INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Two different mechanisms of movable mandible for vocal-tract model with flexible tongue

Takayuki Arai

Sophia University, Tokyo, Japan
arai@sophia.ac.jp

Abstract

In 2017 and 2018, two types of vocal-tract models with physical materials were developed that resemble anatomical models and can physically produce human-like speech sounds. The 2017 model is a static-model, and its vocal-tract configuration is set to produce the vowel /a/. The 2018 model is a dynamic-model, and portions of the articulators including the top surface of the tongue are made of a gel-type material. This allows a user to manipulate the shape of the tongue and articulate different vowels and a certain set of consonants. However, the mandible of the model is fixed, making it difficult to manipulate different sounds with different jaw openings, such as high vs. low vowels. Therefore, in 2019, two types were developed by adding an additional mandible mechanism to the 2018 model. For the first type, the mandible was designed to move between the open and closed positions by creating an arc-shape rail. For the second type, the mandible moves the same trajectory with an additional support. As a result, various speech sounds with a flexible-tongue and moveable mandible can be easily produced. These models are more realistic than the anatomical models proposed in 2017 and 2018 in terms of articulatory movements.

Index Terms: vocal-tract model, movable mandible, flexible tongue, anatomical-type model

1. Introduction

A series of vocal-tract models were previously proposed that are useful for pedagogical and clinical purposes (e.g., [1-3]). They are not only used for education in acoustics and speech science but also for pronunciation training as well as basic research applications. During the process of vocal-organ modeling, an issue of simplification became a key issue. Sometimes, simplifying the mechanisms of vowel production for teaching purposes is necessary, such as the sliding three-tube model [2], for example. Such a model is straight, the cross-sectional shape of the outer tube is a circle, and the major degree of freedom is only the location of the tongue constriction. However, the two most recent models have different simplification strategies.

In 2017, the first anatomical model was designed [4]. The goals with the 2017 model were as follows: 1) to produce an intelligible /a/ vowel; 2) to resemble an anatomical model, so that students can learn at least the rough positions of each speech organ, such as lips, teeth, tongue, mandible, uvula; and 3) to have the speech organs be partially visible from the outside. For 1), the mandible was opened, and the tongue was placed at the back position to form a vocal-tract configuration when producing /a/. For 2), simplified versions of the speech organs were designed. Each speech organ was also colored. Finally, for 3), windows were created on the left and right cheeks and the posterior pharyngeal wall. Figure 1(a) shows the complete version of the 2017 model.

Another anatomical model was designed in 2018 [5]. An additional goal with the 2018 model to those with the 2017 model was as follows: 4) to have the lower articulators be movable (except mandible), so that the model can produce different types of sounds. To achieve 4), a part of the tongue was made of gel material. The shape of the tongue was an abstraction of an actual tongue, as with the 2017 model, but the top surface of the tongue, the bottom of the oral cavity, and anterior pharyngeal wall were also made of gel material. The shape of the pharyngeal cavity can be changed by pushing the tongue root towards back to create low vowels. The tongue can be pushed up against the palate for creating high vowels. Figure 1(b) shows the 2018 model.

The 2018 model was successful in terms of the goals described above. However, there was a problem in that the height of the tongue has to be changed with a fixed jaw opening. It is more natural to produce low vowels with an open jaw and high vowels with a closed jaw. This is true in terms of the manipulability and their acoustics. Therefore, the 2018 model was improved and called the 2019a and 2019b models (right two models in Fig. 1) with the following additional goal: 5) to have a mandible also be movable, so that the model can produce low vowels with an open jaw and high vowels with a closed jaw. These models were evaluated for this study.

2. 2019 Models

The design of the 2019 models is based on the 2018 model, so that the neutral/resting positions of these models are fundamentally the same. To achieve additional goal 5) in the previous section, the 2019a and 2019b models were designed.

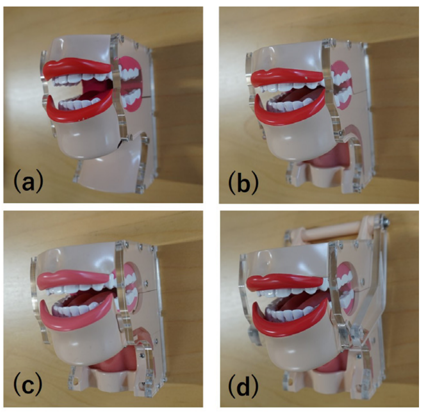

Figure 1: *(a) 2017 model, (b) 2018 model, (c) 2019a model and (d) 2019b model.*

http://dx.doi.org/10.21437/Interspeech.2020-1159

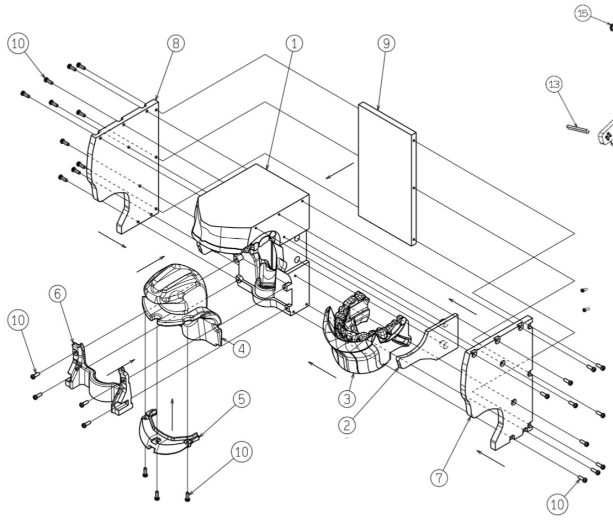

Figure 2: *Design of 2019a model.*

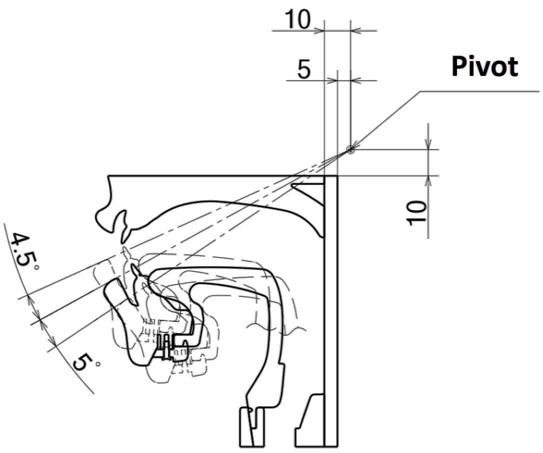

Figure 3: *The pivot of movable mandible is located at slight off position of upper corner of 2019a model.*

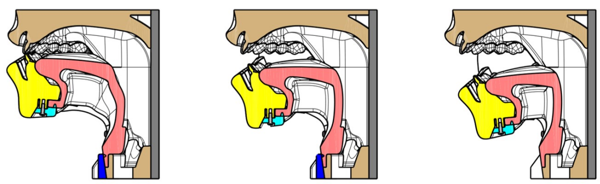

Figure 4: The *angle of mandible changes from -4.5 degrees (left) through +5.0 degrees (right) via neutral position (center).*

2.1. 2019a model

Figure 2 shows the assembly drawing of the 2019a model. In this figure, the fixed upper jaw is Part No. 1. To make the inside of the oral and pharyngeal cavities visible, windows were designed with transparent acrylic plates (Part Nos. 7-9), as in the 2017 and 2018 models. The flexible tongue (Part No. 4) also plays an important role in this model, as in the 2018 model, so that goal 4) is still targeted.

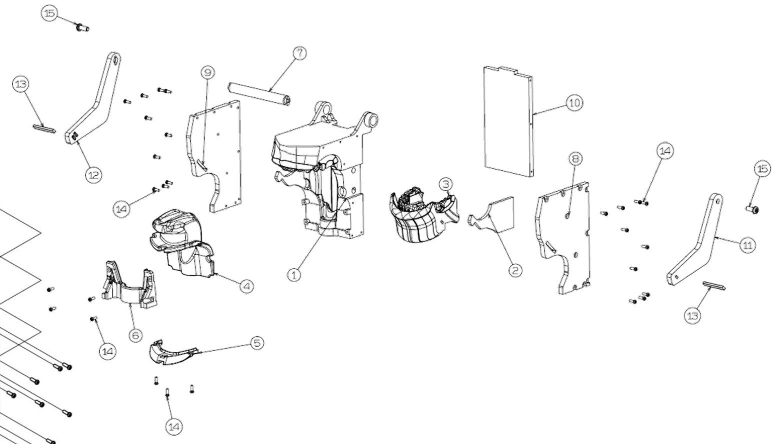

Figure 5: *Design of 2019b model.*

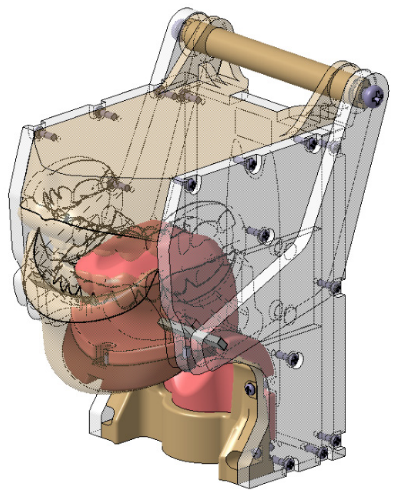

Figure 6: *2019b model has the outer support around the pivot.*

The main difference from the 2018 model is the movable mandible, Part No. 3. The mandible was designed to move up and down by moving it along a rail. The rail is located at Part No. 2 in Fig. 2. The curvature of the rail is an arc with a circular trajectory. The pivot shown in Fig. 3 is based on the human anatomy [6]. The rotation angle of the mandible is -4.5 degrees and +5.0 degrees from the neutral/resting position (Fig. 4).

2.2. 2019b model

Figure 5 shows the assembly drawing of the 2019b model. The main difference from the 2019a model is the outer support (Part No. 11 in Fig. 5). This support rotates around the pivot, of which the location is exactly the same as in Fig. 3. The actual pivot is Part No. 7 in Fig. 5 of the 2019b model. Figure 6 shows the outer support, which can be rotated around the pivot.

2.3. Completed models

The completed models are shown in Fig. 1(c) and 1(d). The models look like an "anatomical model," so that the human anatomy related to speech production can be easily explained with the rough positions of each speech organ, such as lips, teeth, tongue, mandible, and uvula (goal 2). The speech organs are partially visible from the outside (goal 3). Furthermore, different speech sounds can be produced with the flexible tongue and movable mandible. The flexible tongue is of a polyethylene-styrene copolymer, the same material used in previous studies [5, 7] (two degrees of hardness was tested: 2 and 4 in ASKER-C hardness). Figure 7 shows photos when vowels /i/, /e/, and /a/ are produced.

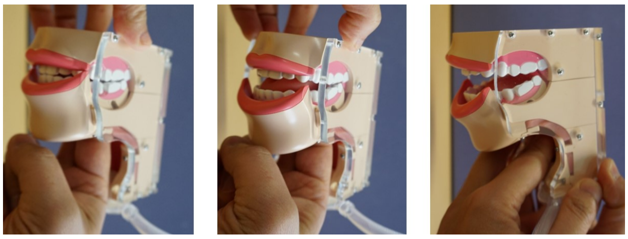

Figure 7: *2019a model produces three vowels: /i/, /e/, and /a/.*

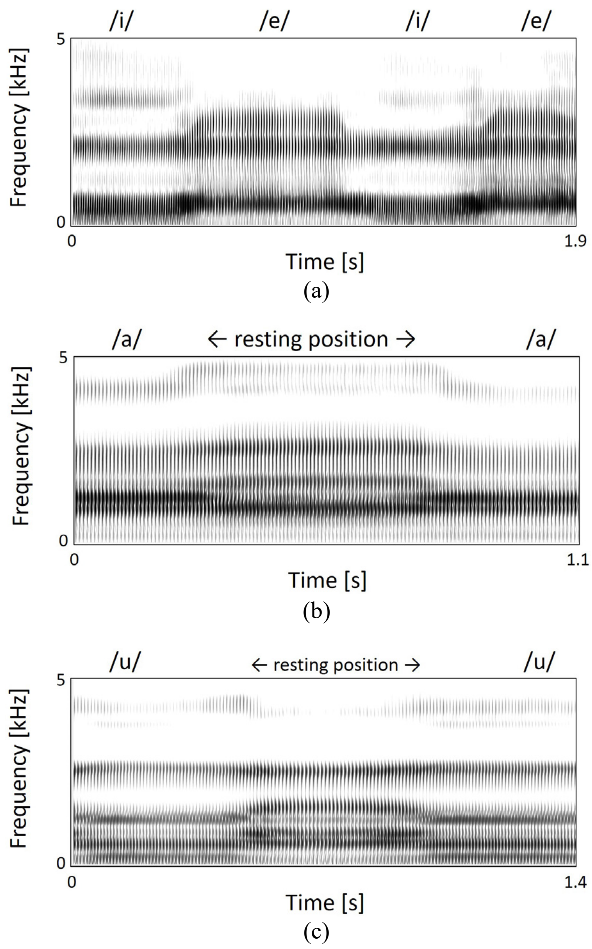

Figure 8: *Spectrogram of output sounds recorded with 2019a model and whistle-type artificial larynx.*

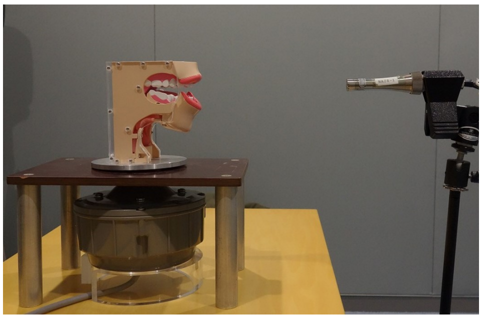

Figure 9: *Recordings for 2019a model.*

3. Acoustic analysis

3.1. Five vowels

Output sounds produced with the 2019a model were recorded. The sample positions of the articulators are shown in Fig. 7. A whistle-type artificial larynx was used as a sound source. The source was fed into a hole at the glottal end of the 2019a model. Output sounds were recorded using an audio interface (Roland, Rubix 44) via a microphone (Rode, NT6).

The spectrographic representations of different configurations are shown in Fig. 8. Figure 8(a) was the case when the vocal tract was set to the configuration of /i/ in Fig. 7 (left) and /e/ in Fig. 7 (middle). Figure 8(b) was the case when the vocal tract was set to the configuration of /a/ in Fig. 7 (right) and its resting position. Figure 8(c) was the case when the vocal tract was set to the configuration of (unrounded) /u/ and its resting position. The formant frequencies well describe the difference in vowel qualities.

3.2. Impulse responses

The impulse responses of the 2018 and 2019a models with their neutral vocal-tract configurations were measured. The input signal for this measurement was a swept-sine signal with a sampling frequency of 48 kHz and the length of this signal was 65536 samples. The input signal was fed into the driver unit (TOA, TU-750) via an audio interface (RME, FireFace UC) and power amplifier (Sony, TA-V777ES) as shown in Fig. 9.

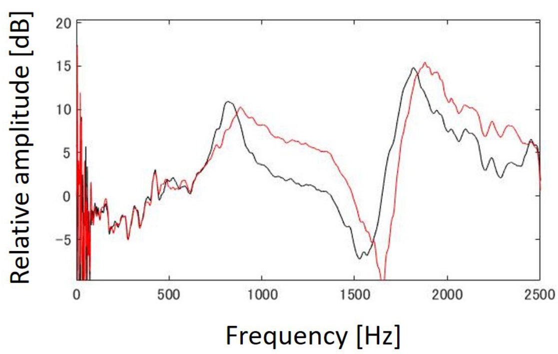

Figure 10: *Spectra computed from impulse responses. Black: 2018 model. Red: 2019a model. Their vocal-tract configurations were at the neutral/resting positions.*

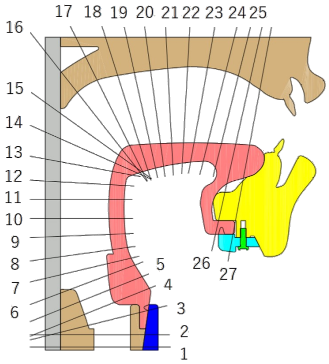

Figure 11: *Mid-sagittal cross-sectional view of 2019 model with lines for measurements of cross-sectional areas as function of distance from glottis along midline of vocal tract. The vocal-tract configuration was at the neutral/resting position.*

To avoid unwanted coupling between the neck and area behind the neck of the driver unit and achieve high impedance at the glottis end, a close-fitting metal cylindrical filler was inserted inside the neck, and a hole was made in the center of the metal filling with an area of 0.13 cm^2. The glottis end of the vocal-tract model as placed on top of a thin metal plate. The output sounds were recorded using a microphone of the sound-level meter (RION, NA-28) and an audio interface (RME, FireFace UC) with a sampling frequency of 48 kHz. The microphone was placed approximately 15 cm in front of the output end in a sound-treated room (Fig. 9). The signals recorded were synchronously averaged multiple times to obtain a higher signal-to-noise ratio.

Figure 10 shows the spectra computed from the impulse responses of the two models.

4. Discussion

As described above, the 2019 model is an extension of the 2018 model. The 2018 model [5] has a gel-type flexible tongue, as in [7], as well as the 2019 model, so that they are able to produce different vowels with different vocal-tract configurations. The 2018 model has a mandible, but it is fixed and the degree of the jaw opening is constant. For the 2019 model, however, the jaw opening is changeable with a movable mandible, and the degree of the jaw opening can be adjusted depending on the vowel height. This allows a user better manipulability. Furthermore, the degree of the mouth opening produces more realistic acoustics, such as the contrast between open vs. closed vowels.

To achieve the open and close movements of the mandible, two parts (Part Nos. 5 and 6 in Figs. 2 and 5) are key in the 2019 models. For the 2018 model, the corresponding parts is a single piece. This part is necessary to fix the flexible tongue to the underneath of the oral cavity. When the mandible becomes movable, this part has to become two; upper and lower. These two parts fix the flexible tongue to the mandible in the 2019 models and are a little bit shorter, so as not to interfere the upper and lower parts with each other, especially when they are close at positive angles of the mandible.

The mandible movement with the 2019a model is currently not very smooth. This is because the movement is done along the arc-shaped rail (Part No. 2 in Fig. 2). On the other hand, because of the outer support, the 2019b model was able to achieve better movements. In previous studies, the mandibles were mostly directly connected to the pivot [8-10]. Such an “arm” between the mandible and pivot allow smooth movement of the mandible. Further discussions on the temporomandibular joints can be found in the study by Brady [10].

Table 1: *Cross-sectional areas as function of distance from glottis along midline of vocal tract with their neutral/resting vocal-tract configurations.*

No.	Area (cm²) 2018 model	Area (cm²) 2019a model	No.	Area (cm²) 2018 model	Area (cm²) 2019a model
1	1.13	1.13	15	5.84	5.84
2	1.13	1.13	16	6.15	6.15
3	1.53	1.53	17	5.41	5.41
4	1.48	1.48	18	5.29	6.91
5	1.67	1.67	19	5.43	5.84
6	3.29	3.29	20	6.01	6.01
7	3.75	3.75	21	7.49	9.1
8	4.23	4.23	22	8.53	10.52
9	4.16	4.16	23	8.94	11.11
10	4.07	4.07	24	8.94	11.19
11	4.06	4.06	25	8.55	10.79
12	4.16	4.16	26	8.22	10.31
13	4.43	4.43	27	8.68	10.45
14	4.97	4.97			

There is always a risk that the more complicated models become, less intelligible sounds are produced. This is generally due to the acoustic leakage caused by complicated mechanisms. To test this, the two spectra of the 2018 and 2019a models computed from the impulse responses measured in Section 3.2 were compared. In general, the similarity between the spectra of the two models shown in Fig. 10 was found. The difference between them is probably due to the cross-sectional area functions of their resting positions. Table 1 shows the cross-sectional areas as function of the distance from the glottis along the midline of the vocal-tract configuration in the 2018 and 2019a models at neutral/resting position. Figure 11 shows a mid-sagittal view of the model with the lines for the measurements. There was a dip at around 1600 Hz in the spectra of the models in Fig. 10, and this might indicate that sound leaked from the cavities.

5. Conclusions

An anatomical model with a flexible tongue and movable mandible was proposed. The previous models, [5] and [7] also have flexible tongues. However, with the proposed model, the jaw opening can be adjusted with the movable mandible. Like the previous models, this model can be used as an educational tool not only for acoustics, acoustic phonetics and speech science, but also for pronunciation training for language learners as well as patients. Designing a speaking robot [8,11] based on the mechanisms discussed in this study is for future work. On the other hand, investigating the relation between the mandible and the larynx [e.g., 12] is also going to be our future work.

6. Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Number 18K02988 and Sophia University Special Grant for Academic Research (Research in Priority Areas).

7. References

[1] T. Arai, “Education system in acoustics of speech production using physical models of the human vocal tract,” *Acoust. Sci. Tech*., vol. 28, no. 3, pp. 190–201, 2007.

[2] T. Arai, “Education in acoustics and speech science using vocal-tract models,” *J. Acoust. Soc. Am.*, vol. 131, no. 3, Pt. 2, pp. 2444–2454, 2012.

[3] T. Arai, “Vocal-tract models and their applications in education for intuitive understanding of speech production,” *Acoust. Sci. Tech*., vol. 37, no. 4, pp. 148–156, 2016.

[4] T. Arai, “Vocal-tract model with static articulators: Lips, teeth, tongue, and more,” *Proc. of INTERSPEECH*, pp. 4028–4029, 2017.

[5] T. Arai, “Flexible tongue housed in a static model of the vocal tract with jaws, lips and teeth,” *Proc. of INTERSPEECH*, pp. 171–172, 2018.

[6] S. Standring, *Gray's Anatomy: The Anatomical Basis of Clinical Practice*, Elsevier, 2015.

[7] T. Arai, “Gel-type tongue for a physical model of the human vocal tract as an educational tool in acoustics of speech production,” *Acoust. Sci. Tech*., vol. 29, no. 2, pp. 188–190, 2008.

[8] K. Fukui, T. Kusano, Y. Mukaeda, Y. Suzuki, A. Takanishi and M. Honda, “Speech robot mimicking human articulatory motion,” *Proc. of INTERSPEECH*, pp. 1021–1024, 2010.

[9] M. C. Brady, “Prosodic timing analysis for articulatory re-synthesis using a bank of resonators with an adaptive oscillator,” *Proc. of INTERSPEECH*, pp. 1029–1032, 2010.

[10] R. Hofe, “Biomimetic vocal tract modelling: An artificial speaker to investigate the energetics of speech production,” Ph.D. Dissertation, the University of Sheffield, U.K., 2011.

[11] T. Mochida, M. Honda, K. Hayashi, T. Kuwae, K. Tanahashi, K. Nishikawa and A. Takanishi, “Control system for talking robot to replicate articulatory movement of natural speech,” *Proc. of INTERSPEECH*, pp. 1533–1536, 2002.

[12] J. R. Westbury, “Mandible and hyoid bone movements during speech,” J. Speech Hear. Res., vol. 31, pp. 405-416, 1988.