Acoust. Sci. & Tech. **28**, 3 (2007)

PAPER

Education system in acoustics of speech production using physical models of the human vocal tract

Takayuki Arai*

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

(*Received 30 May 2006, Accepted for publication 16 February 2007*)

Abstract: In this paper, we present and discuss an educational system in the fields of acoustics and speech science using a series of physical models of the human vocal tract. Because education in acoustics is relevant for several fields related to speech communication, it hosts students from a variety of educational backgrounds. Moreover, we believe that an education in acoustics is important for students of different ages: college, high school, middle school, and even elementary school students. Because of the varied student populations, we develop an educational system that instructs students intuitively and effectively and consists of the following models: lung models, an artificial larynx, Arai's models (cylinder and plate type models), Umeda and Teranishi's model (a variable-shape model), and head-shaped models. These models effectively demonstrate several principal aspects of speech production, such as phonation, source-filter theory, the relationship between vocal-tract shape/ tongue movement and vowel quality, and nasalization of vowels. We have confirmed that combining the models in an effective way produces complete education in the acoustics of speech production. The examinations and questionnaire surveys conducted before and after using our proposed system revealed that the learners' understanding of what improves with the use of the system. The system is also effective for voice and articulatory training in speech pathology and language learning.

Keywords: Education in acoustics, Vocal tract model, Vowel production, Speech science, Acoustic phonetics

PACS number: 43.10.Sv, 43.70.Bk [doi:10.1250/ast.28.190]

1. INTRODUCTION

Speech communication and related fields intersect with acoustics in crucial ways. The series of events involved in speech communication is called the "Speech Chain" [1], and it is a basic concept in speech and hearing sciences. Speech and hearing sciences and acoustics are related to other fields as well [2], such as, speech pathology, phonetics and phonology, psychoacoustics, and speech technology. Because acoustics is related to so many fields in speech communication, the backgrounds of students are extraordinarily diversified. The author is teaching acoustics at Sophia University not only to technical students but also to students majoring in Linguistics, Psychology, and Speech Pathology. Furthermore, we believe that acoustics education is important for college, high school, middle school, and even elementary students, which makes age another significant source of diversity in acoustics student populations [3]. We are motivated to develop an intuitive and effective system for educating all these students [2–14].

These days, a variety of tools for teaching acoustics is widely available including textbooks, recordings, databases, physical tools, and computer-based tools [10,11]. Moreover, people are now paying more attention to education in acoustics and speech communication (e.g., [15] and [16]). However, physical models designed for educational purposes in speech science can hardly be found in the literature. Because we feel that physical models of the human vocal tract should be widely used for education, especially in speech sciences, we have been developing an educational system since I designed early models of the human vocal tract [2–14].

In this paper, we propose a total system of education in the acoustics of speech production using the physical models of the human vocal tract summarized in Fig. 1. Each component of this total system exemplifies something both unique and instructive so that if used together they are powerful. In Section 2, we discuss a series of physical models of the proposed educational system including

---

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

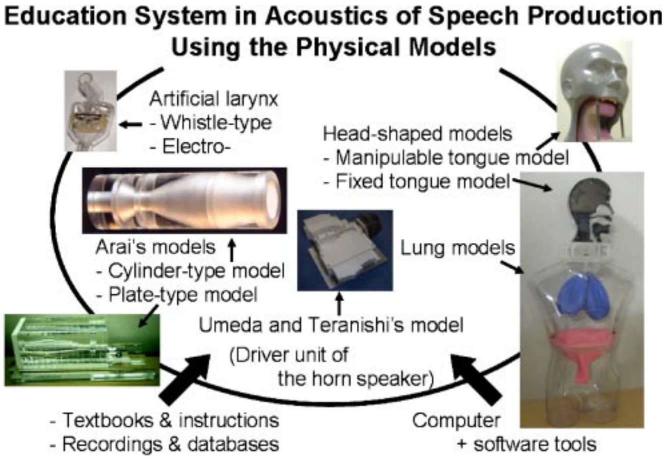

Fig. 1 Diagram of education system in acoustics of speech production using physical models of human vocal tract.

several new ones, with explanations of how we implemented them effectively for educational purposes. Then, in Section 3, we propose an integration of those models with related theories that organize them systematically, enabling students to grasp concepts in acoustics in speech science more intuitively. Finally, we summarize our proposed education system and discuss some of our evaluation results.

2. PHYSICAL MODELS AS EDUCATIONAL TOOLS IN EDUCATIONAL SYSTEM FOR SPEECH PRODUCTION

2.1. History of Vocal Tract Models

Creating mechanical models of the human vocal tract is not new. In the 18th century, Kratzenstein (1781) and von Kempelen (1791) proposed mechanical models for vowels and consonants [17–20]. Kempelen's machine, operated with the right forearm, consisted of bellows simulating the lungs. An oscillating reed stimulated by the airflow from the bellows, excited a single, hand varied resonator for producing voiced sounds. Consonants, including nasals, were simulated by four separate constricted passages controlled by the fingers of the other hand. Willis (1830) claimed to have synthesized all the human vowel sounds as a continuum using a uniform diameter tube with varying length. In the early 19th century, Faber demonstrated his "talking head" [21]. In the late 19th century, Bell and Wheatstone duplicated von Kempelen's machine [21]. More recently, Riesz have designed a mechanical talker [19]. Paget (1923) successfully simulated artificial vocal folds, and also discovered the closest approximation of the shape of an actual vocal cavity by conducting a comparative study of sounds emitted by model cavities of various shapes that he constructed [22].

Chiba and Kajiyama (1942) made mechanical models of the human vocal tract based on their cross-sectional measurements, compared mechanically produced sounds with the sounds of natural vowels, and confirmed that vowel quality is governed by the overall shape of the vocal tract [22]. After Chiba and Kajiyama published their book in 1942, many researchers designed models for their own research purposes. For example, Umeda and Teranishi (1966) developed a mechanical model for investigating vowel and voice qualities [23]. Dang and Honda (1995) designed a mechanical model for measuring the effect of pyriform fossa side branches at the larynx [24]. Recently, Honda *et al.* designed precise models of a human vocal tract using magnetic resonance images [25].

Unfortunately, not many studies use mechanical speech synthesizers for education purposes in speech science, however, some museums have physical models of the human vocal tract. The final version of von Kempelen's machine is kept at the Deutches Museum in Munich [26]. There is a set of vocal tract models at the Exploratorium in San Francisco under the supervision of John Ohala [27,28].

Although some professors use physical models of the human vocal tract in their classes [28], such use is hardly discussed in the literature. Recently, Arai (2001) replicated Chiba and Kajiyama's mechanical models of the human vocal tract from transparent materials and showed that the models are useful as educational tools [4–6]. Since then, we have continued to develop educational tools [2–14]. The following sections describe the original and new models comprising our education system.

2.2. Sound Sources

2.2.1. Artificial larynx as sound source

A sound source is needed to excite any vocal tract. An artificial larynx [29], which is used to simulate the glottal source in patients with laryngectomy, is one of the best sound sources because it produces a reasonably clear vowel sound with our models. We found the electrolarynx to be useful for education. Alternatively, a whistle-type artificial larynx is available, including a reed type [29–31]. The first description of an artificial larynx was given by Czermak in 1859 [29]. Since then, different artificial larynxes were developed, such as Tapia's artificial larynx (the so-called "duck call" also has a similar mechanism). Figure 2(a) shows the whistle-type artificial larynx by Hankou-kai and used in Arai's models described in [5]. With this artificial larynx, one end of the tube is attached to the patient's trachea cannula, while pulmonic air flows through the tube. At the other end, a small plastic box with a hole covered with a rubber membrane is vibrated in the presence of a certain pressure drop in front of and behind the membrane.

Both of the above-mentioned artificial larynxes are useful for our purposes in the proposed educational system. First, they produce reasonable glottal source sounds, which

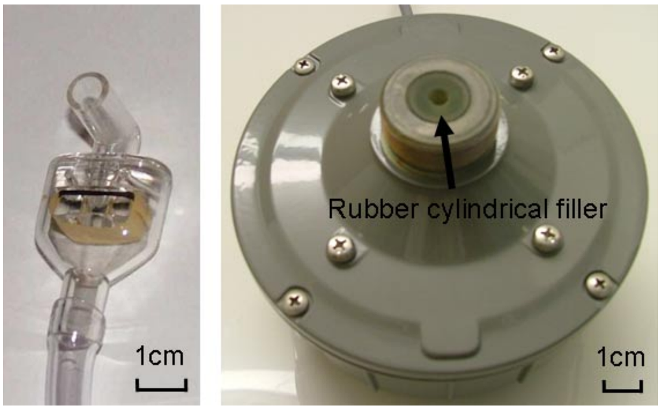

Fig. 2 Sound sources: (a) whistle-type artificial larynx, and (b) driver units of horn speaker with hard-rubber cylindrical filling (note small hole at center).

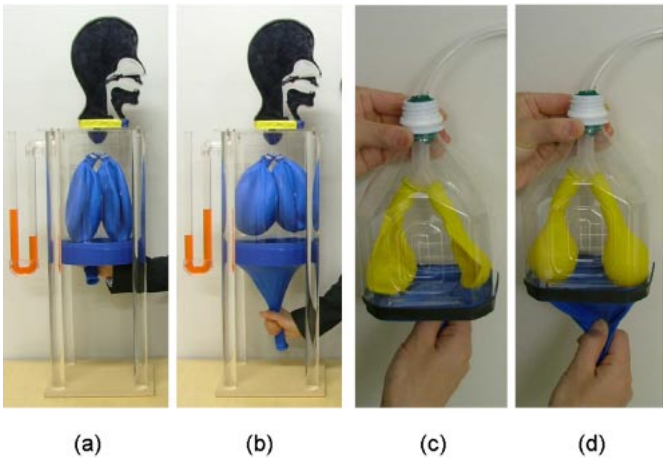

Fig. 3 Large (a, b) and small (c, d) lung models. (a), (c) exhalation; and (b), (d) inhalation. The head-shaped models are set on the top of the large lung models.

is important because the quality of vowel production demonstration depends on the quality of the sound source. Second, they are simple and portable, and are therefore suitable for classroom demonstrations. Third, because they are originally designed for laryngeal prosthesis, they are effective tools for speech science. With the electrolarynx, students can observe that the original buzz sound changes immediately to a humanlike vowel when the device is gently pressed on the outside of the neck, just opposite the larynx. Furthermore, artificial larynxes including electrolarynx are specially designed to prevent sound from leaking from the body parts surrounding the device, so that the source sound does not directly mix with the output sound.

2.2.2. Lung models

The whistle-type artificial larynx requires airflow to produce sounds. For airflow, one may either blow into the larynx or use bellows. Bellows may be preferable, in order to avoid the incorrect impression that a person is phonating a vowel rather than simply blowing into the device. Bellows are useful because they are not affected by or confused with the human anatomy.

To imitate the human respiratory system with a simple device, we adopted a functional model of the lung and diaphragm in our education system (Fig. 3). With this model, students can slowly pull on a knob attached to the "diaphragm" (a rubber membrane covering the bottom of the cavity) to inflate the "lungs" (two balloons). The balloons are connected to a Y-shaped tube simulating the trachea. Pulling down the diaphragm increases the volume of the thoracic cavity, thereby creating a negative pressure in the air inside the thoracic cavity. Air flows into the lungs to equalize the pressure inside the lungs with the atmospheric pressure, simulating inhalation. Pushing up on the diaphragm decreases the thoracic cavity volume, causing air to flow out of the lungs, simulating exhalation.

We designed large lung models for longer phonation, as in Figs. 3(a) and (b). The volume of these lungs is slightly smaller than that of the lungs of the average adult male. Our lung model yields a maximum phonation time of about 15 seconds. The small-lung model shown in Figs. 3(c) and (d) is also effective for education, and is easy to make. One only needs a plastic bottle for the thoracic cavity and a couple of balloons for the lungs and diaphragm. Small children can make these lung models as a hands-on activity. Another type of lung model is the human-shaped model shown in Fig. 1. In this case, a plastic torso is used.

2.2.3. Other sound sources

A driver unit of a horn speaker can be used as a transducer for producing an arbitrary sound source. One can feed signals to the driver unit not only from an oscillator, but also from a computer by means of a digital/analog converter and amplifier, so that any arbitrary signal can be a source signal.

One drawback with the driver unit of a horn speaker is that it is difficult to couple the unit to a vocal tract model. If attachment is not performed carefully, unexpected resonances appear in the output signal due to unwanted coupling between the neck and the area behind the neck. To avoid unwanted coupling and achieve high impedance at the glottis end, Umeda and Teranishi [23] filled the neck of their model with nails. The total area of the resulting opening was about 0.3 cm^2. For our education system, we inserted a close fitting, hard, rubber cylindrical filler inside the neck of our model [32], and made a hole in the center of the rubber filling with an area of 0.3 cm^2. Figure 2(b) shows the driver unit with the filling. The driver unit was the model TU-750 used for a horn speaker by TOA.

The sound source for a driver unit could virtually be any signal. We typically use the following: an impulse train, triangular glottal pulses, white noise, an LPC residual

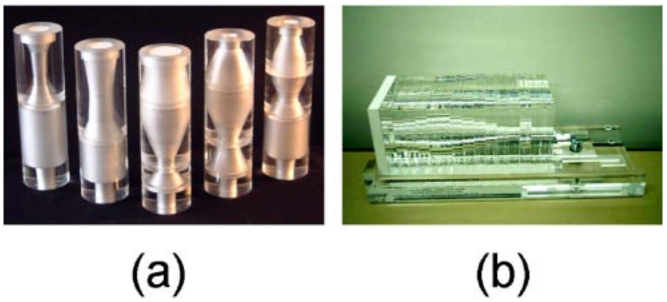

Fig. 4 Arai's models of human vocal tract: (a) cylinder-type models (from left, /i/, /e/, /a/, /o/ and /u/); and (b) plate-type model.

Table 1 Diameters (in mm) for the five Japanese vowels used in plate-type model.

Vowel	Index no. from lips															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
/i/	24	14	12	10	10	10	16	24	32	32	32	32	32	32	12	12
/e/	24	22	22	20	18	16	16	18	24	28	30	30	30	30	12	12
/a/	32	28	30	34	38	38	34	30	26	20	14	12	16	26	12	12
/o/	14	22	26	32	38	38	34	28	22	16	14	16	22	30	12	12
/u/	16	14	20	22	24	26	22	14	18	26	30	30	30	30	12	12

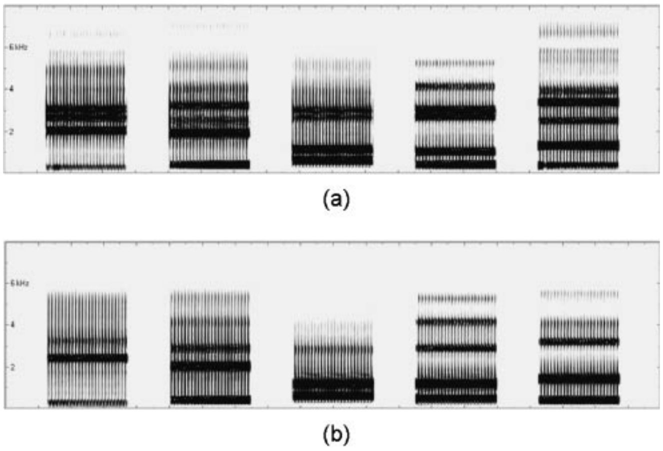

Fig. 5 Spectrograms of produced sounds of five Japanese vowels (from left, /i/, /e/, /a/, /o/, and /u/) for each type of model; (a) cylinder-type model and (b) plate-type model. An electrolarynx was used as a sound source in each case.

signal and an acoustic signal recorded just outside a speaker's larynx during phonation. Another possibility is to use a model for the glottal waveform, such as, the KLGLOTT88 voicing source model [33].

2.3. Arai's Models

Arai proposed two types of models: the cylinder-type model (Fig. 4(a)) with the precise reproduction of the original vocal-tract shapes of Chiba and Kajiyama [22], and the plate-type model, (Fig. 4(b)) a stepwise approximation of these original shapes. Arai's models are important components of our proposed education system because they help learners understand the relationship between vocal tract shape and output. In both models, a vowel-like sound is produced at the lip end when a sound source excites the glottis end.

2.3.1. Cylinder-type model

In Chiba and Kajiyama (1942) [22], polygonal lines show the radius for five Japanese vowels based on their measurements. These lines are the first-order approximation of the original measurements. This type of approximation may lead to some loss of speaker-specific spectral information while still maintaining vowel quality. This simplification is preferable for our education system because it produces a better sketch of vowel shapes for learners, and emphasizes speaker-independent aspects of vowel production. Therefore, we used the same approximation, that is, the polygonal lines of the radius along the length of the vocal tract. Cylinder-type models were produced by rotating the radius curves around a pivot.

For the cylinder-type model, we made an acrylic cylinder with a diameter of 50 mm. Resin was sculpted from the center of the cylinder to make a round bottle-shaped cavity. Prior to sculpting the cavity, we cut each cylinder into segments to reach less accessible areas with the sculpting tool and then glued the cylinder back together.

2.3.2. Plate-type model

An easier way of designing a similar vocal tract model is to use the zeroth-order approximation of radius curves. We approximate the radius curves at a 10 mm resolution in a stepwise manner. The plate-type model consists of a set of acrylic plates, each with a hole at the center. When placed side by side the holes in the plates form a tube, the cross-sectional area of which changes in a stepwise manner. Each plate is 75 mm × 75 mm × 10 mm. Table 1 shows the diameters of the holes varying based on the zeroth-order approximation of the radius curves at a 10 mm resolution.

2.3.3. Acoustic analysis of Arai's models

Figure 5 shows spectrograms of the output signals of the five vowels for Arai's two model types. An electrolarynx was connected to the glottis end of each model, and output signals were recorded and analyzed. This figure confirms that there is little difference between the outputs of the two models, especially at the first and second formant frequencies.

Maeda *et al.* (2004) investigated the differences between the cylinder and plate-type models [34,35]; they found that the formant frequencies in the lower frequency range, that is, those governing vowel quality, were almost identical in both cases, although the plate-type model had

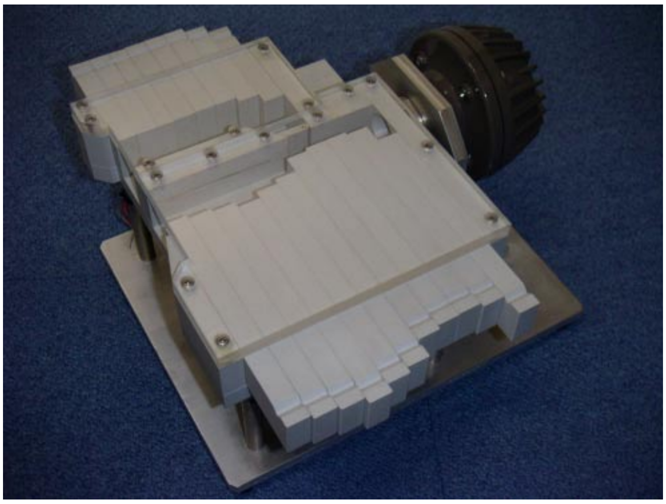

Fig. 6 Umeda and Teranishi's model of human vocal tract. In this figure, each plastic strip is 15 mm wide; 11 strips are used to form an oral cavity, the total length of which is 165 mm. This model has a nasal cavity (the top part). The velopharyngeal coupling is controlled by rotating a valve.

broader formant bandwidths. Because of their similarities, the outputs produced from the plate and cylinder-type models have nearly the same vowel qualities. The differences seem to result from a loss of acoustic energy between adjoining plates. At higher frequencies, the relative amplitude of the plate-type model is also different, but this difference is negligible for our educational purposes.

2.4. Umeda and Teranishi's Model

Umeda and Teranishi (1966) developed a simple device that acoustically simulates the human vocal tract [23]. 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, as shown in Fig. 6. The model also has a nasal branch.

Various vowels and other sustained sounds may be produced by configuring their model differently. Glottal sounds are sent into the glottis end of the model and emitted from the mouth end. The driver unit of a horn speaker is used as a sound source. Using the model, Umeda and Teranishi investigated phonemic and vocal features of speech [23].

We obtained a replica of their model and incorporated it into our education system as an educational tool (Fig. 6) [36]. For sustained vowels, the strips were properly positioned prior to making sounds. For transient sounds, the strips were moved while the sound source is fed. Although it was difficult to make perfect diphthongs by sliding the strips by hand, we were able to produce several degrees of time-varying sounds.

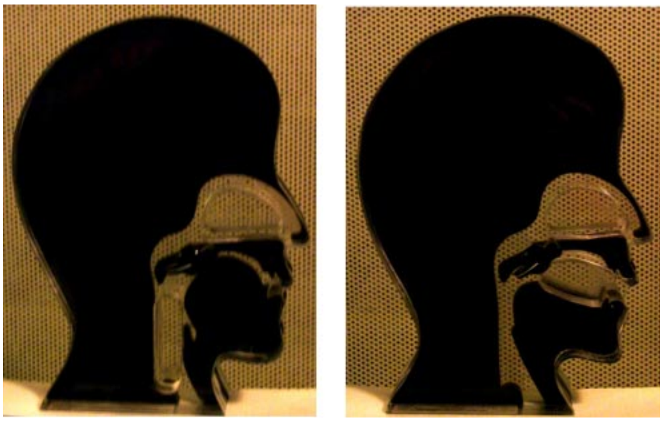

Fig. 7 Vocal tract and simplified nasal passage as part of head-shaped models: /i/ (left) and /a/ (right).

2.5. Vocal Tract and Simplified Nasal Passage in the Head-Shaped Model

None of the models that have been discussed so far seemed helpful for students to visualize how the vocal tract is positioned in the head. In response to this need, we designed another model in our education system (Fig. 7). This figure shows head-shaped models for /i/ and /a/. Each model is constructed from five acrylic plates.

For the education system, we have two goals for the head-shaped models: 1) to make the midsaggital cross-section visible from the outside, and 2) to produce actual vowel sounds from the models. We achieved the first goal with a 1-cm-thick center (black) plate with a schematic midsaggital cross-section for each vowel. All of the other plates are made transparent so that the midsaggital cross-section is visible from the outside. For the second goal, we need to have a proper area function for each vowel to form a three-dimensional model. To achieve this, two transparent 2-cm-thick plates on both sides of the center plate are necessary. These plates should have holes for achieving proper area functions for the vocal-tract configurations of the vowels /i/ and /a/, with nasal cavities.

The velum can be rotated around a pivot located at the boundary of the soft and hard palates. This movable velum acts as the velopharyngeal port and allows the simulation of nasal coupling. The velopharyngeal opening is controlled by the rotating valve. Figure 8 shows spectrographic representations of the output signals from the head-shaped models for the vowels /i/ and /a/ when fed with a glottal waveform by the KLGLOTT88 voicing source model [33] ($F_0 = 100\,\text{Hz}$). In both cases, the velopharyngeal port was gradually opened in the middle of the utterance. Once the velopharyngeal port is opened, several signs of nasalization become evident: weakened F_1 amplitude, widened F_1 bandwidth, shifted F_1 frequency, and nasal poles and zeros (note the nasal pole especially around 1,000 Hz).

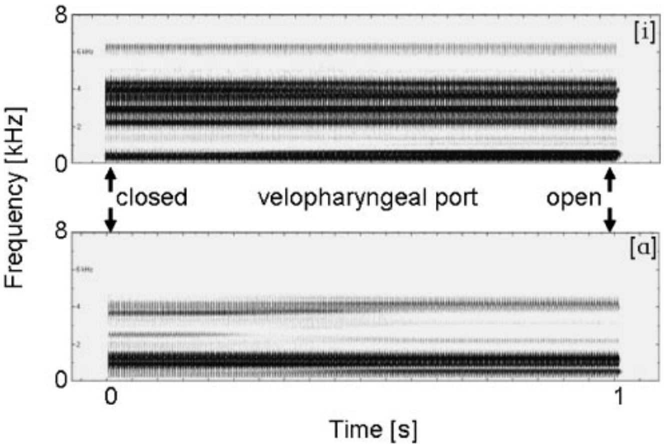

Fig. 8 Spectrographic representations of output signals from head-shaped models for vowels /i/ and /a/. In both cases, the velopharyngeal port was gradually opened in the middle of the utterance.

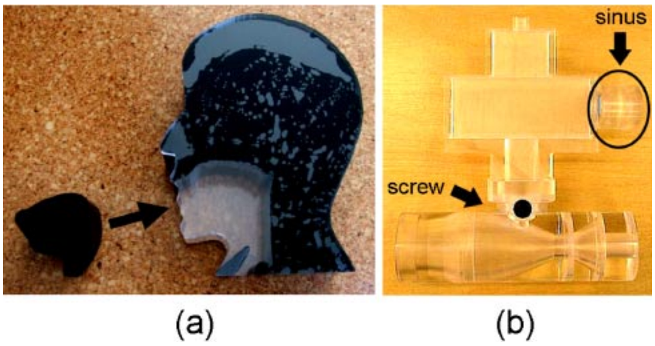

Fig. 9 (a) Head-shaped model with manipulable tongue. (b) Model for nasalized vowel of /a/ (dimension for nasal cavity was based on Chen [37]).

2.6. Other Vocal-Tract Models

We also developed several advanced models in our educational system. Fig. 9(a), as well as Fig. 1, shows an extended version of the head-shaped model described in the previous section, the main difference being tongue manipulability. In the first "tongue fixed" head-shaped model, the vocal tract shape was fixed, so one had to use different models for different vowels. For the newly proposed "manipulable tongue model," a learner can change the vocal tract shape by manipulating the tongue in a single model, and thereby better understand the dynamics of speech production [13,14]. The tongues in the models described in [13,14], are made of acrylic resin, a bent rubber tube, and felt. In each case, the tongue position may be changed by sliding the tongue between the two acrylic plates in the rest of the head-shaped model. Figure 9(a) shows the model with the tongue made of acrylic resin. In this case, although the tongue shape is fixed, we can slide the acrylic tongue by rotating and shifting the location so that the vocal-tract shape is changed. In the case of the bent rubber tube or the felt, the tongue itself is flexible. Another model with a flexible tongue is shown in Fig. 1; the tongue can be made of different materials such as rubber, silicone, urethane, and gel. Again, the tongue position may be changed by sliding the tongue between the acrylic plates.

Fig. 10 Models for consonants /r/ (left) and /l/ (right). The designs are based on Stevens [39].

The model in Fig. 9(b) simulates the nasalized vowel /a/ and is operated by attaching a side branch to the acoustic tube, simulating the nasal cavity [37]. The dimension of the nasal cavity was based on Chen [38]. The velopharyngeal opening can be adjusted using two screws. A resonator for a sinus is available, so that we can use the model either with or without the sinus.

We are currently developing new models, such as models for sonorants /r/ and /l/, and models for other consonants including stridents. Although dynamic models are ideal for /r/ and /l/, we first developed the static models of /r/ and /l/ shown in Fig. 10. The designs of these models are based on Stevens [39]. The models are made of transparent (exterior) and black (interior) acrylic plates, so that the internal shape of the oral cavity can be seen. Figure 11 shows the spectrograms of the sounds produced by these models. From these spectrograms, one can observe that the third formant of the sound from the /r/ model is very low (less then 2 kHz), whereas that from the /l/ model is approximately 2.5 kHz.

3. HOW THE EDUCATIONAL SYSTEM WORKS

In this section, we describe how we should combine the theory behind the models with each of the proposed models in our educational system. We also discuss how to instruct learners in particular topics on the basis of our experience with students in the classroom.

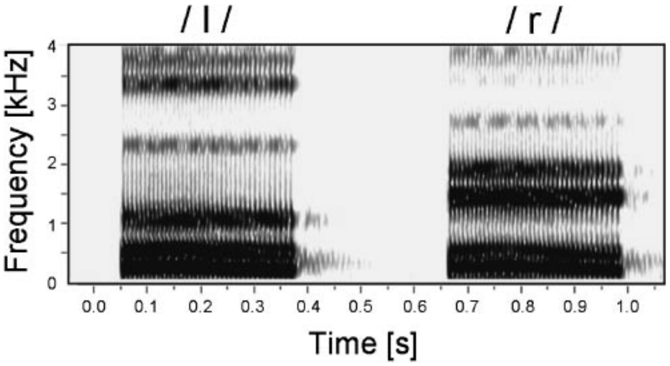

Fig. 11 Spectrograms of sounds produced by models for /r/ (right) and /l/ (left).

3.1. Phonation

By combining the lung model with the whistle-type artificial larynx we can effectively demonstrate human phonation. The subglottal pressure is typically within 6–10 cm H_2O in modal phonation. To visualize a pressure range appropriate for the proper functioning of the artificial larynx, we prepared a U-shaped tube in our education system. By filling this tube with water and mounting it to the thoracic cavity, as shown on the left hand side of Figs. 3(a) and (b), we can observe the vertical movement of the surface of the water corresponding to inhalation and exhalation as a result of the equalizing pressures inside and outside of balloons. Just as the subglottal pressure must be within a certain range for the vocal folds to vibrate, the pressure under the artificial larynx must be within a certain range for the rubber membrane of the whistle-type artificial larynx to vibrate. If the pressure is too low, the membrane does not vibrate. If the pressure is too high, the membrane stops vibrating. If by measuring the height of the water column, we can determine the pressure in cm H_2O. This height for the artificial larynx to vibrate stably is within the range of 6–10 cm. During inhalation, a water column is created on the other side of the U-shaped tube, reflecting negative pressure in the thoracic cavity.

3.2. Source-Filter Theory

Vowel production can be approximated by a source-filter concept [39,40]. The source-filter theory is easily taught with the vocal-tract models. This theory states that a vowel can be treated as an output from a linear time-invariant system (the vocal tract) with a sound source (glottal source). To demonstrate this theory, we begin by producing sounds from a source. The source sound must be a buzzlike sound, such as that emitted at the actual glottis, with no vowel quality, but which produces our sensation of pitch. Once we input the sound source into a vocal-tract model for a vowel, we can show students that the sound now has vowel quality, but the pitch has not changed. By feeding several different sound sources into a model, we can demonstrate that pitch and voice quality are mainly determined by the sound source (vocal fold vibration) not by the shape or type of model (vocal tract). At the same time, students are able to see that the quality of a vowel is mainly determined by the shape of the resonator (vocal tract) independent of the sound source (vocal fold vibration). Finally, we are able to show students that the harmonic structure is independent of the resonance of the acoustic tube.

When we combine a physical model of /i/ and a whistle-type artificial larynx, we often observe an interaction of the source and filter due to aerodynamics. In other words, if the constriction of /i/ is too small, it causes some degree of pressure drop before and after the constriction. As a result, because transglottal pressure decreases, the vibration of the rubber membrane of the artificial larynx becomes slow. This is a real phenomenon that can happen at the larynx, and it may be described in advanced classes of speech production.

3.3. Relationship between Vocal-Tract Shape and Vowel Quality

Because the models in our education system are transparent, the location of constriction is visible to the naked eye, as is the overall shape of the cavity. This consistent transparent design for all the vocal-tract models in our education system helps observers intuitively associate the quality of a vowel with the location of the constriction in the model. Both the cylinder-type model and the head-shaped models are particularly effective for quickly demonstrating vowel production because they are "sound-source-ready," that is, there is no set-up time. By producing sounds from such models, we can demonstrate the relationship between vocal-tract shape and vowel quality.

3.4. Changes in Vocal-Tract Shape and Dynamic Speech Production

The plate-type model, Umeda and Teranishi's model, and the manipulable tongue model in our education system are suitable for demonstrating changes in vocal-tract shape. Although the cylindrical models also demonstrate certain different shapes of the vocal tract, they cannot be changed, where as the former models can be manipulated if one wishes to investigate how changes in vocal tract shape affect acoustic output. Measurements obtained using the plate-type model are reproducible, so students can go back to an arbitrary measurement and get the same result, helping them test their hypotheses as they learn the concepts. The plate-type model is more appropriate for hands-on laboratory experiments because students can change the shape of the model piece-by-piece. Although it requires a longer set-up time to line up the plates, students can work with the plates as if playing a game.

By using the manipulative tongue model as well as Umeda and Teranishi's model, time-varying sounds, such as diphthongs, can be produced by sliding the strips or the tongue to change the configuration of the vocal tract. Sliding the strips or tongue also enables us to produce transient sounds. Diphthongs and transient sounds often seem more humanlike than the steady vowel sounds produced by the other models. Another benefit of these models is that students are able to see and hear the change from one sound to another in real time. This model is also valuable when measuring how sensitive sound quality is to changes in configuration. From our teaching experience, we confirmed that these models are particularly useful when teaching the dynamic aspects of speech production. A learner can manipulate the tongue by hand, so that he/she can intuitively learn how we produce speech sounds. The combination of the simultaneous sensations of tactility, somatosensory and auditory perception helps learners understand the phenomenon more naturally and easily.

3.5. Nasalized Vowels

Umeda and Teranishi's model, the head-shaped models, and a model based on Chen [38] in our education system are effective tools for demonstrating the nasalization of vowels. There are certain features in the models that help students grasp these concepts. In particular, in the first two models, the velopharyngeal opening is adjustable with a rotating valve. For the Chen-based model, the velopharyngeal opening can be adjusted using two screws. The Chen-based model also includes an optional resonator representing a sinus.

It has been reported that the velum is higher in high vowels and lower in low vowels [41]. To achieve the same degree of perceived nasality, we need more velopharyngeal opening for low vowels, and less opening for high vowels [42]. We control different degrees of velopharyngeal opening for the vowels /i/ and /a/, which is easily demonstrated with the models in our education system.

4. DISCUSSION

4.1. Comprehensive Acoustics Education from Lungs to Head

In this section, we discuss our pedagogical system using several models for education in speech production: 1) lung models, 2) artificial larynxes, 3) cylinder-type (straight-tube) models, 4) variable-shape models (Umeda and Teranishi's model, plate-type model, and manipulable tongue model), and 5) head-shaped models. As discussed in Section 3, each model has its own advantages, and if combined effectively, those advantages can be underscored, producing a systematic and comprehensive education in acoustics from the lungs to the head.

The following are some of the strengths of the comprehensive education proposed using our system, which is facilitated at every step by the hands-on models. For example, an electrolarynx itself is an effective tool because we can demonstrate simple vowel production by attaching it to a vocal-tract model as well as to a person's neck to excite an actual vocal tract. (Vowel production is most effectively demonstrated when the glottis is closed but not vibrating; the resultant speech sounds become more intelligible.) The whistle-type artificial larynx is more realistic than the electrolarynx because it produces glottal-type sounds by means of airflow. Students can then learn about human phonation by connecting the lung model (the respiratory system) to the whistle-type larynx. The plate-type models are more appropriate for hands-on activities because students can change the shape of the model and investigate how shape change affects acoustic output. Using Umeda and Teranishi's model, one can change the shape of the vocal-tract model while the sound is in progress. The cylinder models are good introductory models because they enable students to quickly grasp how vocal output relates to resonator shape. The cylinder-type model is convenient because it does not require any setup; once the sound source is attached, the vowel sounds clearly. With the head-shaped models, students can easily understand the anatomy of the vocal tract, how and where the vocal tract is situated within the head, and how the tongue forms constrictions needed to produce each vowel. Furthermore, students can learn that it, instead of the bend of the vocal tract that is important in determining vowel quality, is the area function because with the head-shaped models we can demonstrate that the same vowel qualities are produced by straight and bent tubes.

4.2. Teaching in Actual Situations

We believe that mechanical models of the human vocal tract are useful for educating students of various backgrounds and ages (e.g., [2]). Physical models such as ours are especially suitable for nontechnical students, because they are intuitive. Also, we believe that education in acoustics is important not only for college students, but also for high school or even middle school and elementary school students, for whom such hands-on models are even more important.

4.2.1. College and graduate levels

We have already used the proposed educational system 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 when we used them in a series of lectures at Sophia University. Arai's models were also used several times for instruction in an Acoustic Phonetics course taught at both the Northwest Christian College and the University of Oregon, in Eugene, Oregon, USA, as part of the Oregon Summer Institute of Linguistics [43]. These models were found to help students grasp basic concepts in acoustic phonetics, particularly the source-filter theory. Furthermore, Arai's models as well as the lung and head-shaped models are used in Acoustics and Speech Communication classes at the Massachusetts Institute of Technology. The students are interested to see and hear real models of vowels with an excitation source. Some are mainly interested in different shapes for vowels, and the capability to simulate a variety of other shapes [44]. Arai's models are used at the University of California, Berkeley as well [28].

4.2.2. Voice and articulatory training

We are also attempting to apply of our education system to voice training and/or articulatory training such as those in speech pathology, language learning, and musicology. For those who have speech disorders or hearing impairments, it is sometimes very useful to show them the ideal shape of the vocal tract and/or to produce sounds through the model. It is especially important to show patients that a slight shift of articulators may result in a huge acoustic difference. Our system can be useful in a clinical situation for children who have speech disorders, because they can learn more about phonation and articulation by seeing, touching, and playing with the vocal tract models.

Not only can speech pathologists use these tools for their patients, but they can also train themselves using the system. For example, it is often difficult to imagine the relationship between the velopharyngeal opening and the acoustic output outside the mouth. The models for nasalized vowels, however, allow us to associate the acoustic output with the visual observation of the velopharyngeal opening. Using the models, we can experience how much opening is needed to make a certain degree of nasalization. Furthermore, it also becomes obvious that different vowels need different degrees of opening at the velopharyngeal port to achieve the same nasality. For example, with the models, we can easily see that the vowel /i/ requires only a small degree of velopharyngeal coupling, whereas the vowel /a/ needs a wider opening.

The proposed education system can also be useful for those who are learning a new language and have difficulty pronouncing sounds that do not exist in their own language. Many of our early models are suitable for this purpose for vowel production. Moreover, the models for the sonorants /r/ and /l/, stridents, and nasalized vowels should also be useful. For instance, native speakers of Japanese often have difficulty pronouncing /r/ and /l/. Showing them the ideal shape of the vocal tract might help learners to configure their own articulators to produce these sounds [45]. The system can also be used for voice training. For example, one important aspect of singing is correct breathing, and the lung models can be used to demonstrate abdominal respiration. Students can also learn the appropriate articulatory gesture for singing a foreign vowel, e.g., umlauts in German for non-native speakers.

4.2.3. Teaching for children

Finally, we had a chance to teach the acoustics of speech production to high school students using Arai's models as part of a regular Physics class [46–48]. Prior to our lecture, the students had just finished studying basic acoustics. They were familiar with the Doppler effect, and the technique for calculating the velocity of sound by applying the knowledge of resonance in an acoustic tube of variable length. We first used cylinder-type models to demonstrate vowel production. Since the students had studied resonance in a column of air prior to this lecture, they thought that only tube length was important, but this exercise confirmed that the shape of the tube also affects speech output. We then divided the students into groups for hands-on experiments using the plate-type models. The students lined up the plates by hole size based on the table of diameters that comes with the model. At the end of the class, we gave the students a questionnaire regarding the lecture. More than half of the students showed interest particularly in the plate-type models because they were able to manipulate the plates to simulate constrictions.

As mentioned earlier, teaching acoustics is also important for younger generations. In an attempt to reach the youth using these concepts, we contributed to the development of an exhibition at the Science Museum "Ru-Ku-Ru" in Shizuoka City, Japan [49]. This museum has recently opened under the concept of "watch, listen and touch." Arai's cylinder and plate-type models of the human vocal tract are installed in an exhibition at that museum. 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 hands-on experiments. Children can line up the plates in the order indicated to make Japanese vowels (and other vowels). The models are well suited to the theme of this museum.

4.3. Evaluation of System

To evaluate the effectiveness of our proposed educational system using the physical models of the human vocal tract, we conducted an experiment. First, we gave students a regular lecture without using the proposed system, under the assumption that such physical models are typically not available in a class. A second lecture employed the proposed educational system. An identical written exam was administered after each lecture, and the results were compared.

4.3.1. Experimental conditions

Twenty students who studied advanced acoustics as part of their speech pathology course participated in this evaluation experiment. First, we gave students an approximately three-hour introductory lecture in speech science, such as acoustics and speech production, with a set of slides as well as some audio and visual demonstrations that are not part of our physical models. After the lecture, we conducted the first written examination. After a lunch break, we gave them an hour-and-half demonstration of speech production using the proposed educational system. There was a 15-minute hands-on activity using the plate-type model at the end of the demonstration. Afterwards, the second written examination was administered, with exactly the same problem set. Finally, we asked the students to answer a questionnaire.

The problem set consisted of 28 multiple-choice questions. 12 questions (Problem Set A) are directly related to the demonstration and 16 questions (Problem Set B) deal with the basics of speech science covered in the first lecture. In each of the problems, part of the statement was missing and there were multiple fill-in-the blank options, from which the students were asked to choose the best answer.

4.3.2. Results

The overall correct rates of the first and second examinations were 73.9% and 79.6%, respectively. For Problem Set A, they were respectively 70.4% and 82.1%; for Problem Set B, they were 76.6% and 77.8%. Although some of the improvement in test scores may be due to learning effects, based on our experience using these tools in the classroom, it is unlikely that all of improvement is due to learning effects. The use of the tools in other classroom situations coupled with these results suggests that giving an extra demonstration using the models will reveal significant gains in the students' understanding of the basic acoustics of speech production.

4.3.3. Questionnaire

95% of the students were interested in the lung models and 90% thought that the models helped their understanding. Some typical comments on the long models are as follows.

- I appreciated that the models enabled me to see the effect of negative pressure in the thoracic cavity during inhalation.
- I was able to see the simultaneous movements of the lungs and diaphragm.
- The models helped me understand that the lungs change volume passively (not actively).
- Seeing the pressure with the U-shaped tube helped me grasp the concept of transglottal pressure difference.

100% of the students were interested in the cylinder-type models, and 100% thought that the models help their understanding. Some typical comments on the cylinder-type models are as follows.

- I was surprised that the cylinders made sounds like human-pronounced vowel sounds.
- Although I have learned the two-dimensional shapes of vowels, having the three-dimensional forms helped me associate vocal tract shape with the output sound.
- I was surprised that I heard similar vowels even with a straight tube.
- The models illustrated that whether the tube is bent or straight does not make a large difference in the vowel quality, rather the area functions are the key.
- I noticed that the materials from which the vocal tract is made are not crucial for producing similar sounds.

100% of the students were interested in the plate-type models and 90% thought that the models helped their understanding. Some typical comments on the plate-type models are as follows.

- I could participate in a hands-on activity to create my own vowels.
- I tried different shapes by trial and error and checked the sound.
- I learned that the shapes of two vowels, such as /i/ and /e/, are similar up to a certain point.
- I realized that a small difference in shape makes a huge difference in vowel quality.

100% of the students were interested in the "tongue fixed" models and 90% thought that they helped their understanding. Some typical comments on the tongue-fixed models are as follows.

- Combining the tongue-fixed models with the lung models, I easily learned the mechanism of phonation visually.
- I could imagine the vocal tract better with the tongue-fixed models than with straight-tube models.
- The movable velopharyngeal port helped me to understand the mechanism of nasalization.

100% of the students were interested in the head-shaped "manipulable-tongue" model and 95% thought that the model helped their understanding. Some typical comments on the manipulable-tongue model are as follows.

- I understood that the tongue plays an important role in speech production.
- I learned that the vocal tract shape is changed by the position of the tongue.
- I was surprised because a small degree of change in the tongue position produced a huge difference in vowel quality.

95% of the students were interested in Umeda and Teranishi's model and 95% thought that the model helped their understanding. Some typical comments on the Umeda and Teranishi's model are as follows.

- I noticed that shape (i.e., whether square or round) of cross section has no effect on vowel quality.
- I noticed that the area function is the key.
- I was surprised to observe that vowel quality can be changed only by sliding plastic strips, although it is difficult to slide them without sliding the adjacent strips.

Overall, it seems that through the simple models the students were able to see "the essence." Many students became interested and realized more about the mechanisms and phenomena, even if their scores did not always improve.

5. SUMMARY

We proposed and discussed an education system in acoustics and speech production using various physical models of the human vocal tract. We confirmed that physical models, when used in a classroom, are particularly effective for increasing students' understanding of the theories of speech production. The combination of the lung models, artificial larynx, Arai's cylinder-type and plate-type models, Umeda and Teranishi's model, and the head-shaped models with nasal cavities provides the basis for an effective, comprehensive, and systematic education in speech production. When this system is integrated with computer models, learning will be further enhanced. In the future, we can incorporate prosodic aspect into our educational system. If we use a sound source with a time-varying fundamental frequency, we can produce pitch-accented words in Japanese.

Our ultimate goal is to make the above-mentioned tools and this comprehensive system of education widely available, so that others can benefit from the models and the educational philosophy we propose in this paper. We hope that our vocal tract models will spread widely throughout the world in an effort to promote education in acoustics and speech science. The audio and visual demonstrations of the education system is partly available at

http://www.splab.ee.sophia.ac.jp/Vocal_Tract_Model/ or
http://www.splab.net/Vocal_Tract_Model/.

ACKNOWLEDGMENTS

I acknowledge the assistance of members of Arai Laboratory, the Phonetics Laboratory, and the Center for Speech and Hearing at Sophia University, especially Nobuyuki Usuki, Eri Maeda, Michiko Toyama-Yoshida, who helped me to work on this project; Kikuo Maekawa, who gave me an opportunity to write on the project of replicating Chiba and Kajiyama's models; Noriko Umeda, who gave us a replica of Umeda and Teranishi's model along with helpful advice; Shuichi Itahashi, Ken-ichi Kido, and Shun-ichi Nakamura, who gave me many valuable suggestions; John Ohala, who showed me his physical models at the Exploratorium and was a rich source of information, especially historical; Dawn Behne and Yue Wang, who gave me information on the lung model; Terri Lander, who agreed to use Arai's models in her class as part of our project. I also would like to thank members of the Speech Communication Group at Massachusetts Institute of Technology, especially Kenneth N. Stevens, who gave me several opportunities to teach on the acoustics of vowel production using the models in his course; Hirokazu Sato and Naoki Ishii of NTT Advanced Technology Co., who commercialized Arai's models, VTM-10 (http://www.sp4win.com/); and staff members of the Shizuoka Science Museum. Finally, I would like to thank my family members, Takashi Arai, Michiko Arai, Tomomi Arai and Tomoe Arai, who helped me in various ways. This research was supported in part by Grants-in-Aid for Scientific Research (A-2, 16203041 and C-2, 17500603), the Moritani Scholarship Foundation, and the Sound Technology Promotion Foundation.

REFERENCES

[1] P. B. Denes and E. N. Pinson, *The Speech Chain: The Physics and Biology of Spoken Language*, 2nd ed. (W. H. Freeman, New York, 1993).

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

[3] T. Arai, E. Maeda, N. Saika and Y. Murahara, "Physical models of the human vocal tract as tools for education in acoustics," *Proc. 1st Pan-American/Iberian Meet. Acoust.*, Cancun (2002).

[4] N. Usuki, T. Arai and Y. Murahara, "Prototype of vocal-tract models for education in acoustics of vowel production," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 399–400 (2001).

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

[6] 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).

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

[8] M. Yoshida, N. Usuki, T. Arai, Y. Murahara and T. Sugawara, "A challenge for visualizing vocal tract resonance: Devising a tool for acoustic education in speech science," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 399–400 (2002).

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

[10] T. Arai, "Physical and computer-based tools for teaching Phonetics," *Proc. Int. Congr. Phonet. Sci.*, Vol. 1, pp. 305–308, Barcelona (2003).

[11] T. Arai, "Education in acoustics using physical models of the

human vocal tract," *Proc. Int. Congre. Acoust.*, Vol. III, pp. 1969–1972, Kyoto (2004).

[12] T. Arai, "Visualizing vowel-production mechanism using simple education tools," *J. Acoust. Soc. Am.*, **118**, Pt. 2, 1862 (2005).

[13] T. Arai, "Lung model and head-shaped model with visible vocal tract as educational tools in Acoustics," *Proc. Spring Meet. Acoust. Soc. Jpn.*, Vol. 1, pp. 273–274 (2005).

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

[15] Eurospeech Education Arena (e.g., http://www.dcs.shef.ac.uk/spandh/ed_arena/ and http://www.dcs.shef.ac.uk/˜sas/ed_arena/).

[16] Symposium on Tools for Teaching Phonetics at International Congress of Phonetic Sciences 2003 (http://shylock.uab.es/icphs/plenariesandsymposia.htm).

[17] W. von Kempelen, *Mechanismus der menschlichen Sprache und Beschreibung einer sprechenden Machine* (Wien, 1791).

[18] H. Dudley and T. H. Tarnoczy, "The speaking machine of Wolfgang von Kempelen," *J. Acoust. Soc. Am.*, **22**, 151–166 (1950).

[19] J. L. Flanagan, *Speech Analysis, Synthesis and Perception* (Springer-Verlag, 1965).

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

[21] D. Lindsay, "Talking Head," *Am. Heritage Inventi. Technol.*, **13**, 56–63 (1997).

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

[23] 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.*, **22**, 195–203 (1966).

[24] J. Dang and K. Honda, "Acoustic effects of the pyriform fossa on vowel spectra," *Tech. Rep. IEICE*, SP95-10, pp. 1–6 (1995).

[25] K. Honda, H. Takemoto, T. Kitamura, S. Fujita and S. Takano, "Exploring human speech production mechanisms by MRI," *IEICE Trans. Inf. Syst.*, **E87-D**, 1050–1058 (2004).

[26] http://www.deutsches-museum.de/

[27] http://www.exploratorium.edu/exhibits/vocal_vowels/

[28] J. Ohala, personal communication (2001).

[29] Y. Lebrun, *The Artificial Larynx* (Swets & Zeitlinger B.V., Amsterdam, 1973).

[30] R. R. Riesz, "Description and demonstration of artificial larynx," *J. Acoust. Soc. Am.*, **1**, 273–289 (1930).

[31] W. H. Gardner and H. E. Harris, "Aids and devices for laryngectomees," *Arch. Otolaryngol.*, **73**, 145–152 (1961).

[32] 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).

[33] D. H. Klatt and L. C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.*, **87**, 820–857 (1990).

[34] E. Maeda, N. Usuki, T. Arai, N. Saika and Y. Murahara, "Comparing the characteristics of the plate and clynder type vocal tract models," *Acoust. Sci. & Tech.*, **25**, 64–65 (2004).

[35] E. Maeda, N. Usuki, T. Arai and Y. Murahara, "The importance of physical models of the human vocal tract for education in acoustics in the digital era," *Proc. Chin-Jpn. Jt. Conf. Acoust.*, pp. 163–166, Nanjing (2002).

[36] 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.*, pp. 341–342 (2003).

[37] N. Saika, E. Maeda, N. Usuki, T. Arai and Y. Murahara, "Developing mechanical models of the human vocal tract for education in speech science," *Proc. Forum Acusticum Sevilla* (2002).

[38] M. Y. Chen, "Acoustic correlates of English and French nasalized vowels," *J. Acoust. Soc. Am.*, **102**, 2360–2370 (1997).

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

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

[41] F. Bell-Berti, "Velopharyngeal function: A spatial-temporal model," in *Speech and Language: Advances in Basic Research and Practice 4*, N. J. Lass, Eds. (Academic, New York, 1980), pp. 291–316.

[42] A. S. House and K. N. Stevens, "Analog studies of the nasalization of vowels," *J. Speech Hear. Disord.*, **21**, 218–232 (1956).

[43] T. Lander and T. Arai, "Using Arai's vocal tract models for education in Phonetics," *Proc. Int. Congr. Phonet. Sci.*, Vol. 1, pp. 317–320, Barcelona (2003).

[44] K. N. Stevens, personal communication (2004).

[45] E. Maeda and T. Arai, "Additional physical models of the human vocal tract as tools for education in language learning," *Proc. Int. Congr. Acoust.*, Vol. III, 2321–2322, Kyoto (2004).

[46] E. Maeda, T. Arai, N. Saika and Y. Murahara, "Lab experiment using physical models of the human vocal tract for high-school students," *Proc. 1st Pan-American/Iberian Meet. Acoust.*, Cancun (2002).

[47] E. Maeda, T. Arai, N. Saika and Y. Murahara, "Education in acoustics using mechanical models of the human vocal tract in high school," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 299–300 (2002).

[48] E. Maeda and T. Arai, "Education in acoustics for high-school students using mechanical vocal tract," *Trans. Tech. Comm. Educ. Acoust.*, Acoust. Soc. Jpn., EDU-2003-09, pp. 1–6 (2003).

[49] http://www.rukuru.jp/

Takayuki Arai received the B.E., M.E. and Ph.D. degrees in electrical engineering from Sophia Univ., Tokyo, Japan, in 1989, 1991 and 1994, respectively. In 1992–1993 and 1995–1996, he was with Oregon Graduate Institute of Science and Technology (Portland, OR, USA). In 1997–1998, he was with International Computer Science Institute (Berkeley, CA, USA). He is currently Professor of the Department of Electrical and Electronics Engineering, Sophia Univ. In 2003–2004, he is a visiting scientist at Massachusetts Institute of Technology. His research interests include signal processing, acoustics, speech and hearing sciences, and spoken language processing.