750) for a horn speaker was used. An impulse train was fed into the driver unit via the digital-to-analog (D/A) converter of a digital audio amplifier (Onkyo MA-500U); the sampling frequency was 16 kHz. To avoid unwanted coupling between the neck and the area behind the neck of the driver unit and to achieve high impedance at the glottis end, we inserted a close-fitting hard rubber cylindrical filler inside the neck [13]. We made a hole in the center of the rubber filling with an area of 0.07 cm^2. A flange with the diameter of 25 cm was attached at the open end of the tube. The output sounds were recorded using a microphone (Sony ECM-23F5) and a digital recorder (Marantz PMD670) with sampling frequency of 16 kHz. The microphone was placed approximately 20 cm in front of the output end in a sound-attenuated room. The room temperatures were 25.1 and 25.2 degrees centigrade when measured for Settings 1 and 2, respectively.

The dots in Fig. 2 show the measured resonance frequen-

[*] e-mail: arai@sophia.ac.jp



(a)

(b)

Fig. 1 Sliding three-tube (S3T) model with an electrolarynx attached to the closed end.

Table 1 Values of parameters for each setting (in [mm]).

	D	d	L	l_1	l_2	l_3
Setting 1	34	10	175	0–125	50	125-0
Setting 2	34	24	175	0–125	50	125-0

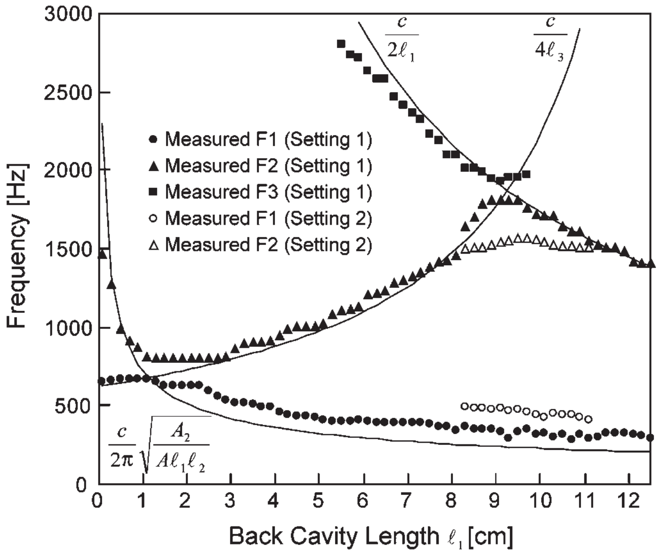


Fig. 2 Measured formants and underlying resonances produced by the S3T model.

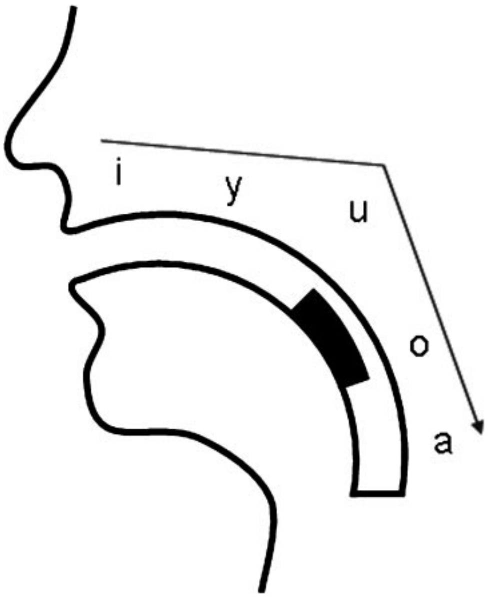


Fig. 3 Tongue constriction and vowels on the two sides of the vowel quadrilateral.

cies (up to 3 kHz) of the S3T model as a function of the back tube length l_1, where l_1 was shifted from 1 to 125 mm in 2 mm step. After downsampling to 8 kHz, the measurement was done by linear prediction on the software, XKL, which is a revision of the software package developed by Klatt [14]. The filled and open dots correspond to the "Setting 1" and "Setting 2" in Table 1, respectively.

In this figure, we also plotted the underlying three resonance curves (solid lines) based on Stevens [15]; they are all monotonic either upwards or downwards as a function of the back tube length. If A_2 is sufficiently small, the impedance of the acoustic mass of the constriction is large compared to the characteristic impedances of the front and back tubes. The line sloping down to the right in Fig. 2 represents the uncoupled resonance of the back tube, i.e.,

$$\frac{c}{2l_1},$$

where c is the velocity of sound. The line sloping up to the right in Fig. 2 represents the uncoupled resonance of the front tube, i.e.,

$$\frac{c}{4l_3}$$

(the end correction of 0.82 r with a flange, where r is the radius of the third tube, was applied at the open end). The back tube and the constriction also form a Helmholtz resonator, i.e.,

$$\frac{c}{2\pi}\sqrt{\frac{A_2}{Al_1l_2}},$$

and the resonance curve is also plotted at the bottom of this figure. In this case, we decoupled and neglected the effect of the front tube, and this might account for the error in Fig. 2 between the theoretical and measured frequencies. The resonance curves are crossing each other, and as a result, when we assign peak numbers from the lower frequency in an overall spectrum as "formants," a different part of the trajectory of a given formant comes from a different resonance curve due to switching the order of the resonance.

3. Discussion

When $l_1 = 9.1$ cm in Setting 1, the resultant vowel was the closest to [i]. This corresponds to when the constriction is located in the front cavity and the second formant is highest, as shown in Fig. 2. These facts are consistent with our knowledge of vowel production. The vowel output was closest to [a] when $l_1 = 0.5$ cm in Setting 1. This corresponds to constriction in the back cavity, when the first and the second formants, F_1 and F_2, are closest together. Again, these facts are consistent with previous knowledge.

The vowel is more like [o] when $l_1 = 2.5$ cm in Setting 1. When the constriction is positioned in the middle of the tube ($l_1 = 7.5$ cm) in Setting 1, the output is more like [u]. In fact, the vowel also sounds like [u] when $l_1 = 12.5$ cm. This observation is reasonable because F_1 and F_2 are lowered when the mouth end is narrowed. When $l_1 = 9.1$ cm in Setting 2 instead of Setting 1, the vowel becomes [e] not [i]. When we move the constriction a little bit further back from $l_1 = 9.1$ cm to 8.3 cm in Setting 1, the vowel changes from [i] to [y].

Thus, as we change l_1 from 9.1 to 0.5 cm in Setting 1, the vowel is changed from [i] → [y] → [u] → [o] → [a]. This movement corresponds to the arrow in Fig. 3. When l_1 moves from 9.1 to 7.5 cm, the constriction moves from the front to the back with almost the same tongue height, where as the constriction moves from high to low when l_1 moves from 7.5 to 0.5 cm. These vowels are located on the two sides of the vowel quadrilateral.

For the F_1 frequency region of vowel [a] and the F_2 frequency region of vowel [i], the two underlying resonance curves intersect with each other. At the intersection, the systems avoid taking exactly the same resonance frequency [15], and as a result, the trajectory of a formant usually has a plateau at such a frequency region. In other words, it yields a stable region in terms of formant frequency, and in that region, the formant frequency is less sensitive to the position of the constriction. This is frequently seen in the relationship between some articulatory parameters and their resulting acoustics as the quantal theory of speech production, and many languages tend to use such stable regions for their

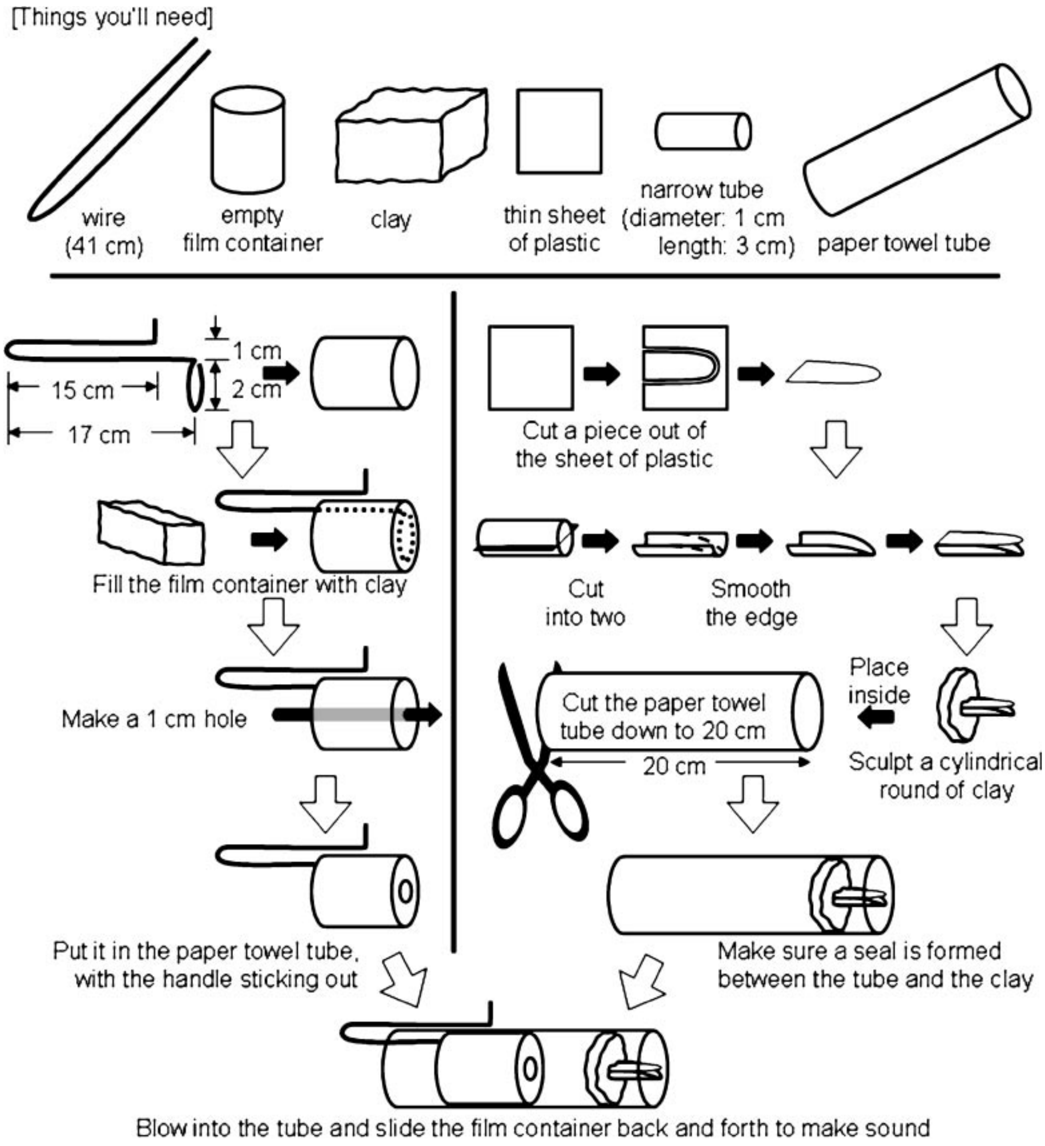


Fig. 4 Illustration of how to handcraft a sliding three-tube model.

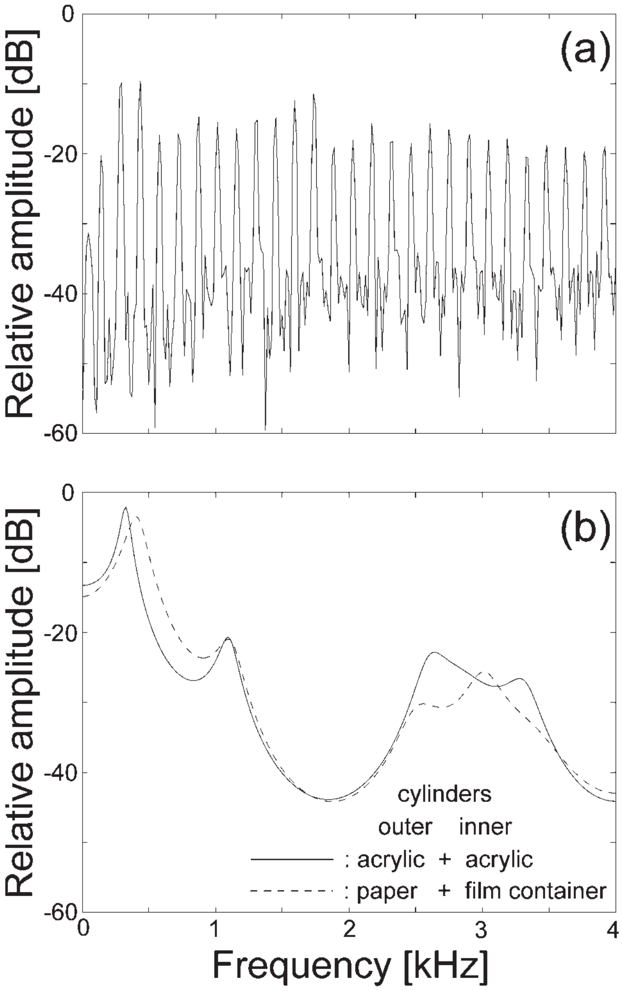


Fig. 5 Spectral characteristics of the handcrafted S3T: (a) an FFT spectrum of the reed-type sound source, and (b) LPC-based spectral envelopes of vowels when the reed-type sound source was fed into the acrylic S3T (solid line) and the handcrafted S3T (dashed line).

vowels [16,17]. Such plateaus are well represented in Fig. 2.

In the measurements, we observed that the back cavity resonance has a much narrower bandwidth. This might account for why the model yielded a perception of a biphonic-singing sound in some positions of the constriction. In [18], it is proved that the high melody pitch in biphonic-singing is caused by the back cavity resonance. This report also corroborates our observation above.

The S3T model is highly suitable for hands-on activities in acoustics education. Not only using the model, but creating such a model could be a part of a pedagogical activity, as well. Figure 4 shows how we can handcraft such a simple S3T model. The sound source could be an electrolarynx, another type of artificial larynx, such as whistle type, or a driver unit of a horn speaker, etc. In the case of Fig. 4, a reed-type sound source was used. Figure 5(a) shows an FFT spectrum of the reed-type sound source. As you can see in this figure, the amplitudes of the harmonics fluctuate within approximately 10 dB. Figure 5(b) shows spectral envelopes of vowels when the reed-type sound source was fed into the acrylic S3T model (solid line) and the handcrafted S3T model with the paper towel tube and the film container (dashed line). In this case, the length of the outer cylinder L was 18 cm and the length of the inner cylinder l_2 was 5 cm. The inner cylinder was located around the middle of the outer cylinder, so that the produced vowel was close to /u/. For making Fig. 5(b), 12th-order linear predictive analysis was applied to the signals after downsampling to 8 kHz. From this figure, we confirmed that the two spectral envelopes are close each other, particularly F_1 and F_2 frequencies. The reason that the F_1 frequency of the handcrafted S3T model is slightly higher can be because the wall of this model is less hard compared to the acrylic model [15]. This difference might also caused the difference in the bandwidths of the formants (the formant bandwidths of the handcrafted model are wider than the acrylic model).

4. Conclusions

In this study, we designed a sliding three-tube model that has a simple mechanism for producing several different vowels. We confirmed that the proposed S3T model can compensate for what we were not able to do with our previous models, as there is now a simple mechanism for producing several different vowels. This model can be used for many activities from science workshops for children to demonstrations of quantal theory for graduate students.

Acknowledgements

This research was partly supported by Grants-in-Aid for Scientific Research (A-2, 16203041 and C-2, 17500603) from the Japan Society for the Promotion of Science, and the Moritani Scholarship Foundation.

References

[1] T. Arai, “The replication of Chiba and Kajiyama’s mechanical models of the human vocal cavity,” *J. Phonet. Soc. Jpn.*, **5**(2), pp. 31–38 (2001).

[2] T. Arai, N. Usuki and Y. Murahara, “Prototype of a vocal-tract model for vowel production designed for education in speech science,” *Proc. 7th Eur. Conf. Speech Commun. Technol.*, Vol. 4, pp. 2791–2794, Aalborg (2001).

[3] T. Arai, “Incorporating more intuitive acoustic education into speech science,” *Proc. Spring Meet. Acoust. Soc. Jpn.*, Vol. 2, pp. 1219–1220 (2002).

[4] T. Arai, “An effective method for education in acoustics and speech science: Integrating textbooks, computer simulation and physical models,” *Proc. Forum Acusticum*, Sevilla (2002).

[5] T. Arai, E. Maeda, N. Saika and Y. Murahara, “Physical models of the human vocal tract as tools for education in acoustics,” *J. Acoust. Soc. Am.*, **112**, 2345 (2002).

[6] T. Arai and E. Maeda, “Acoustics education in speech science using physical models of the human vocal tract,” *Trans. Tech. Comm. Education in Acoustics, Acoust. Soc. Jpn.*, EDU-2003-08, pp. 1–5 (2003).

[7] T. Arai, “Education in Acoustics using physical models of the human vocal tract,” *Proc. Int. Congr. Acoust.*, Vol. 3, pp. 1969–1972 (2004).

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

[9] T. Chiba and M. Kajiyama, *The Vowel: Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo, 1942).

[10] N. Umeda and R. Teranishi, “Phonemic feature and vocal feature: Synthesis of speech sounds, using an acoustic model of

vocal tract," *J. Acoust. Soc. Jpn.* (*J*), **22**, 195–203 (1966).

[11] T. Arai, E. Maeda and N. Umeda, "Education in Acoustics using Umeda and Teranishi's mechanical model of the human vocal tract," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, Vol. 1, pp. 341–342 (2003).

[12] G. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, Netherlands, 1960).

[13] E. Maeda, T. Arai, N. Saika and Y. Murahara, "Studying the sound source of a mechanical vocal tract using a driver unit of horn speaker," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 417–418 (2003).

[14] D. H. Klatt, "The new MIT speech VAX computer facility," *Speech Communication Group Working Papers IV*, Research Laboratory of Electronics, MIT, Cambridge, pp. 73–82 (1984).

[15] K. N. Stevens, *Acoustic Phonetics* (MIT Press, Cambridge, Mass., 1998).

[16] K. N. Stevens, "The quantal nature of speech: Evidence from articulatory-acoustic data," in *Human Communication: A Unified View*, P. B. Denes and E. E. David Jr., Eds. (McGraw Hill, New York, 1972), pp. 51–66.

[17] K. N. Stevens, "On the quantal nature of speech," *J. Phonet.*, **17**, 3–46 (1989).

[18] S. Adachi and M. Yamada, "An acoustical study of sound production in biphonic singing, Xöömij," *J. Acoust. Soc. Am.*, **105**, 2920–2932 (1999).