Vocal-tract models and their applications in education for intuitive understanding of speech production

Takayuki Arai*

Department of Information and Communication Sciences, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan

Abstract: This paper describes vocal-tract models that we have developed and their applications in education in acoustics. First, we grouped the representative models into two major categories depending on their configuration: straight vs. bent. Then, within each category, we discussed the characteristics of each model in terms of its degree of freedom of movement. Subsequently, we review lectures using the vocal-tract models and report the results of tests and questionnaires carried out simultaneously with the lectures, some of which were re-evaluated in this paper. On the basis of the review, we further discuss how education should be carried out using the vocal-tract models, and we made the following conclusions: 1) the models are useful not only for education on sounds themselves but also for phonetic education; 2) it is important that appropriate models should be selected depending on specific purposes; and 3) it is necessary to continuously develop more models having different properties with wider variations in the future.

Keywords: Vocal-tract model, Speech production, Education in acoustics, Phonetic education, Acoustic phonetics

PACS number: 43.10.Sv, 43.70.Bk [doi:10.1250/ast.37.148]

1. INTRODUCTION

In the modern era, The Vowel [1] by Chiba and Kajiyama concluded an argument on the mechanisms of vowel production [2] and founded the modern field of speech science and acoustic phonetics. Later, with a focus not only on vowels but also consonants, this field was systematically and rapidly developed by core founders, such as, Fant [3] and Stevens [4]. Thus, the book, The Vowel [1], was the basis of modern speech science and acoustic phonetics; in this book, for the first time in a history, Chiba and Kajiyama conducted a three-dimensional (3D) measurement of a configuration of the human vocal tract during vowel production. Furthermore, in the section on “Artificial Vowels,” they made clay models based on their 3D measurements and simplification. Finally, they showed that vowel-like sounds can be produced by feeding some types of sound sources from the glottis end of the tubes, or the vocal-tract models. It has already been shown that mechanical models, such as the one by von Kempelen in the 18th century (1791) [5,6], produced vowels with a similar mechanism as described above. However, Chiba & Kajiyama further investigated the relationship between 3D configurations of the vocal tract and the spectral characteristics of vowels, and proved that the difference in vowel qualities are caused by the difference in configurations. Later, Fant [3] established the fundamental principal of vowel production as “the source-filter theory,” and furthermore, Stevens modeled any sounds of speech as a combination of resonance with one of the four types of sound source: phonation (glottal) source, aspiration source, transient source, and friction source [7]. Based on this concept, Klatt developed his formant synthesizer [8].

Studies using vocal-tract models are also done by several researchers. Umeda & Teranishi (1966) investigated phonemic and vocal features with a vocal-tract model, in which bars were inserted from one side and their location was adjusted to achieve arbitrary configurations (step-wise approximation) [9]. More recently, Honda et al. [10] made a physical model based on the precise 3D configuration of a vocal tract with magnetic resonance imaging (MRI) technology. Honda et al. (e.g., [11]) has been developing a series of speaking robots. Although it is suspected that researchers realized such vocal-tract models are useful for education, it is probably true that no one tried to investigate “education in acoustics by using vocal-tract models” from the science education point of view.

The author of this paper was the first to replicate the vocal-tract models in the book The Vowel [12], and is...
continuously developing vocal-tract models with proper improvements for various types of educational applications in acoustics [13–16]. The main models which result from such development are shown in Fig. 1. As you can see, this figure displays widely ranging vocal-tract models with various characteristics. Through careful study, we have learned how the models are best used to achieve success in education, and what aspects of the models still need to be improved. Therefore, in this paper, we first review the characteristics of the vocal-tract models and discuss how they are effective for education. In that process, we also discuss the possibility that special-purpose models will be selectively developed in the future.

2. VOCAL-TRACT MODELS

In Fig. 1, nine models are displayed. They are grouped into two categories; the bottom panel shows the straight models and the top panel shows the bent models. The horizontal axis of this figure is the degree of freedom in terms of their movement from static models with the null degree of freedom through dynamic models with moderate and high degrees of freedom. In this section, we will review the characteristics of each model based on this categorization.

2.1. Straight Models

The bottom panel of Fig. 1 shows the category of straight models. The five models in the bottom panel will be reviewed in each of the following sections.

2.1.1. VTM-C10

The cylinder-type vocal-tract models [12,14] shown as VTM-C10 are replicas based on Chiba & Kajiyama’s configurations of five Japanese vowels in *The Vowel* [1]. Like other models, VTM-C10 can intuitively show the source-filter theory as well as the relationship between the vocal-tract configuration and the quality of vowels. By feeding a sound source, such as sounds produced by an artificial larynx from the hole at the glottis end, we can demonstrate vowel production instantaneously. For this reason, they have a wide application in various types of lecture and science workshops. In fact, VTM-C10 is installed in the exhibitions for vowel production at Shizuoka Science Museum “Ru-Ku-Ru” (Shizuoka, Japan) and Hitachi Civic Center Science Museum (Hitachi, Japan) under my supervision. The exhibition of vocal-tract models at the Exploratorium in San Francisco (California, USA) was supervised by Prof. John J. Ohala at the University of California, Berkeley. Although it seemed that the configurations of the models at the Exploratorium were based on Chiba & Kajiyama [1], the models were not a precise reproduction of [1], but were deformed.

2.1.2. VTM-T20

The configurations of VTM-C10 are based on the precise replication of the 3D measurements of the vocal tract by Chiba & Kajiyama and their simplification. However, due to the round bottle shape based on the first-order approximation of the cross-sectional area functions, the production process of the models was complicated, and the cost of making the models was high. Therefore, we further simplified the configurations of VTM-C10 and obtained VTM-T20 [16]. In VTM-T20, variations of the diameters of the outer tubes and constriction holes are minimized, and the designs were made of fewer numbers of tubes with different lengths and
diameters. Therefore, we succeeded in minimizing production costs. The quality of the vowel /o/ was also improved with VTM-T20.

It was necessary to use transparent materials, such as, acrylic resin, to see the vocal-tract configurations from the outside of the models. However, in VTM-T20, the vocal-tract configurations are reflected in the outer shapes, so that the materials do not have to be transparent anymore. Therefore, the model can be made of plastic, metal (e.g., aluminum), wood, paper and so on, in addition to acrylic resin. Furthermore, people with visual impairment can also touch and feel the configurations with their hands. The usage of VTM-T20 for explaining vowel production is almost the same as VTM-C10. Because of overall discussions of VTM-T20 including the quality of vowel (/o/ with slight improvement) and the cost, this model is more appropriate for more applications than VTM-C10.

2.1.3. VTM-S20

The configurations of VTM-T20 feature a slider, which slides inside the long outer tube, and a hole for the constriction. Based on this concept, we designed VTM-S20 [13,16]. While VTM-C10 and VTM-T20 are static models, VTM-S20 has a degree of freedom in motion, in that the constriction part slides, which is the first degree of freedom. In addition, the diameter of the constriction hole (degree of constriction) is variable, allowing us to produce more vowels, which is the second degree of freedom. Finally, the lip aperture (the third degree of freedom) helps to cover the wide area in the vowel space with minimal degrees of freedom (a maximum of three). For a quick demonstration, we do not use the mechanism for lip constriction (rounding) but can use our thumb and index finger to form lips to simulate lip rounding. On the other hand, if the sliding part with a variable degree of constriction (see [16] more details) is not available, multiple column parts with different diameters can be used one by one. For example, when the diameter of the outer tube is 34 mm and two sliding parts are used, the model can produce five vowels: the sliding part with the diameter of 32 mm is used for the vowels /i/, /a/, /o/, and /u/; and a sliding part with the diameter of 27.5 mm is used for the vowel /e/ with proper positioning of the sliding parts. Thus, even if we do not use the lip-rounding mechanism, with only a few degrees of constriction, five intelligible vowels can quickly be demonstrated by preparing one outer tube and two sliding parts. That is an advantage of this model.

As we just saw, VTM-S20 can achieve similar configurations to those of VTM-T20; because each of these two models have different characteristics, it is important to understand how and which model we use, and how we combine them, depending on the purpose of the demonstration, in order to highlight the advantages of each model. For example, if we start with the production of the schwa vowel and move onto the five vowels, we can use VTM-S20 with no sliding part at the beginning (for schwa). At this point it would be good to compare the sound of the sound source alone with the sound source fed into the outer tube. With VTM-S20, a user has to know the right location of the sliding part with the right degree of constriction. VTM-S20 is basically one model, whereas VTM-T20 is five different models corresponding to five vowels. Because VTM-S20 has this simple mechanism but can produce different vowels easily, it is valuable especially when multiple models like VTM-T20 cannot be carried. Yet another advantage of VTM-S20 is that it allows us to produce vowels with dynamic variations. As a result, the “overshoot” effect in vowel perception yields more intelligible vowels. In other words, it is often observed that a vowel sequence, such as, /a/ → /i/ → /u/, can be more clearly produced, even if each vowel is less intelligible in isolation. When we look closely at our production mechanism, a similar phenomenon happens. In other words, even if our articulator does not reach a target position but continuously moves to the next target position (target undershoot), our auditory mechanism compensates for such undershoot by perceiving the target sound [17]. Such phenomenon underlying speech dynamics can be demonstrated by the dynamic models.

2.1.4. VTM-P10

VTM-P10 provides a set of 10-mm thick plates with holes of different diameters in the center of each plate, and a user can line them up together to form an arbitrary configuration of the vocal tract (step-wise approximation) [12,14]. It has a higher degree of freedom because the diameter of each of the 10+ plates can vary, and as a result, a huge number of vowels can be achieved. In addition, children can play with them actively like playing with a puzzle. However, due to the higher degree of freedom, children need to be guided by prototypes of standard configurations, otherwise they will combine them randomly. At the Shizuoka Science Museum mentioned earlier, VTM-P10 and VTM-C10 are installed as an exhibition of the mechanism of vowel production.

2.1.5. Umeda and Teranishi’s model

This model is based on the study by Umeda & Teranishi [9] and has a set of bars inserted from one side of the vocal-tract model to form an arbitrary vocal-tract configuration as mentioned in Sect. 1. Like VTM-P10, the degree of freedom is high, so this model can produce many different vowels; however, training is needed to produce a target vowel. Furthermore, manual manipulation is difficult, because when one bar is moved, an adjoining bar may move inadvertently. Therefore, in Arai (2010) [18], each bar of Umeda and Teranishi’s model is connected to each actuator, and the movement of each bar was fully
controlled by a computer to change the vocal-tract configuration in real-time.

2.2. Bent Models

The top panel of Fig. 1 shows the category of bent models. The four models in the top panel will be reviewed in each of the following sections.

2.2.1. Head-shaped (HS) model

This model is an outline of the human head, and from the side, the vocal-tract configuration is the same as the mid-sagittal cross-section typically appearing in textbooks. Furthermore, as shown in Fig. 1, this model has not an only oral cavity but also a nasal cavity, and the rotating valve at the velo-pharyngeal port changes the airflow to these two cavities. As a result, the non-nasalized vowel /a/ is produced when the velo-pharyngeal port is closed, and a nasalized vowel is produced when velo-pharyngeal port is opened by rotating the valve. In demonstrations using only straight models, as shown in the bottom panel in Fig. 1, we received the comment that it is hard to imagine how the straight model corresponds to the actual vocal tract in the head. This HS model addresses such comments. The author uses this model in his lectures along with the straight models as much as possible. Moreover, by combining the lung model [14] with the HS model, it is possible to explain the whole system of vowel production from breathing and phonation to articulation. In addition, vowels produced by these models sound human-like, and this makes the demonstrations more impressive, which leads to better understanding. Please note that the static HS model corresponds to a single vowel, so that different models must be prepared for different vowels.

2.2.2. Block-type (BL) model

For the static HS model, we need different models for different vowels. Therefore, we tried to design a single bent vocal-tract model that can produce more than one vowel. This concept gave us the block-type (BL) model [19]. This model is bent at a right angle in the middle of the vocal tract, and blocks are placed inside the vocal tract to produce different vocal-tract configurations. The typical blocks are for the front cavity, back cavity, and the lips; the blocks allow us to reproduce any variety of vowel. Like many previous models, the outer plates are made of transparent acrylic resin, so the blocks are visible from the outside. Placing a head-shaped urethane cover on top of the model produces an effect similar to the HS model, which allows one to see where the vocal tract is located inside the head, for example.

2.2.3. Flexible-tongue (FT) model

The flexible-tongue (FT) model resulted from our desire to have flexible tongue movement [15]. This model has a tongue made from flexible materials, and the shape of the tongue can be changed by manipulating the tongue from underneath. A user can manipulate and adjust the position of the tongue manually, check the produced sounds aurally, and observe the configurations visibly. Because the materials of the tongue are flexible, the vocal-tract configuration is variable with practically infinite degrees of freedom, and the resulting sounds are rich in variety. However, it is also true that one needs a huge amount of practice to achieve a target sound. It is worth noting that possible applications are wide ranging because the model can produce both vowels and consonants.

2.2.4. Approximant (AP) model

Many of the models discussed so far produce vowels, with the exception of the FT model. The demand for the vocal-tract model is high. It is true that the approximants, such as, English /r/ and /l/, may be produced by the FT model, and such sounds can easily be produced by the vowel models. In the case of the FT model, however, the degree of freedom was too high, so that we need special training to produce a target sound. Therefore, we designed the AP model [20]. Although possible sounds that this model produces are limited, it is designed to produce approximants with a low degree of freedom. This model focuses on the English retroflex approximant /r/ and alveolar lateral approximant /l/. We already made models of /r/ and /l/ in the previous study [14]. However, because these models were static, they only produced steady sounds and were problematic because the sounds produced were less intelligible. Therefore, this AP model is designed to achieve a simple motion yielding a sequence like vowel $\rightarrow$ consonant $\rightarrow$ vowel by folding the first half of the tongue towards the palate and unfolding it back to the original position. This motion is done by rotating a lever, which only has a single degree of freedom. Besides this AP model, the FL model for the flap sounds was also developed with a similar concept [20].

3. EVALUATION OF EDUCATIONAL BENEFITS OF THE VOCAL-TRACT MODELS

The authors have been conducting lectures and science workshops, demonstrating the vocal-tract models explained so far. The audience and participants have included people of all ages, including children and graduate students. In addition to our lectures, multiple educational institutions are using the vocal-tract models we developed. In this section, the benefits of the models for education will be qualitatively and quantitatively discussed based on the data collected in the lectures and workshops.

3.1. Example of Lectures on Acoustics

Among the proposed vocal-tract models, we used the lung model, the HS model, the VTM-C10 and the VTM-P10 in a lecture on acoustics and speech production for college students who aim to be pathologists in speech and
hearing. We reported on this lecture in [14]. During the lecture, we gave a written exam on acoustics, and the results were reported in that study. In this paper, we review the results of the written exam and we re-evaluate the educational benefits using statistical analysis.

3.1.1. Contents and exams conducted in the lecture

Twenty college students who aim to be pathologists in speech and hearing participated in the lecture. They had all taken a basic acoustics class beforehand. In the first half of this class, Lecture 1, the author spoke for three hours on basic speech science including acoustics and speech production. This was a typical lecture-type class. Lecture 1 used mainly slides and verbal explanations, and we limited demonstrations to widely-used audio and video files (such as a video clip showing vocal fold vibration). However, we did not use the vocal-tract models in any of the explanations. After Lecture 1, we conducted the first exam.

After lunch, we had the second half of the lecture (Lecture 2). In Lecture 2, we spoke about speech production for approximately 1.5 hours, but we included a demonstration of the vocal-tract models. The last 15 minutes of the class was a hands-on practice session using the VTM-P10 model. Right after Lecture 2, we conducted the second exam.

The contents of the first and second exams were exactly the same. There were 28 questions, and they were divided into two groups, G1 and G2; G1 questions were related to Lecture 1 but not directly related to Lecture 2, whereas G2 questions were related to Lecture 2. G2 questions were divided into G2a and G2b, where G2a questions were only indirectly related to the explanations using the vocal-tract models, and G2b questions were directly related to the demonstrations using the vocal-tract models. Each question on the exam consisted of a short sentence with one missing word. The students were given multiple choices to fill in the blanks, and were asked to choose the best option.

3.1.2. Overall results from the exams

As reported in [14], the results of the two exams showed an improvement from 73.9% correct on the first exam to 79.6% correct on the second exam. Part of this improvement may be due to a learning effect, as we gave the same exam twice. However, even if there was a learning effect, by comparing the difference in the improvements of correct rates in the G1 and G2 groups we can show that the model demonstration increased learning. Therefore, we evaluate the degree of improvement in each question group.

3.1.3. Definition of the degree of improvement

We define the degree of improvement as follows: When a particular question had a correct answer in the first exam and an incorrect answer in the second exam, the degree of improvement is -1 to reflect that the understanding deteriorated. When a particular question had incorrect answers in both the first and second exams, the degree of improvement is 0 to reflect that the understanding was not good. When a particular question had correct answers for both the first and second exams, the degree of improvement is +1 to reflect that the understanding was sufficient, although not improved. Finally, when a particular question had an incorrect answer in the first exam and a correct answer in the second exam, the degree of improvement is +2, to reflect significant improvement.

Please note what these categorizations mean. The change from a correct answer to an incorrect answer indicates that the lecture does not show any effect. The change from an incorrect answer to an incorrect answer indicates the lecture does not show sufficient effect. The change from a correct answer to a correct answer indicates an effect of a deeper understanding. The change from an incorrect answer to a correct answer indicates the lecture functions sufficiently. Because we treat them as non-parametric, the only importance to these degrees of improvement is their order; the values do not indicate qualitative measure. That is, +2 does not mean twice as much as +1.

3.1.4. Comparing the degrees of improvement between G1 and G2

First, we compared G1 and G2. In this comparison, we analyzed the degrees of improvement in groups G1 and G2, and we applied the unpaired Mann-Whitney U test. As a result of the test, G2 had significantly better improvement than G1 ($p = 0.004$). Figure 2 shows the proportions of the degrees of improvement for each question group.

3.1.5. Comparison between G1 vs. G2a and G1 vs. G2b

Next, we divided G2 into G2a and G2b, and a test similar to that applied in the previous section was conducted. As a result of the unpaired Mann-Whitney U
test, G2b had significantly better improvement than G1 ($p = 0.011$), whereas G2a only showed a significantly better tendency than G1 ($p = 0.054$). Figure 3 shows the proportions of the degrees of improvement for each question group.

### 3.2. Lecture in Phonetics

A lecture was given at Rutgers University in the USA using the vocal-tract models VTM-T20 and VTM-S20, reported in [16]. In this section, we discuss the results of the questionnaire given in that lecture.

#### 3.2.1. Contents of the lecture

Our lecture spanned three classes as a part of a larger course; the first part of the lecture took place during the first half of the course, and the second two parts of the lecture took place during the second half of the course. Twenty to twenty-five students participated in the lecture. The first part of the lecture explained the topic of articulation using the vocal-tract models in a demonstration lasting approximately 30 minutes. First, the outer tube of VTM-S20 with no constriction was used to produce the schwa sound. Then, the constriction was inserted, and front and back vowels were produced. Next, using the VTM-T20 model, the change in sound due to difference in tongue height was explained by comparing the vowels /i/ and /e/.

Finally, lip protrusion (rounding) was explained using the vowels /u/ and /o/, again using the VTM-T20 model.

During the second half of the course, the topic of the final two classes was acoustic phonetics. In the first class, we first analysed the schwa sound produced from the outer tube of the VTM-S20 model. We determined whether the output resonances of this uniform tube with a constant cross-sectional area having one end open and one end closed measure the same as those predicted by acoustic theory. Secondly, during that class we used the two-tube model VTM-T20 to produce the vowel /a/. Again, analysis was done to determine whether the resonance frequencies of the output sound matched those predicted by acoustic theory. In the final class, we used the three-tube model VTM-T20 to produce the rest of the vowels, and analysis determined whether the resonance frequencies measured from the output sounds matched those predicted by acoustic theory.

#### 3.2.2. Results of the questionnaire

After each class in the first and second half of the course, a questionnaire was given asking the students to describe their experience. The summary appears below. The numbers in the parentheses show the frequency of similar answers.

**[The results of the questionnaire after the first class in the first half of the course]**

- The models were interesting, helpful and useful for speech education. (10)
- The models helped me understand the relationship between vocal-tract shape and vowel quality. (3)
- The models produced realistic (human-like) vowels that could easily be differentiated. (5)
- The models enabled one to visualize sound and gain a better understanding of speech production. (6)
- The models helped me to visualize the location of constrictions in the vocal tract. (1)
- The demonstration helped me understand what the glottal source sound is like. (1)

**[The results of the questionnaire after the classes in the second half of the course]**

- The acoustic models were interesting, helpful and/or useful. (9)
- Whereas the vocal tract is usually invisible, these models make it visible. (4)
- The models provided a 3D visual, which is more effective than a picture. (1)
- The models illustrated how the shape of the vocal tract affects the output. (3)
- The models helped me understand how sound is produced in the vocal tract. (2)
- It was helpful to record the sounds with speech analysis software and then match the measurements up to the equations we had just solved. (3)
- The models produced realistic sounds, with measurable formants which were almost identical to the vowels they represent. (1)
- The models helped me visualize the size of the resonator and see how resonance is affected by the size of the connecting tube. (1)
- The models are enjoyable because they activate the senses and they turn an abstract idea into a more solid one. (1)
3.3. Lecture to Graduate Students Aiming to be Pathologists in Speech and Hearing

A lecture using the AP vocal-tract model for approximants was done as a part of a graduate course on acoustic phonetics. About one half of the class were language teachers or students who hope to be pathologists in speech and hearing. The results from the questionnaire conducted after the lecture were reported in [21]. We discuss the results below.

[Advantages of the models]
- One can experience the articulation by listening to the sound. It is easy to be interested.
- It is easy to imagine where each part of the tongue is supposed to be located.
- We can clearly understand the difference in articulation.
- The model is visible from various angles.
- We can experience the direct relationship between tongue movement and change of sound.

[Areas for improvement]
- Coloring the full body of the tongue might look better.
- The more realistic the tongue is, the more easily we can compare it with our own tongue and the more benefits for learning can be expected.
- It would be nice if more precise instructions were made as to where we are supposed to look during the demonstration.
- It would be even better if it were possible to switch between /r/ and /l/ instantaneously.
- The model itself is too small, so that it is hard to see from a distance.
- The manipulation looks a bit difficult; it seems that the lever has to be manipulated quickly.

4. GENERAL DISCUSSIONS

In this paper, we discussed the vocal-tract models developed so far. We divided them into two groups: bent and straight, and we arranged them in order of degrees of freedom with respect to movement of the model. Then, we reviewed how each of the vocal-tract models is used during lectures as well as the effect the model had on education. We also discussed the results of the questionnaire participants took after using the models. In this section, we will discuss how the vocal-tract models affect education in acoustics.

4.1. What We Mean by “Education in Acoustics” and Purpose-Specific Vocal-Tract Models

When we say Education in Acoustics in a strict sense, we often focus on the study of sound itself. This would include the fact that sounds are produced by vibrating objects, the vibrations are transmitted through a medium such as air, the wavelength of sound waves is related to the frequency of the vibration and the speed of sound, and it also includes the resonance and spectrum of sounds. Because speech is a significant part of daily life, it is reasonable to use speech for an example of sound to study. However, the more narrow definition of Education in Acoustics does not necessarily include how vowels and consonants are produced. Nevertheless, even if there are no explanations of vowel and consonant production, using speech as an example of sound is still useful, and from our experience we also know that the vocal-tract models are effective for teaching about sound in the narrow sense [14,16].

The term “Education in Acoustics” may also be used in a broad sense. For example, phonetic education or pronunciation training, where one teaches pronunciation, is included in the broad definition of Education in Acoustics, as is music training, where we teach music. The vocal-tract models, which we focused on in the present paper, are especially effective for phonetic education. In addition, they are also useful for speech pathologists in speech therapy for speech and language disordered people. In such situations, giving instructions and training to the ‘target’ sound is important. The vocal-tract models that the author has proposed help a learner understand how the sound is produced. For this application, the ultimate goal is to produce the target sound, and therefore an explanation of acoustic theory is not necessary. In such an environment, phonetic education has little in common with the narrow definition of Education in Acoustics. However, it is a part of the broader definition. And certainly, in any application where it is necessary to understand acoustic theory, the vocal-tract models are, of course, helpful. Additionally, we can use them for lectures in graduate courses and studies by researchers as a part of one’s experimental equipment. (See [16] for examples in undergraduate and graduate courses). Learners and teachers will also find the vocal-tract models to be useful when studying foreign languages.

Thus, the vocal-tract models are useful under the narrow and broad definitions of Education in Acoustics; the former regards what sound is, and the latter includes phonetic education and pronunciation training. The issue, then, is which model is optimal for a given application? Naturally, which vocal-tract model one chooses will vary depending on his/her goals, and it is rather important to select the appropriate models given one’s objective. In this sense, it behooves us increase the variations of models to fit the needs of the many target usages one might conceive of.

In general, when we explain a phenomenon, there are great educational benefits to using simple educational tools, with features that are directly related to a target phenomenon, but which reduce irrelevant factors as much as possible. This is true, even if the tools represent articulation differently than what actually occurs in the vocal tract. For
example, VTM-S20 is different from the actual vocal tract in that its shape is straight, and an actual tongue does not slide as the slider of VTM-S20 does. Nevertheless, it is extremely effective when explaining the importance of location of constriction for vowel production. This is so, because irrelevant factors were substantially eliminated from the model. As another example, the realistic tongue and its movement are easy for learners to identify when practicing their pronunciation. Using a model with a bent vocal tract, where the tongue is relatively realistic and its movement can simulate realistic dynamical change helps one to imagine the phenomena, even if the model is simplified to a certain degree. For this application, we would expect the AP model to be useful. Of course, with the FT model it is possible to produce many types of sounds because the tongue can be moved relatively flexibly; however, the model is hard to manipulate. Thus, we have found that models having fewer degrees of freedom, with limitations in terms of movements, as with the AP model, are especially useful for such applications as pronunciation training. And it is not an exaggeration to say that there is an urgent need for the development of bent models targeting sounds frequently observed in articulation disorders by speech and language disordered patients. For example, it is reported that the /r/ sound in Japanese is difficult for children to acquire [22]. Therefore, showing the flap sound, typically perceived as Japanese /r/, using the FL model, or showing its allophones with the AP model are both effective [22]. In addition, models for other sounds, such as /s/ and /ʃ/, would be very beneficial.

Thus, it is not necessarily the case that models having lower degrees of freedom are less beneficial than ones with higher degrees of freedom. Likewise, models having more complicated mechanisms for speech production are not necessarily more beneficial than simple ones. Fewer degrees of freedom make it possible for everyone to produce a target sound quickly and simply. Furthermore, the simplicity of a model makes it possible to show the essence of certain phenomena more effectively. In sum, each model has its own advantages, and we can increase educational effects by combining the models wisely to emphasize the relative values of each one, depending on one’s purpose and use.

4.2. How to Demonstrate the Effectiveness of Education

In Sect. 3, we introduced the results from examinations as well as a questionnaire conducted in some lectures. If we want to demonstrate how effective the vocal-tract models were for education, we need to verify that learners understood the contents of the lecture. One of the best ways to show such effectiveness is by having two lectures: one with vocal-tract models and the other without. Then, we conduct the same exams and compare the learners’ performance to see whether there is a significant difference in test results. However, in this case, a problem might arise if there is unfairness due to different qualities of lecturers between the two classes. To address this potential inequity, we need to have a follow-up lecture using vocal-tract models for the class where no models were originally used. However, it is often the case that there is no extra time to do this.

In the class introduced in [14], the whole class listened to a lecture without vocal-tract models and then took the first written exam. Then they heard a second lecture in which the models were used, and they took the same exam a second time. For a result, student performance on the second exam was better than that of the first exam with a statistically significant difference. However, this would be due, in part, to a training effect. We, therefore, divided the whole set of questions into two groups: G1 and G2, where G1 is the question group containing questions only related to the first lecture, and G2 is the question group containing questions related to the second lecture. Then we investigated how performance improved. From the result shown in Fig. 2 in Sect. 3, the degree of improvement of +2 shows the proportion of 6.6% for G1 and that of 16.3% for G2. This means that the proportion of students who incorrectly answered a particular question in the first exam and correctly answered the same question in the second exam is higher for the questions in G2 than those in G1. As a result, a significant difference was obtained from the entire cross-tabulation. Thus, we consider that learners’ understanding was promoted by the use of vocal-tract models in the second lecture (Lecture 2).

Strictly speaking, it is still not clear whether the second lecture improves the performance if we conduct the same contents in Lecture 2 without using vocal-tract models. Therefore, we will examine the results comparing G1–G2a and G1–G2b. The degree of improvement of +2 shows the proportion of 6.6% for G1, 8.8% for G2a, and 20.0% for G2b. This means that the proportion of students who incorrectly answered a particular question in the first exam and correctly answered the same question in the second exam is higher for the questions in G2b than those in G2a. As a result, a significant difference was obtained in G2b compared with G1, whereas only a marginal difference was obtained in G2a compared with G1. Thus, even though an additional lecture was conducted, students’ understanding did not increase much for questions in G2a, which did not directly correspond to the explanations using vocal-tract models. On the other hand, the degree of improvement did increase remarkably for questions in G2b, directly corresponding to lecture content related to the vocal-tract models. Thus, in our consideration, this proves that improvement in test results was due to using the vocal-
tract models and not simply to the fact that there was an additional lecture.

Through this kind of analysis, we can evaluate the educational effectiveness, even if the whole class listens to the lectures, without being divided into two groups. In this case, all participants in a lecture are treated uniformly, and inequity can thus be minimized. Please note, however, there is an assumption that the influences on the learning effect among question groups are uniform.

5. SUMMARY

In this paper, we described vocal-tract models and their application in education in acoustics. We have shown that the vocal-tract models are not only useful for education of sound itself, but also for pronunciation training and phonetic education. In addition, we have shown that it is important to use the vocal-tract models properly, depending on one’s goals, and we noted that we need to develop many different variations of models because of the many potential uses. In particular, we mentioned the urgent need to develop vocal-tract models for sounds that are acquired only with difficulty and delay for speech/language disorders.

We are currently collecting feedback and comments on the vocal-tract models, including anticipated improvements, from collaborators at educational institutions around the world. We also collaborate with museums to use them at science workshops as a part of outreach activities (e.g., [16]) and to install new exhibitions (Okinawa Zoo & Museum in Japan, Swiss Science Center Technorama outside Japan, etc.). We will continue to spread our wings and develop more models with varied characteristics for a wider range of applications.

ACKNOWLEDGMENTS

We would like to thank people involved in the project of developing the vocal-tract models who are using them for education and in various other ways; especially, Kikuo Maekawa, Hirokazu Sato, Shigeto Kawahara, Noriko Umeda, John J. Ohala, Kenneth N. Stevens, Terri Lander, members in Arai Laboratory, and faculty members and students at Sophia University. This work was partially supported by JSPS KAKENHI Grant Number 24501063.

REFERENCES


