Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Education in Acoustics
Session 2pED: Teaching Methods in Acoustics

2pED3. Mechanical bent-type models of the human vocal tract consisting of blocks

Takayuki Arai*

***Corresponding author's address: Dept. of Information and Communication Sciences, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, 102-8554, Tokyo, Japan, arai@sophia.ac.jp**

In our previous work, we developed several physical models of the human vocal tract and reported that they are intuitive and helpful for students studying acoustics and speech science. Models with a bent vocal tract can achieve relatively realistic tongue movements. These bent-type models had either a flexible tongue or a sliding tongue. In the former case, the tongue was made of a flexible gel-type material so that we could form arbitrary tongue shapes. However, this flexibility meant that training is needed to achieve target sounds. In the latter case, the tongue was made of an acrylic resin, and only a limited number of vowel sounds can be achieved because so few sliding parts are available to change the tongue shape. Therefore, in this study, we redesigned the mechanical bent-type models so that they now consist of blocks. By placing the blocks at the proper positions, the block-type model can produce five intelligible Japanese vowels. We also designed a single bent-type model with sliding blocks that can produce several vowel sounds. [This work was partially supported by a Grant-in-Aid for Scientific Research (24501063) from the Japan Society for the Promotion of Science.]

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4800980]
Received 8 Jan 2013; published 2 Jun 2013

INTRODUCTION

Several physical models of the human vocal tract have been developed for educational purposes in our previous studies, and they have proven to be intuitive and helpful for students in acoustics and speech science[1-4]. Collecting different types of vocal-tract models is important because each addresses a different set of needs. All of our models demonstrate 1) the relationship between vocal-tract configuration and vowel quality, and 2) the source-filter theory of speech production[5]. Even with the connected-tube (CT) model[2], one of the simplest types, we can demonstrate these two points. However, if we want to teach about a third concept, that of tongue position and tongue movement, we need bent and dynamic models of the vocal tract.

One of the earlier dynamic models is the von Kempelen's machine with hand varied resonator[6]. A recent talking robot can also change the vocal tract configuration dynamically[7]. We have also developed several dynamic models of the human vocal tract for educational purposes: the sliding-three-tube (S3T) model[2,8], Umeda and Teranishi's computer-controlled model[9,10], the gel-type tongue model[3,4], and the head-shaped models[1,10]. The former two models mentioned have a straight vocal tract, and the area functions are roughly approximated. In these cases, we can demonstrate changing vocal-tract configuration yields different vowel qualities in real time, and learners can compare the changes in the configuration by eyes as well as the changes in the output sounds by ears. However, due to the simple designs, the dynamic movements with these models are less realistic.

On the other hand, bent vocal-tract models are suited, when we demonstrate how we produce vowels from our speech organs. One of the common situations that learners often come across is to ask the question of where the vocal tract is placed in our head. Static bent models with head shapes[1] often give us an enough answer to that question. However, our previous static bent models were limited to vowels /a/ and /i/. Therefore, the first goal of this study is to find an appropriate set of the vocal-tract configurations for static bent models for classroom demonstrations.

Although the dynamic straight models are useful for simple demonstrations, they do not directly simulate the movements of tongue as mentioned above. Therefore, we have developed dynamic bent models, such as the gel-type tongue model[3,4] and the head-shaped model with the sliding tongue[10]. With these models, relatively realistic tongue movements can be simulated. The dynamic bent models are useful when we demonstrate a tongue movement between vowels /a/ and /i/, for instance. In this case, learners can see the downward / backward tongue movement is needed for vowel /a/, and the upward / forward tongue movement for vowel /i/. Furthermore, the gel-type tongue model[3,4] has many advantages, including flexibility of the tongue, enabling one to produce many different vowels. One disadvantage of the gel-type tongue model is that it is difficult to manipulate, making it a challenge to reproduce the same configuration repeatedly. Due to this difficulty, this model is mainly used when the author demonstrates the vowel production to learners in a class or workshop, but learners are never asked to manipulate this model.

The head-shaped model with the sliding tongue[10] has the advantages of both the S3T and the gel-type tongue models. In other words, the sliding tongue model has limited degrees of freedom, so that it is simpler to produce the target vowel. In addition, the vocal tract is bent in the middle at a right angle, so that we can move the tongue more realistically. The degrees of freedom for this model are as follows: The 1st degree of freedom is the diagonal movement of the tongue; the 2nd degree of freedom is the protrusion of the tongue dorsum; the 3rd degree of freedom is lip rounding. This model was able to produce the vowel sequence between /a/ and /i/ relatively easily; however, the other vowels were hard to deal with. Therefore, the second goal of this study is to redesign the mechanical bent-type models, so that a single bent-type model with sliding blocks to cover all five Japanese vowels, /i/, /e/, /a/, /o/, and /u/.

In this study, we started with the CT model, which was originally designed with cylindrical tubes, and redesigned it using square tubes. Then, we design two bent-type models with sliding blocks: the one for front vowels and the other for back vowels. Finally, we designed a single bent-type model with sliding blocks that produces all five vowels.

BENT-TYPE MODELS WITH SLIDING BLOCKS

Two Bent-type Models with Sliding Blocks

To find an appropriate set of the vocal-tract configurations for static bent models for classroom demonstration, we designed two bent-type models with blocks. By redesigning the CT models with square tubes (see Appendix), we made bent-type models with a rectangular cross-section. The basic dimension of the cross-section was 45 mm x 20 mm for the neutral vowel, *schwa*. The length of the oral and pharyngeal cavities was 90 mm and 70 mm, respectively. There was a narrow constriction at the larynx, the length of which was 20 mm. The dimension of its cross-section was 9 mm x 9 mm. We designed two bent-type models with sliding blocks: one was for front vowels (Model A) and the other was for back vowels (Model B). In Figs. 1 (a)-(e), the left panel shows the three-dimensional representations of the vocal-tract shape (the numbers are the lengths of sections in mm along the vocal-tract length) and the right panel is a picture of the actual model (the front plate was removed in order to photograph the model).

Front Vowels (Model A)

The models shown in Figs. 1 (a) and (b) are based on the bent-type model A for front vowels. A block (shown in yellow) inserted from the floor of the oral cavity controls tongue height for the front vowels /i/ and /e/. On the top surface of the block there is a groove running along the vocal-tract length. The cross sectional dimension of the groove is 9 mm x 9 mm (the location of the groove is indicated by the red dashed line). When the surface of the block reaches the roof of the oral cavity, the area of the constriction becomes minimal and it simulates the vowel /i/ (Fig. 1a). When the block shifts 6-8 mm downwards, it simulates the vowel /e/ (Fig. 1b; in this case, the downward shift is 8 mm). From Figs. 1 (a) and (b) we can observe that the tongue constriction for /i/ and /e/ is in the same position, but the area of the constriction is wider for /e/ than for /i/.

Back Vowels (Model B)

The models shown in Figs. 1 (c)-(e) are based on the bent-type model B for back vowels. In Figs. 1 (c) and (d), the block shown in yellow was placed inside the pharyngeal cavity to control tongue height for the back vowels /a/ and /o/. There is a groove on the surface of the block facing the pharyngeal wall, the cross-sectional dimension of which is 9 mm x 9 mm, running along the length of the vocal tract. When the block is placed 5 mm above the bottom, as in Fig. 1 (c), we can produce /a/. When the block is shifted up to 11 mm and we use an additional block for lip rounding (leftmost yellow block), we can produce the vowel /o/, as shown in Fig. 1 (d). The block for lip rounding has a square hole, of which the dimension is 18 mm x 18 mm (the location of the hole is also indicated by the red dashed line). By replacing the straight sliding block for /o/ with a right-angle shaped block and positioning it at the corner of the vocal tract as shown in Fig. 1 (e) (rightmost yellow block), we can produce the vowel /u/. In this case, there is also a groove, 9 mm x 9 mm, on the surface of the block facing the pharyngeal wall and the palate. The block for lip rounding is also used in Fig. 1 (e), just as it was for /o/ in Fig. 1 (d).

A Single Bent-type Model with Sliding Blocks

Finally, we designed a single bent-type model with sliding blocks for all five vowels (Model C). The design advancement in this model is the combination of the former two bent-type models into one model. When designing the single model, we considered the following four points:

1) Front vowels have a block for tongue constriction located in the palatal region, but back vowels do not. Vertical movement of the block takes care of this.
2) Back vowels must have a block for tongue constriction in the pharyngeal cavity, but front vowels do not. Horizontal movement of the block takes care of this.
3) Diagonal protrusion of the tongue dorsum is necessary for the vowel /u/.
4) An extra block at the mouth end is necessary for lip rounding.

Figure 2 shows a picture of this single bent-type model with sliding blocks. As shown in this figure, there are four sliding blocks. Figure 3 shows the configuration for all five vowels using this single model. (The front plate was removed for the photographs in Figs. 2 and 3.)

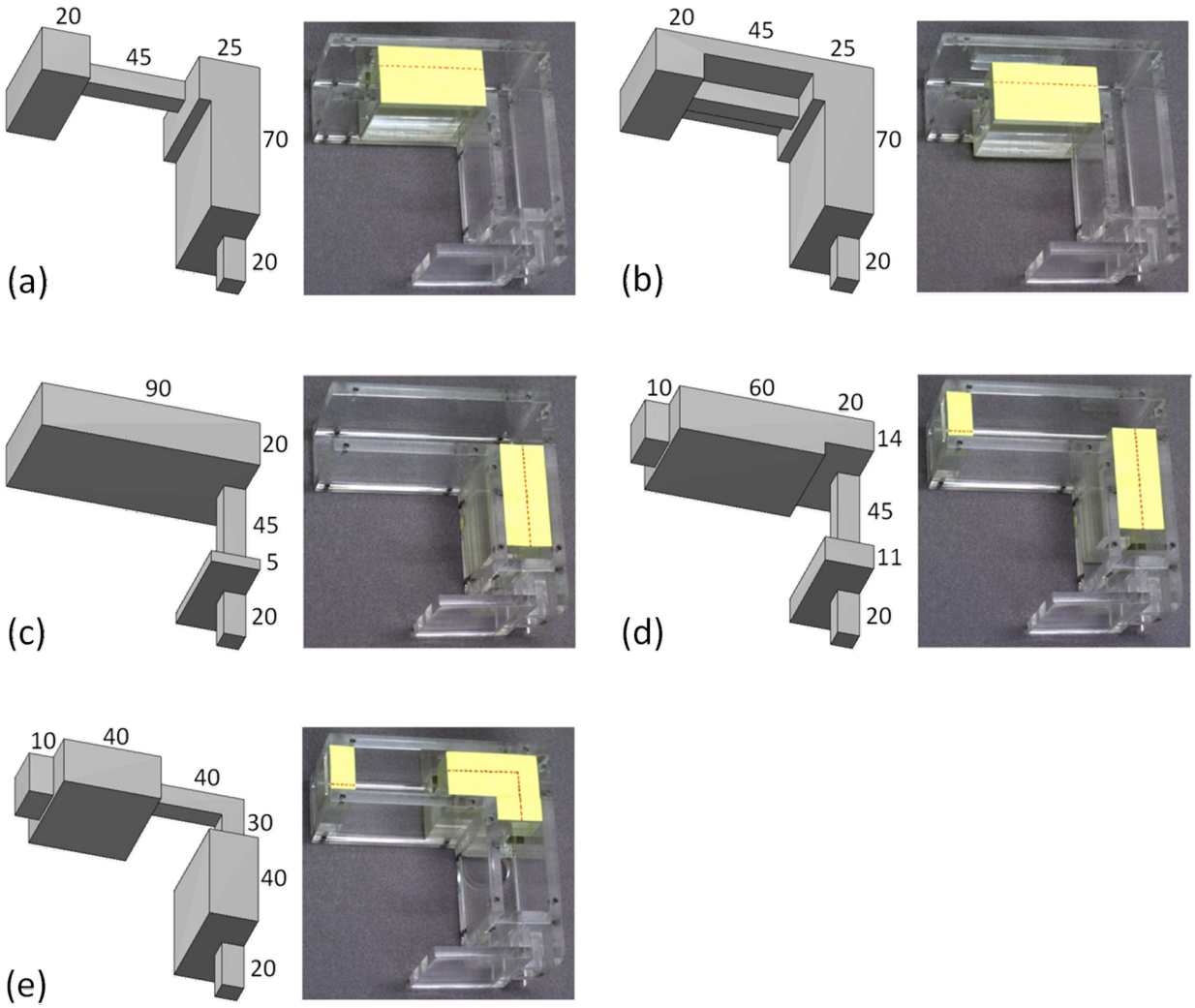

FIGURE 1. Bent-type models A (a, b) and B (c-e): (a) vowel /i/, (b) vowel /e/, (c) vowel /a/, (d) vowel /o/, and (e) vowel /u/.

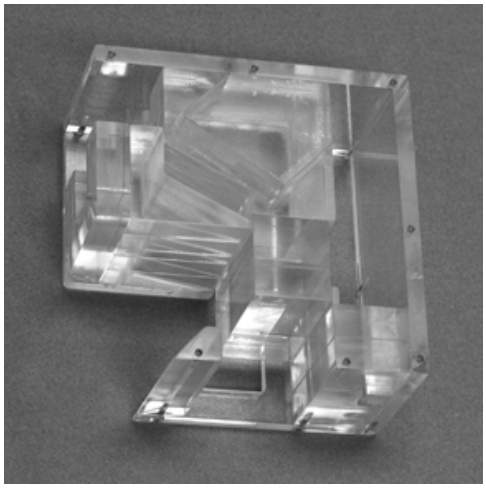

FIGURE 2. Single bent-type model with sliding blocks (Model C).

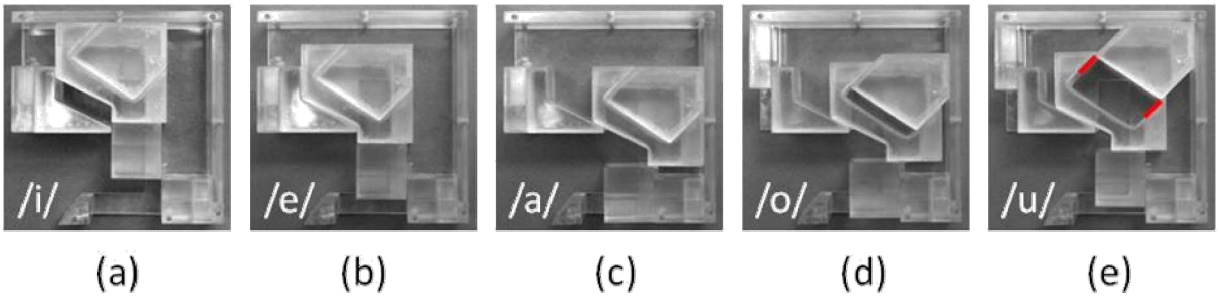

FIGURE 3. Single bent-type model (Model C) for five vowels.

MEASUREMENTS

We recorded sounds from Models A, B and C for the five Japanese vowels and analyzed the output sounds by inspecting the spectrograms. The sounds produced by these models are also used for informal listening tests.

Two Bent-type Models with Sliding Blocks (Models A and B)

A driver unit (TOA TU-750) for a horn speaker was attached to the glottis end of the model. Input signals were fed into the driver unit via an audio interface (RME Multiface) and a power amplifier (FOSTEX AP1020). There were two types of input signals. The first signal was an impulse train with an original sampling frequency of 16 kHz; later, the signals were upsampled to 48 kHz. Its fundamental frequency, f_0, increased from 100 to 125 Hz within the first 100 ms, and then decreased to 100 Hz within the next 200 ms. The total duration of this signal was 300 ms. The second type of input signal was a swept-sine signal with a sampling frequency of 48 kHz. The length of the swept-sine signal was 65536 samples.

To avoid unwanted coupling between the neck and the area behind the neck of the driver unit and to achieve high impedance at the glottis end, we inserted a close-fitting metal cylindrical filler inside the neck. We made a hole in the center of the metal filling with an area of 0.13 cm^2. The output sounds were recorded using a microphone from the sound level meter (RION NL-18) and an audio interface (RME Multiface) with a sampling frequency of 48 kHz. The microphone was placed approximately 20 cm in front of the output end in a sound-treated room (Fig. 4). The signals recorded were synchronously averaged multiple times to gain the signal-to-noise ratio.

Figure 5 is a sound spectrogram of output signals recorded with the first type of input signal. The output signals from the five configurations were concatenated for this analysis. As shown in this figure, we can observe clear formants, especially the first and second formants (F1 and F2) in the lower frequency region. The five vowels were clearly distinguishable during an informal listening test.

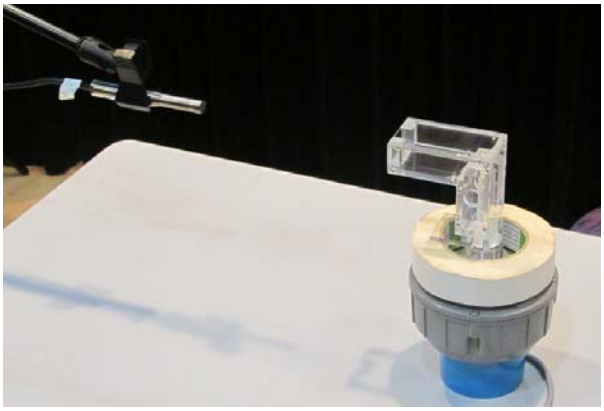

FIGURE 4. Recordings for the bent-type models with sliding blocks (Models A and B).

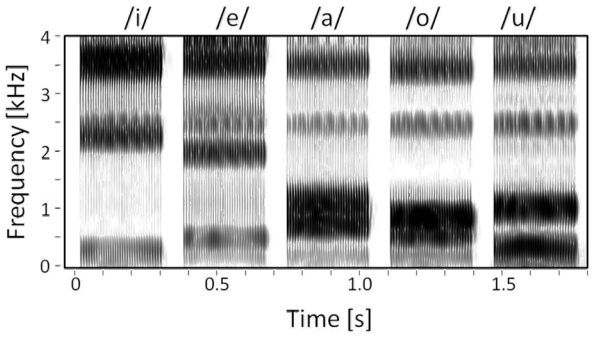

FIGURE 5. Spectrogram of output signals from the bent-type models with sliding blocks (Models A and B).

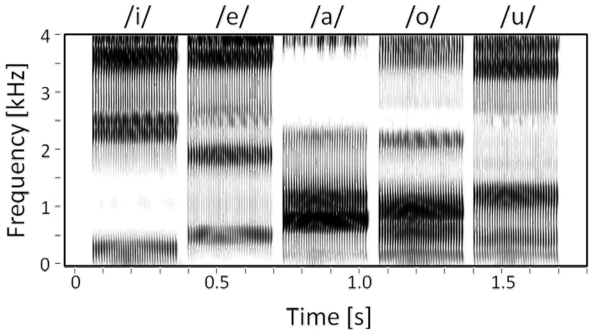

FIGURE 6. Spectrogram of output signals from the single bent-type model with sliding blocks (Model C).

A Single Bent-type Model with Sliding Blocks (Model C)

A driver unit (TOA TU-750) for a horn speaker was again attached to the glottis end of the model; a close-fitting metal cylindrical filler was also used. The first input signal for Models A and B was fed into the driver unit via a USB audio amplifier (Onkyo MA-500U).

The output sounds were recorded using a microphone (Sony ECM-23F5) and a digital audio recorder (Marantz PMD660) 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.

The vocal-tract configuration shown in Fig. 3 was used for the recordings of each vowel. For the vowel /u/, the groove of the main block (the largest one) was, unfortunately, connected to the central hole that was created when sliding the block for the tongue dorsum protrusion diagonally. Therefore, small thin plates were used to block the two connections (the short red lines in Fig. 3).

Figure 6 is a sound spectrogram of output signals concatenated for this analysis and recorded from the five configurations. As shown in this figure, we can observe a similar spectrogram to the one in Fig. 5. However, F1 is less clear in vowels /u/ and /o/. From an informal listening test, vowels /i/, /e/ and /a/ were clearly heard. The vowels /o/ and /u/ had reasonable quality but were a bit less intelligible.

DISCUSSION AND CONCLUSIONS

In this paper, we redesigned mechanical bent-type models. Two bent-type models with sliding blocks were tested: one for front vowels and one for back vowels. As a result, we confirmed that the five Japanese vowels were produced clearly. We then designed a single bent-type model with sliding blocks and confirmed that the five vowels were again produced with reasonable quality. However, it is not easy to slide the blocks in the current model, so improving the design for better usability is a future goal. Ultimately, users should be able to simultaneously

manipulate for tongue height and advancement and check the actual sounds. We would also like to objectively evaluate the usefulness of this new model in a pedagogical situation in acoustics, speech science, phonetics, etc.

Based on our teaching experience, we need both straight-type and bent-type models for different purposes for education in acoustics and speech science. The straight models are much simpler than the bent models, and we can demonstrate the area function is the most crucial factor which determines the vowel quality, but not bending itself. The CT model, one of the simplest models, is a static model, but if you compare the shapes and the sounds of different types of the CT model, we can demonstrate that vocal-tract configuration is associated with the quality of vowels. The S3T model, another simplest model that can achieve similar shapes comparing to the different types of the CT model, is a dynamic model, so that a single S3T model can produce different vowels by changing the position and the size of the slider. Thus, this model partially simulate the tongue movement in our vocal tract.

However, the straight models have disadvantages: 1) it is not easy for learners to imagine how vocal tract is placed in our head; and 2) the tongue movement is less realistic. The bent models, on the other hand, can cover these two points. 1) The vocal tract starts from the throat and ends at the lips, and it is very natural to understand for learners that the vocal tract is bent in between them. Furthermore, 2) more realistic tongue movement can be achieved by the bent-type models. As a result, they can produce a vowel sequence, such as /aia/, with a natural tongue movement. The flexible tongue model with a gel-type material is also a bent-type model, but it is very difficult to reproduce the same vocal-tract configuration multiple times with such model. The bent-type models with blocks as proposed in this study, however, can easily reproduce the same configuration repeatedly. In addition, the bent-type models can extend the coverage of sounds into some types of consonant, such as “glides” and “liquids.” In any case, we should select what type of models that we use for educational purposes depending on what we want to teach. Therefore, we need to collect more different types of vocal-tract models for different purposes.

ACKNOWLEDGMENTS

I acknowledge the anonymous reviewers for their helpful comments. This work was partially supported by Grant-in-Aid for Scientific Research (No. 24501063) from the Japan Society for the Promotion of Science. I would also like to thank Keiichi Yasu for his assistance. Although I am the sole author of this paper, I have used “we” and “our” in the text to refer to myself and my colleagues.

REFERENCES

1. T. Arai, “Education system in acoustics of speech production using physical models of the human vocal tract,” Acoust. Sci. Tech. **28**, 190-201 (2007).
2. T. Arai, “Education in acoustics and speech science using vocal-tract models,” J. Acoust. Soc. Am. **131**, 2444-2454 (2012).
3. 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. **29**, 188-190 (2008).
4. T. Arai, “Physical models of the human vocal tract with gel-type material,” Proc. of Interspeech, 2651-2654 (2008).
5. G. Fant, *Theory of Speech Production*, Mouton, The Hague, Netherlands (1960).
6. W. von Kempelen, *Mechanismus der menschlichen Sprache und Beschreibung einer sprechenden Machine*, Wien, Austria (1791).
7. 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, 1533-1536 (2002).
8. T. Arai, “Sliding three-tube model as a simple educational tool for vowel production,” Acoust. Sci. Tech. **27**, 384-388 (2006).
9. 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).
10. T. Arai, “Mechanical vocal-tract models for speech dynamics,” Proc. of Interspeech, 1025-1028 (2010).

APPENDIX

Connected-tube Models with Square Tubes

Before designing bent-type models, we started with straight-tube models. The CT model was originally designed with straight cylindrical tubes. However, we redesigned it with square tubes. Figure A-1 shows the CT models with square tubes. From an informal listening test, it was confirmed that the shape of the cross-section does not matter for vowel quality (either circular or square, in this case), but the area function matters.

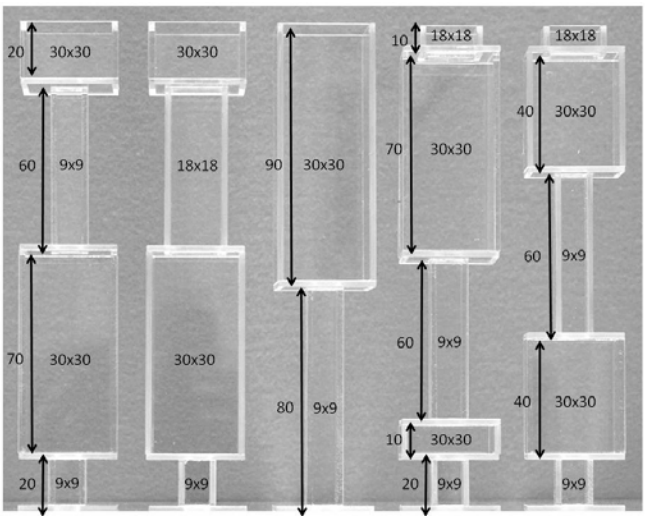

FIGURE A-1. Connected-tube models with square tubes. The vowels are /i/, /e/, /a/, /o/, and /u/ from left to right (the glottis end is the bottom, the lip end is the top). The material is acrylic resin. The numbers in this figure denote the dimension of each part of the models.