Physical Models of the Human Vocal Tract with Gel-type Material

Takayuki Arai

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

Abstract

Beginning in 2001, we have been developing models of the vocal tract to promote a more intuitive understanding of the theories for speech science for technical and non-technical students. In this paper, we compared and contrasted four newer models of the talking heads: our original model with a gel-type tongue, a similar model including teeth and palate, a third model with teeth and palate having a default low tongue height, and a fourth ultra-malleable model made completely of gel material. Results are discussed regarding their strengths and weaknesses for educational purposes and articulatory training in speech pathology and language learning.

Index Terms: speech production, vocal-tract model, flexible tongue, gel-type material

1. Introduction

Creating mechanical models of the human vocal tract is not new; von Kempelen made an early model in the 18th century [1]. In Arai (2001) [2], we developed physical models of the human vocal tract, as shown in Fig. 1, based on Chiba and Kajiyama [3]. These models have been useful for instructing a wide range of students. They have been used in lectures to college and graduate students in the fields of engineering, linguistics and speech and hearing sciences, as well as in demonstrations for elementary through high school students. Through use, we have confirmed their effectiveness as educational tools [2,4]. Since 2001, we have made additional lung and head-shaped models [4-10]. By combining these models, we can help learners to understand the total speech production system [4]. We also attempted to apply of our models to voice training and/or articulatory training such as those in speech pathology and language learning [4].

Arai's models [2] were based on Chiba and Kajiyama's measurements [3]. The design of the two types of vocal-tract models was based on the measured area function of the human vocal tract. One model is the cylinder-type model in Fig. 1 (left), and the other is the plate-type model in Fig. 1 (right). Arai's models are effective education tools, especially for teaching basic speech science concepts, such as source-filter theory and the relation between vocal-tract shape and vowel quality.

To visualize how the vocal tract is positioned in the head, we designed the head-shaped models (talking heads) with a fixed tongue [4,10] as shown in Fig. 2 (left). In these "fixed-tongue" models, the vocal-tract shape is fixed, so the target sound is easily obtained, but the model must be completely reconfigured in order to produce different vowels. Another set of head-shaped models are the "flexible tongue" models [4,11] shown in Fig. 2 (right).

In both the fixed-tongue and flexible-tongue models the shape of the tongue can be seen from the outside through the transparent materials of the model. Additionally, both models can produce actual sounds; they were designed to achieve three-dimensional area functions for different vowels.

In this paper, we develop several new versions of the original flexible-tongue model (Model 1). Shown in Figures 4 and 5, our new models (Models 2 and 3) have teeth and a tongue, which is shaped differently than that of Model 1. Model 4 is a flexible tube model, the entire vocal tract of which was made from a flexible (gel-type) material. Unlike other "flexible-tongue" models, such as [12], the tongues in Models 1-4 are manipulable by hand.

Figure 1: *Arai's models: cylinder type (left) and plate type (right)* [2,4]
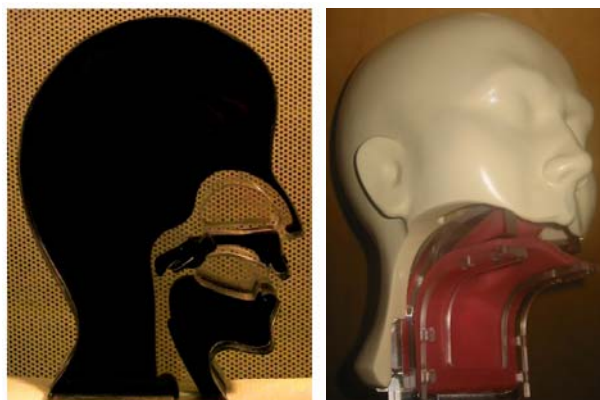
Figure 2: *Talking heads with the fixed tongue (left) [4,10] and the flexible tongue [4,11]*

2. Different models with gel-type material

In this section, we describe four different physical models (Models 1-4) of the human vocal tract made with a gel-type material. All of them can be installed in the talking heads. The gel material used was a polyethylene-styrene copolymer. The material was 4 in ASKER-C hardness. This hardness was used because it was flexible and easy to manipulate.

The entire body of the tongue is not necessarily reproduced in these models. Because we wanted to be able to manipulate the tongue models, we reproduced the upper body of the tongue, i.e., the tongue tip, the upper part of the tongue body including the tongue dorsum, and the surface of the
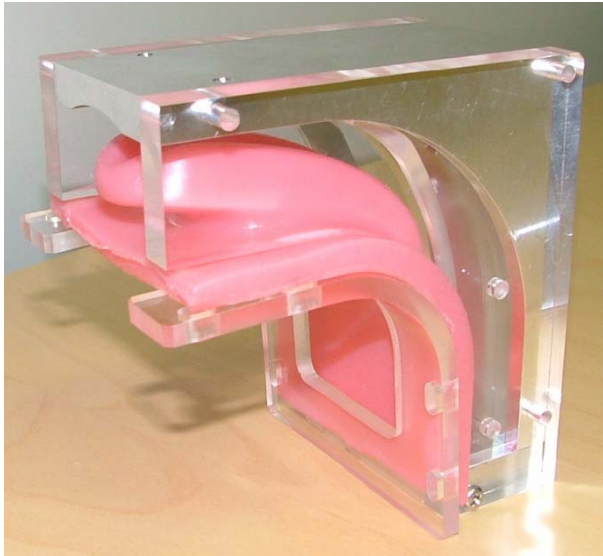
Figure 3: *Model 1 [11]*
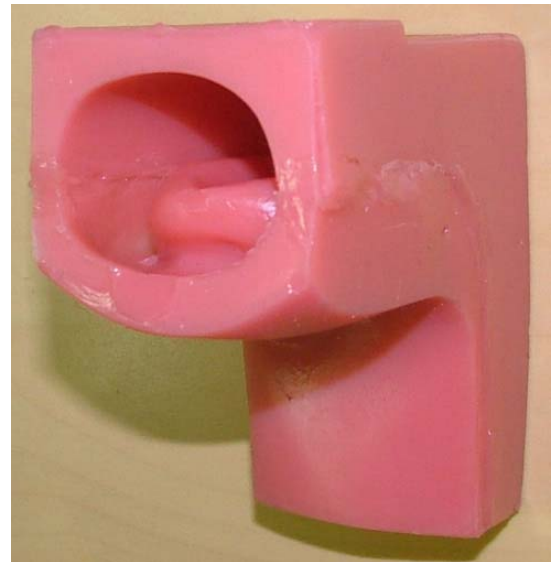

Figure 5: *Model 3*

Figure 4: *Model 2*

Figure 6: *Model 4*

tongue root. Thus, one can change the shape of the tongue manually by manipulating it from the bottom of the tongue, so that the constriction in the vocal tract can easily be moved from front to back and high to low.

2.1. Original model (Model 1) [11]

Model 1 [11] shown in Fig. 3 consists of a gel portion, a transparent outer shell with acrylic plates, and a non-transparent metal portion. The gel portion corresponds to the tongue, the bottom of the oral cavity, and the anterior pharyngeal wall. The transparent portion corresponds to the sides of the oral cavity. The non-transparent metal portion corresponds to the palate and the posterior pharyngeal wall.

There is a semicircular groove (approximately 13 mm in diameter) on the top surface of the tongue. With this groove we can easily create a narrow constriction approximately 50 mm long between the groove and the palate, facilitating the production of the vowel /i/.

2.2. Model 2

Model 2, shown in Fig. 4, consists of a gel portion, (again corresponding to the tongue, the bottom of the oral cavity and the anterior pharyngeal wall), a portion with teeth and palate, and a transparent, acrylic portion.

Unlike Model 1, this model has lower and upper teeth with palate. The rest of the model is transparent, representing the sides of the oral cavity and the posterior pharyngeal wall. The transparent resin material allows us to examine the tongue position not only from the sides but from the rear. The tongue also has a semicircular groove on the top surface, as in Model 1.

2.3. Model 3

Model 3, shown in Fig. 5, consists of a gel portion, a set of teeth with palate, and a transparent portion made of acrylic material, as was seen in Model 2. Like Models 1 and 2, the gel portion corresponds to the tongue, the bottom of the oral cavity, and the anterior pharyngeal wall. As with Model 2, Model 3 has lower and upper teeth with palate, and a transparent portion representing both sides of the oral cavity and the posterior pharyngeal wall. The gel portion corresponding to the bottom of the oral cavity is similar to that of Model 1, except that the default tongue height is low. The semicircular groove on the top surface of the tongue remains similar to Models 1 and 2.

2.4. Model 4

Model 4, shown in Fig. 6, is different from Models 1-3, being made entirely out of gel material. Because the gel material is not transparent, the tongue position is not visible from the sides, but one can examine tongue position through the mouth opening. The strength of this model is that its entire body is flexible, so that one can drastically change the shape of the vocal tract.

3. Discussion

First, let us review how the models are manipulated when /i/, /a/ and /u/ are produced [11]. With /i/, the tongue attaches to the hard palate. The vowel /i/ is produced when the tongue is in the high position, where the groove formed between the top surface of the tongue and the palate forms a long (approximately 50 mm) narrow constriction. For high-front vowels, the sound wave propagates through the narrow constriction and should not leak out of the sides of the tongue. To achieve this effect, the ridges are designed to close the lateral pathways in all models. When the degree of constriction is less severe than vowel /i/, /e/ is produced. In the case of /a/, the tongue body is lowered whereas the surface of the tongue root is pushed backward. For /u/, one can push the middle of the tongue back and up toward the curve of the vocal tract.

Each model produced similar vowel sounds when their tongue positions were the same. Figure 7 shows a spectrogram of an utterance of three concatenated vowels: /a/, /i/, and /u/, produced by Model 1. In this recording, a whistle-type artificial larynx [4] was used as a sound source. The microphone was placed approximately 20 cm in front of the model's lip end, and the resulting sounds were digitally recorded with a sampling frequency of 48 kHz by a linear PCM recorder (PCM-D1, Sony). The signals were downsampled to 8 kHz. The spectrogram was obtained by Wavesurfer. As shown in this figure, the lower formants are clearly observed, and vowel quality is also clear when one listens to each vowel.

Although we observed clear formants as well as natural formant transitions from one vowel to another from the models, each model has strengths and weaknesses from an educational point of view.

Model 1 is the simplest model. It has the highest manipulability and high-quality sounds. However, the tongue is not visible from the back because the rear panel is not transparent.

Models 2 and 3, on the other hand, enable examination of the tongue from the back because the rear portion is made of transparent acrylic material. Furthermore, these models have teeth, so they appear more real. However, making a more life-like model does not necessarily guarantee a better design for education; it depends on one's instructional goals. In general, simplification is more effective because it allows one to emphasize what is important and ignore what is irrelevant. If one's focus is teaching vowel production, attaching teeth is optional. However, if the focus is teaching speech production, the teeth in the model are meaningful because they enable one to produce more realistic sounds, including consonants.

One of the major differences between Models 2 and 3 is the default tongue height. Model 3 has a much lower default tongue position. We expected Model 3 to be more flexible because there is more room in the oral cavity; however, it turns out that the tongue is not as movable as we expected. This is probably due to the long distance between the tongue and the palate.

Being made entirely of gel material makes Model 4 useful when one wishes to change the shape of the vocal tract and express a highly dynamic speech sound. Unfortunately the very characteristic that enables the formation of such dynamic sounds results in the downside of this model, that the tongue position is not visible from the sides but only through the mouth opening. As with Models 1-3, Model 4 can produce vowel sounds. But because of its flexibility, one is also able to produce sonorants such as /l/ more easily than with the other models.

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 [4]. It is especially important to show patients that a slight shift of articulators may result in a huge acoustic difference. Our models can be useful in a clinical situation for adults and children who have speech disorders, because they can learn more about phonation and articulation by seeing, touching, and playing with the vocal tract models of the talking heads.

The proposed models of the talking heads 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. This is not only true for vowels but also for consonants. For instance, native speakers of Japanese often have difficulty pronouncing /r/ and /l/. Showing them the ideal shape of the vocal tract should help learners to configure their own articulators to produce these sounds.

4. Conclusions

The author teaches acoustics at Sophia University not only to technical students but also to students majoring in Linguistics, Psychology, and Speech Pathology. These models need to meet the educational needs of the non-technical as well as technical students. One of the benefits of the models is that
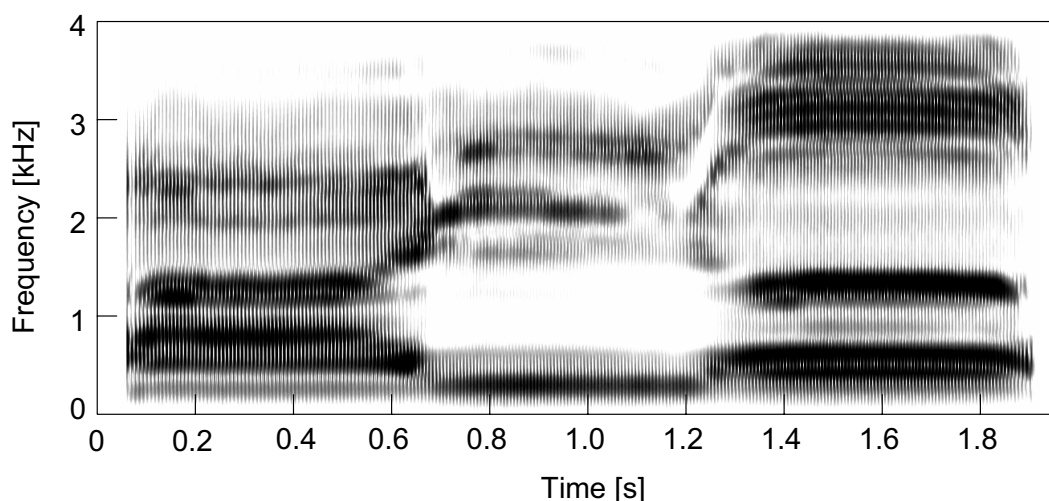
Figure 7: *A spectrogram of an utterance of three concatenated vowels: /a/, /i/, and /u/, produced by Model 1.*

students are able to see and hear the changes taking place in the vocal tract as various sounds are produced in real time. The models are valuable when measuring how sensitive sound quality is to changes in configuration. From our teaching experience for non-technical students, we confirmed that these models are particularly useful when teaching the dynamic aspects of speech production. Furthermore, these models are attractive for children. A science workshop is one of the appropriate places where we can demonstrate the model effectively for them. In any case, a learner can manipulate the tongue by hand, so that he or 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.

These models can also be used for training of the Tadoma method of speech communication [13], in which a deaf-blind receiver places a hand over the face and neck of the talker and uses tactual input to monitor various articulatory actions associated with speech.

5. Acknowledgements

6. References

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

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

[3]   Chiba, T. and Kajiyama, M., "The Vowel: Its Nature and Structure," Tokyo-Kaiseikan Pub. Co., Ltd., Tokyo, 1942.

[4]   Arai, T., "Education system in acoustics of speech production using physical models of the human vocal tract," Acoust. Sci. Tech., 28(3):190-201, 2007.

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

[6]   Arai, T., Maeda, E., Saika, N. and Murahara, Y., "Physical models of the human vocal tract as tools for education in acoustics," Proc. Pan-American/Iberian Meeting on Acoustics, Cancun, 2002.

[7]   Arai, T., "Physical and computer-based tools for teaching Phonetics," Proc. International Congress of Phonetic Sciences, 1:305-308, 2003.

[8]   Arai, T., "Education in Acoustics using physical models of the human vocal tract," Proc. International Congress on Acoustics, 3:1969-1972, 2004.

[9]   Arai, T., "Visualizing vowel-production mechanism using simple education tools," J. Acoust. Soc. Am., 118(3), Pt. 2, p. 1862, 2005.

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

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

[12]   Fukui, K., Ishikawa, Y., Sawa, T., Shintaku, E., Honda, M., and Takanishi, A., "New anthropomorphic talking robot having a three-dimensional articulation mechanism and improved pitch range," Proc. IEEE Int'l Conf. on Robotics and Automation, 2922-2927, 2007.

[13]   Reed, C.M., Rabinowitz, W.M., Durlach, N.I., Delhorne, L.A., Braida, L.D., Pemberton, J.C., Mulcahey, B.D., and Washington, D.L., "Analytic study of the Tadoma method: Improving performance through the use of supplementary tactual displays," J. Speech Hear. Res., 35:450-465, 1992.