BENT VS. STRAIGHT MODELS OF THE HUMAN VOCAL TRACT BASED ON MEASUREMENTS BY CHIBA AND KAJIYAMA

Takayuki Arai

Sophia University
arai@sophia.ac.jp

ABSTRACT

Chiba and Kajiyama measured three-dimensional configurations of the human vocal tract during the production of five Japanese vowels and published the resultant data in "The Vowel: Its Nature and Structure." The data were estimated by palatograms as well as X-ray images and were naturally bent. They also derived cross-sectional area functions for the areas perpendicular to the midline along the length of the vocal tract. The overall area functions were simplified by interpolating them in a piecewise-linear fashion.

In this study, we first compared replicated physical models of the original bent configurations with approximated straight models based on the same area functions. We then created bent models with further simplifications and exhibited them at a museum. These models have pedagogical usefulness because they are simply shaped but also bent, which makes it easy to imagine where they are located in the head.

Keywords: Chiba and Kajiyama, The Vowel, physical models, vocal tract, speech production

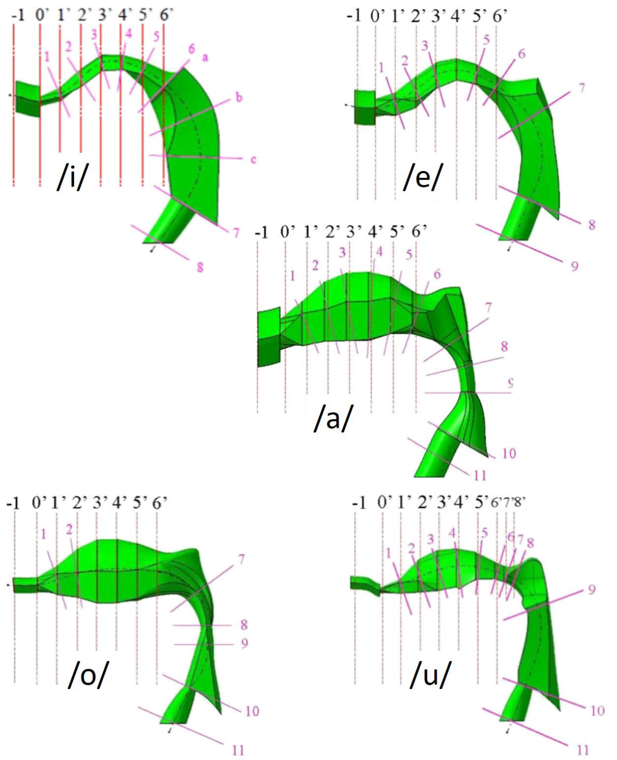

Figure 1: Midsagittal view of five Japanese vowels based on measurements in [1].

1. INTRODUCTION

Chiba and Kajiyama's 1941–42 book "The Vowel: Its Nature and Structure" [1] introduced several fundamental topics on vowel production and perception that paved the way to modern speech science [2–6]. One key issue pointed out in the book is how vocal-tract configurations determine vowel qualities, and to examine it, the authors measured 3D configurations of the vocal tract during the production of five Japanese vowels. These measurements were performed using the state-of-the-art X-ray apparatus at that time (Shimazu Co.), palatography (in which a set of shapes is molded out of plaster), and laryngoscopy (for observing the pharyngeal cavity). By combining these three techniques, they were able to obtain detailed 3D configurations of the vocal tract.

Chiba and Kajiyama derived cross-sectional area functions based on the 3D measurements by 1) drawing the midline of each vocal-tract configuration of each vowel and 2) estimating the cross-sectional areas of points on the midline where the cross-sections are perpendicular to it. They further simplified the area functions and created physical models in round-bottle shapes. After feeding in a sound source, they found that the physical models based on the area functions were close enough to their naturally produced counterparts. Arai [7] replicated their physical models not only for historical but also pedagogical purposes.

One of the options they might have selected was to model vocal tracts by keeping the bent shapes. Therefore, in this study, we compared bent vs. straight models with the same area functions. We also further simplified the configurations for bent models and displayed them at a museum exhibition.

2. PHYSICAL MODELS

Figure 1 shows the midsagittal view of the vocal-tract configurations of five Japanese vowels based on the measurements in [1]. The vertical lines in each plot show the sections (numbered with the prime symbol)

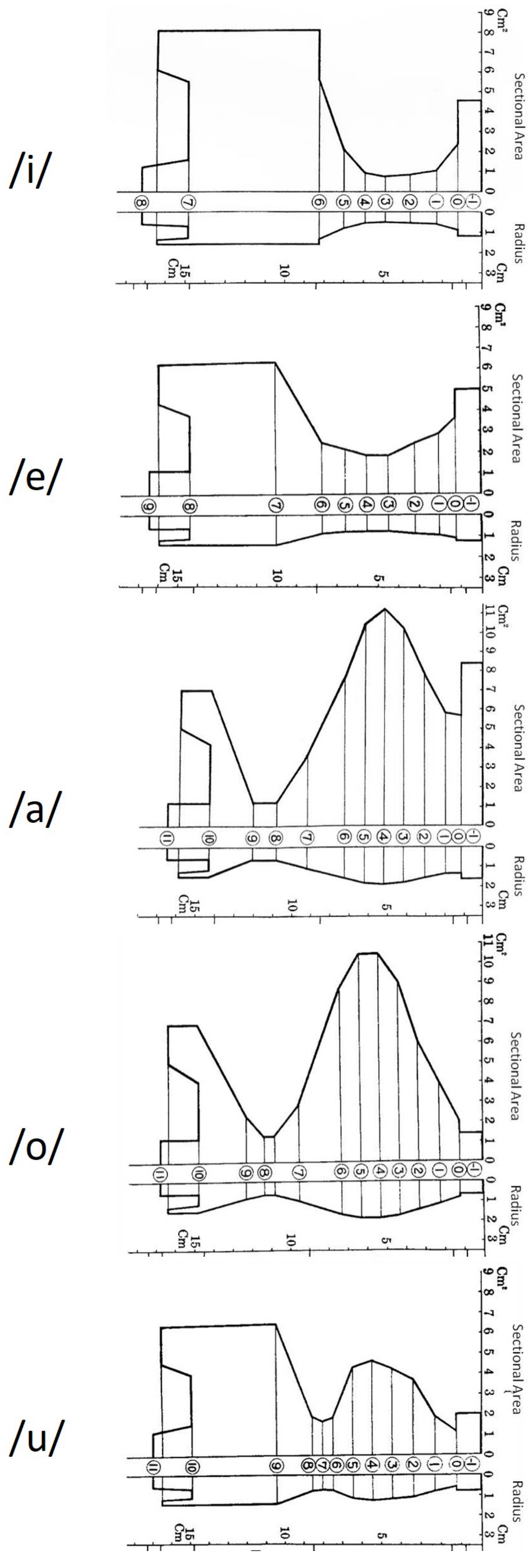

Figure 2: Cross-sectional area functions of five Japanese vowels adapted from [1].

Figure 3: Replicated physical models of five Japanese vowels (from left: /i/, /e/, /a/, /o/, and /u/) [7].

that were originally measured by palatograms. At the same point of the midline in each plot, additional lines show the sections (numbered with no prime symbol) that are perpendicular to the midline. In Fig. 2, the area above the horizontal axis in each plot shows the cross-sectional area function based on the perpendicular lines in Fig. 1, where the numbers in both figures correspond to each other. These area functions were drawn by piecewise linear approximation. The area below the horizontal axis in each plot in Fig. 2 shows the radius function when the cross-sections were modeled as circles.

Using the radius function as a basis, Chiba and Kajiyama created physical models in the “Artificial Vowels” section in [1], and after feeding in a sound source, they showed that artificially produced vowels were close enough to the naturally produced counterparts. However, the original models no longer exist (presumably due to wartime disruption), so the replication of the physical models created by Arai [7] is shown in Fig. 3. Arai [8] tried to model vocal tracts of vowels /i/ and /a/ by keeping the bent configurations based on Chiba and Kajiyama’s original measurements. In [8], there were two major design constraints for the bent models: 1) the midsagittal cross-section must be the same as the one in Fig. 1, and 2) the cross-sectional area function must be the same as the one in Fig. 2. Although Chiba and Kajiyama drew outlines of the uvula, epiglottis, and piriform fossa (as shown in Fig. 1), they were simplified in Fig. 2, so the designs in [8] omitted them. The designs for vowels /i/ and /a/ were successful, but there were no descriptions for the rest of the vowels (i.e., /e/, /o/, and /u/), and no acoustic analysis was ever reported.

This is why, in this study, we expand the design of these bent models so that all five vowels can be implemented. We followed the same design constraints as above and then 3D-printed the bent models, as shown in Fig. 4.

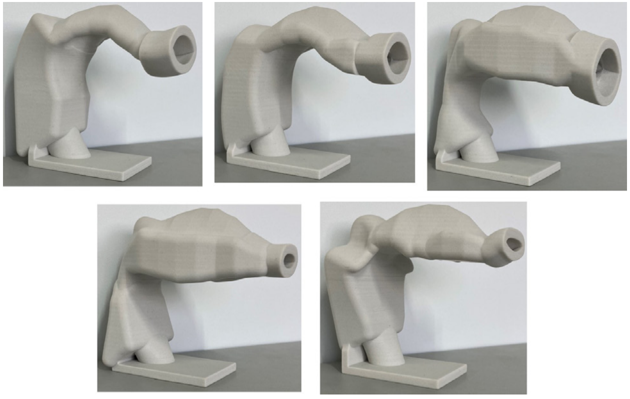

Figure 4: 3D-printed bent models of five Japanese vowels based on measurements in [1].

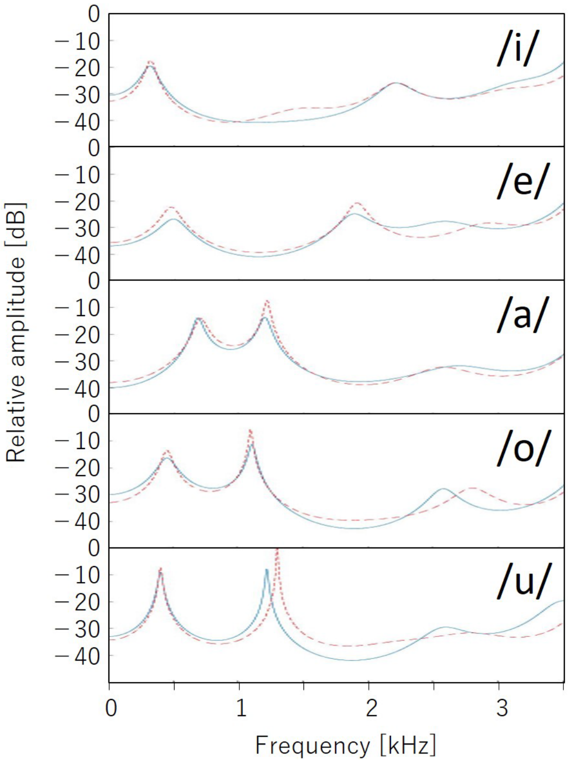

Figure 5: LPC-based spectral envelopes of bent vs. straight models for five Japanese vowels (LPC order: 10).

3. ACOUSTIC ANALYSIS

In this section, we compare the frequency characteristics of the bent vs. straight models.

3.1. Recordings

We wanted to subjectively judge the vowel quality of each output sound, so the recordings were done by feeding in a simple impulse train from a deriver unit of a horn speaker. We digitally recorded in a quiet laboratory at Sophia University and set the original sampling frequency to 44.1 kHz and the quantization to 16 bits. A microphone (Rode, NT6) were placed approximately 150 mm away from the mouth opening.

3.2. Analysis method

We utilized Matlab to analyze each spectrum during vowel production by the linear predictive coding (LPC) method. First, the speech signals were downsampled to 8 kHz, and then the LPC-based spectral envelope was calculated for each utterance.

3.3. Results

Figure 5 shows the comparison between the bent (solid lines) and straight (dashed lines) models for each vowel. As we can see, the first two formant frequencies (F1 and F2) were almost identical between the bent and straight models for each vowel, although there was a slight difference between the two models in high-frequency regions above 2 kHz.

3.4. Discussion

Our findings suggest that Chiba and Kajiyama presumably already knew that the bent and straight models have almost the same frequency characteristics as long as they have the cross-sectional area function in common, where the cross-dimensions are taken perpendicularly against the midline along the vocal-tract length. This is especially evident in lower-frequency regions below 2 kHz. The results in Fig. 5 prove that Chiba and Kajiyama were correct in their original methods. Fant [9] and other researchers have followed the same methodology [2]. The reason for the difference in frequency characteristics might not be caused by the issue of straight vs. bent tubes. When Chiba and Kajiyama converted a bent tube to a straight tube, they also changed the shape of each cross-section to that of a circle. As a result, this might have yielded such differences.

4. PEDAGOGICAL APPLICATIONS

Various previous studies [e.g., 7, 10–12] have demonstrated the usefulness of physical models of the human vocal tract. Here, we introduce yet another set of bent models with extra simplifications.

4.1. Simplification

As discussed in Section 2, there are two design constraints for the bent models in Fig. 5. However, in in Chiba and Kajiyama's work [1], coronal sections were only available in the oral cavities, and not in the pharyngeal cavities. Therefore, some parts of the configurations in Fig. 4 are not realistic and might lead to misunderstandings if the models are applied as educational tools. Therefore, we further simplified the configurations so that the inner cavities (except the hypopharynx and the lips) are approximated by simple box shapes, instead.

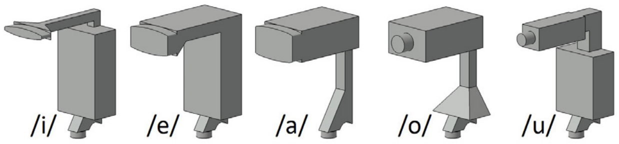

Figure 6: Simplified inner cavities of five Japanese vowels.

The resultant vocal-tract configurations are shown in Fig. 6, where only the inner cavities (i.e., not the outer shapes) for all five vowels are depicted. As we can see, for vowels /i/ and /e/, the pharyngeal cavities are wide open, the oral cavities are narrow, and the lips are spread. For vowels /a/ and /o/, on the other hand, the oral cavities are wide open and the pharyngeal cavities are narrow. For vowel /u/, the major parts of the oral and pharyngeal cavities are relatively wide, and the corner part (the middle part of the vocal tract) is narrow. For vowels /o/ and /u/, the lips are rounded, so each model has a narrow cylinder shape at the mouth end.

4.2. Hypopharynx

When designing the hypopharynx in Fig. 6, there were two options we could have taken. The first was to design the so-called "larynx tube" [13], as shown in Fig. 7(a), which is closer to Chiba and Kajiyama's original measurement. The second option was to design it as a narrow and straight tube, as shown in Fig. 7(b).

We conducted an acoustic analysis between sounds for the two cases in Fig. 7. The results (shown in Fig. 8) indicate a difference around the 3-kHz region in the spectra as well as around the F1 and F2 frequencies. The F1 and F2 frequencies are both within vowel /a/ in Japanese, whereas the frequency range of 2.5–3.5 kHz matches the one for the "singer's formant" in [13]. The case in Fig. 7(a) is similar to the situation where the larynx is lowered [13], which is what caused the higher peak level in this frequency range; the difference was approximately 6.5 dB (Fig. 8). The bandwidths of F1 and F2 for the model in Fig. 7(b) were wider, which might suggest a breathier voice quality.

In the case of vowel /o/, the hypopharynx is much wider than vowel /a/. In addition, the pharyngeal cavity above that section should be sufficiently narrow. These points are reflected in the design of /o/ in Fig. 6, and result in a better sounding quality for /o/.

4.3. Museum exhibition

We 3D-printed the models (Fig. 9) discussed in this work and temporarily exhibited them at the National Museum of Ethnology, Japan in 2022 [14]. Previous

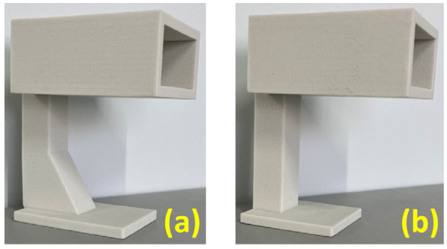

Figure 7: Two versions of simplified bent models of vowel /a/.

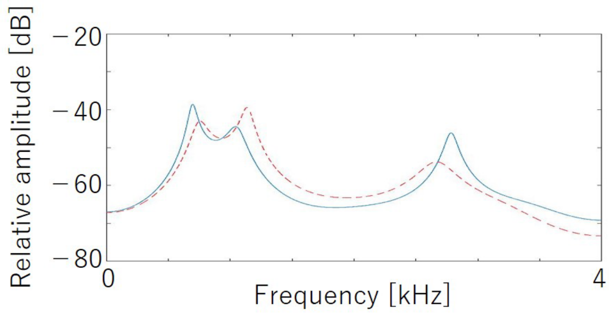

Figure 8: LPC-based spectral envelopes of the two models in Fig. 7 (solid: model in Fig. 7(a); dashed: model in Fig. 7(b)). F1 and F2 frequencies were approximately 693 and 1050 Hz for the Fig. 7(a) model and 761 and 1135 Hz for the Fig. 7(b) model.

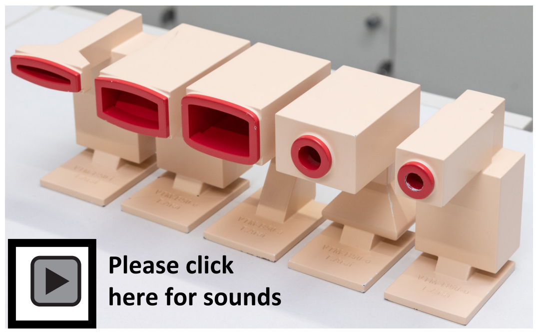

Figure 9: 3D-printed and painted version of VTM-B50 models. A video clip is available here [15].

studies [10–12] have featured mainly straight tubes or head-shaped models. These newly designed models (VTM-B50) have an additional pedagogical usefulness because they are simply shaped but also bent, which makes it easy to imagine where they are located in the head.

5. SUMMARY

In this study, we examined bent and straight models of the human vocal tract for five Japanese vowels based on the measurements by Chiba and Kajiyama [1]. The lower frequency characteristics between the two were similar, thus proving that Chiba and Kajiyama conducted the analyses correctly the first time. We further simplified the design process for pedagogical purposes and exhibited the models at a museum.

6. ACKNOWLEDGMENTS

This work was partially supported by JSPS KAKENHI Grant Numbers 21K02889 and Sophia University Special Grant for Academic Research (Research in Priority Areas).

7. REFERENCES

[1] Chiba, T. and Kajiyama, M. 1941–42. *The Vowel: Its Nature and Structure*. Tokyo-Kaiseikan.
[2] Maekawa, K. and Honda, K. 2001. On the Vowel, Its Nature and Structure and related works by Chiba and Kajiyama. *J. Phonetic Soc. Jpn.* 5(2), 15–30.
[3] Arai, T. 2004. History of Chiba and Kajiyama and their influence in modern speech science. *From Sound to Sense*. MIT.
[4] Maekawa, K. 2002. From articulatory phonetics to the physics of speech: Contribution of Chiba and Kajiyama. *Acoust. Sci. Tech.* 23(4), 185–188.
[5] Stevens, K. N. 2001. The Chiba and Kajiyama book as a precursor to the acoustic theory of speech production. *J. Phonetic Soc. Jpn.* 5(2), 6–7.
[6] Fant, G. 2001. T. Chiba and M. Kajiyama, pioneers in speech acoustics. *J. Phonetic Soc. Jpn.* 5(2), 4–5.
[7] Arai, T. 2001. The replication of Chiba and Kajiyama's mechanical models of the human vocal cavity. *J. Phonetic Soc. Jpn.* 5(2), 31–38.
[8] Arai, T. 2017. Chiba and Kajiyama measured 3D configurations of the human vocal tract and made physical models based on the measurements. *Proc. Second International Workshop on the History of Speech Communication Research (HSCR 2017)*. 55–60.
[9] Fant, G. 1960. *Acoustic Theory of Speech Production*. Mouton.
[10] Arai, T. 2007. Education system in acoustics of speech production using physical models of the human vocal tract. *Acoust. Sci. Tech.* 28(3), 190–201.
[11] Arai, T. 2012. Education in acoustics and speech science using vocal-tract models. *J. Acoust. Soc. Am.* 131(3), Pt. 2, 2444–2454.
[12] Arai, T. 2016. Vocal-tract models and their applications in education for intuitive understanding of speech production. *Acoust. Sci. Tech.* 37(4), 148–156.
[13] Sundberg, J. 1974. Articulatory interpretation of the "singing formant," *J. Acoust. Soc. Am.* 55(4), 838–844.
[14] https://loquens.site/
[15] https://youtu.be/aLyqaafEM38