音声研究　第5巻第2号
2001(平成13)年8月
31-38頁

Journal of the Phonetic Society
of Japan, Vol.5 No.2
August 2001, pp.31-38

The Replication of Chiba and Kajiyama's Mechanical Models of the Human Vocal Cavity

Takayuki ARAI*

Chiba and Kajiyamaによる声道模型の復元

荒 井 隆 行*

要旨：本論文ではChiba and Kajiyamaによる声道模型を復元した。彼らの模型では材料として油粘土が使用されているが，本論文では加工のしやすさと透明性の理由からアクリル樹脂を使用した。復元は，Chiba and Kajiyamaによる日本語5母音に対する声道形状の測定に基づいた。声門側の端に音源を入力することで，各模型が対応する母音を産出することを確認した。さらに，これらの模型を使用することが音声科学における教育の現場で有効であることが示された。

Key words: Chiba and Kajiyama's model, replica, vocal cavity, artificial vowel, education tool

1. Introduction

Chiba and Kajiyama (1941) was foundational in the establishment of the modern acoustic theory of speech production (Fant, 1960; Stevens, 1998). Chiba and Kajiyama's book views the mechanism of vowel production from the viewpoints of physiology, physics and psychology. In the first half of the book, the argument addresses phonation (pertaining to the larynx) and articulation (pertaining to the vocal cavity). In Part One, they investigate the structure and action of the larynx. In Part Two, the theories of vowels are described, and they discuss the mechanism of vowel production. They relate the configuration of the vocal cavity with its resonance based on acoustic theory and electrical analogs. In Part Three, they accurately measure the vocal cavity to create a simple model of a resonator that is acoustically equivalent to the cavity. The measurement is done using X-ray photographs and special solid palatograms. Furthermore, they produce artificial vowels by attaching a larynx-tone emitter to models constructed on the basis of the data. The natural frequencies of the vocal cavity as a non-uniform acoustic tube are also calculated. In Part Four, they deal with the sense of hearing for vowel perception.

In the section "Artificial Vowels" (1941, pp. 128–131) Chiba and Kajiyama describe their mechanical models. They base the shape of their models on sectional measurements they made of vocal cavities. Figure 1 shows their Plasticine model for the vowel [i]. Because it is difficult to calculate the natural frequencies of an actual vocal cavity due to its complex shape, Chiba and Kajiyama converted it into a simpler shape with an acoustically equivalent value.

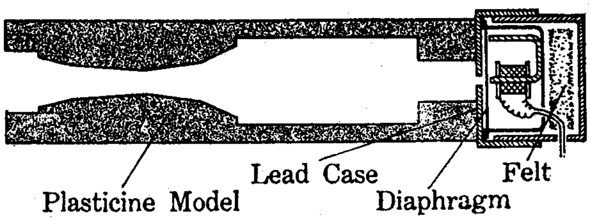


Fig. 1: Chiba and Kajiyama's mechanical model for vowel [i] (Chiba and Kajiyama, 1941, p.129).

In Chiba and Kajiyama's book, Figs. 69, 71, 73, 75 and 77 contain a simplified version of the sectional measurements of the vocal cavities.

* Associate Professor, Dept. of Electrical and Electronics Eng., Sophia University (上智大学理工学部電気・電子工学科助教授)

In the right half of each figure, the radius at every position from the glottis to the lips is shown as a polygonal line. We will obtain an equivalent cavity shape for each vowel by making a round, bottle-shaped model based on the radius curve corresponding to their measurements.

In Section 2, we detail the design of our model, highlighting the process by which we shaped the models based on their measurements. In Section 3, we evaluate our models from acoustic and subjective points of view. Finally, Section 4 discusses the usefulness of our models as a tool in speech science education.

2. Making the replicas

2.1 Design

In the right half of Figs. 69, 71, 73, 75 and 77 in Chiba and Kajiyama (1941), polygonal lines show the radius for five Japanese vowels based on their measurement as described in Section 1. Their round bottle-shaped models were formed by rotating these radius curves around the pivot. Unfortunately, Chiba and Kajiyama do not specify the exact dimensions of the radius curves. Thus, we made a drawing of the model for each vowel by specifying all apexes with resolution of 1 mm (or 0.5 mm) as shown in Fig. A.1 in the Appendix. Any two adjacent apexes are connected with a straight line. Figure 2 shows a computer simulation of the three-dimensional model for the vowel [i].

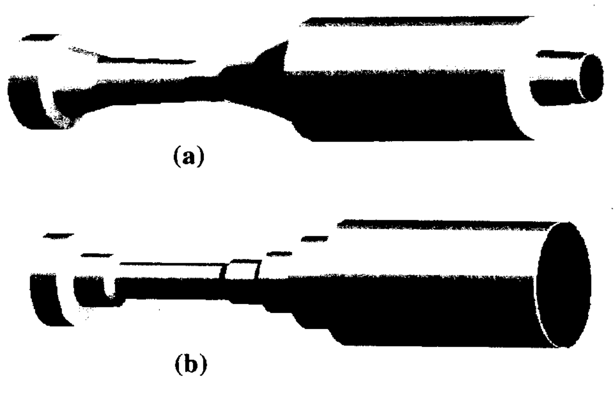


Fig. 2: Computer simulation of round bottle-shaped models for vowel [i]; (a) the precise reproduction and (b) the step-wise approximation (see text for more details).

2.2 Two types of models

Replicas of Chiba and Kajiyama's models can be produced from the radius curves in Fig. A.1. We call this type of model a "precise reproduction." Another (easier) way to make replicas is to use the zeroth-order approximation of the radius curves. We call this type of model a "step-wise approximation." In this case, we approximate the radius curves at 10 mm resolution in a step-wise fashion. We made both a precise reproduction and a step-wise approximation.

2.3 Materials

While Chiba and Kajiyama's original models were made of Plasticine, we use acrylic resin because it is both transparent and easy to sculpt. In Chiba and Kajiyama's discussion of wall effects (1941, p.94) they notice no great difference in when they compare tubes made of several different materials. Therefore, we assume that any acoustic differences resulting from the different materials we used for our models will be minimal.

2.4 Sculpting the models

For the precise reproduction type of model, we used an acrylic cylinder with a diameter of 50 mm. We sculpted the resin so that the cavity formed a round bottle-shape, based on Fig. A.1. Prior to sculpting the cavity we cut each cylinder into two or three portions in order to reach the less accessible areas of the cavity with the sculpting tool. Finally, we glued each portion of the cylinder together.

The step-wise approximation model consists of a set of acrylic plates, each with a hole in the center. When placed side-by-side the holes in the plates form a tube, the cross-sectional area of which changes in a step-wise fashion. Each plate is 100 mm x 100 mm x 10 mm. The diameters of the holes vary based on the zeroth-order approximation with a 10 mm resolution of the radius curves in Fig. A.1 (Table 1). For comparison, we also made several plates having holds of the same diameter to form a uniform tube approximating the central reduced vowel, schwa.

Table 1: Diameters (in mm) for five Japanese vowels used in the step-wise approximation.

index# from glottis	[i]	[e]	[a]	[o]	[ɯ]
1	30	30	34	26	26
2	30	30	20	20	26
3	30	30	12	16	26
4	30	30	14	14	26
5	30	28	16	16	24
6	30	26	20	22	16
7	22	22	26	28	12
8	14	20	30	34	16
9	10	16	34	38	22
10	8	16	38	34	22
11	8	16	34	34	20
12	8	20	30	26	18
13	12	18	26	22	14
14	22	24	32	14	14

[unit: mm]

Fig. 3: Precise reproduction model for five Japanese vowels (from left, [i], [e], [a], [o], and [ɯ]). In order to take the photograph the cavities were filled with flour. This made their shape more visible. The joints (where the cut parts were glued together) can also be seen in this photograph.

Figures 3 and 4 show the two types of models. Figure 3 is the precise reproduction for each of the five Japanese vowels, and Fig. 4 is the step-wise approximation for the vowel [a].

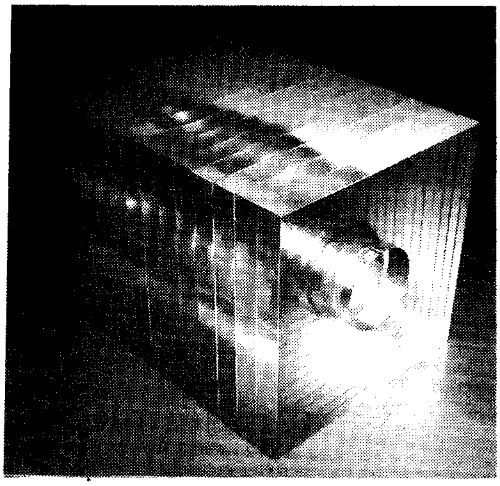


Fig. 4: Step-wise approximation model for vowel [a].

3. Evaluation

We recorded the sounds emitted from the mouth end of the tube and evaluated them with spectral analysis. We also conducted an informal listening test to evaluate the output subjectively.

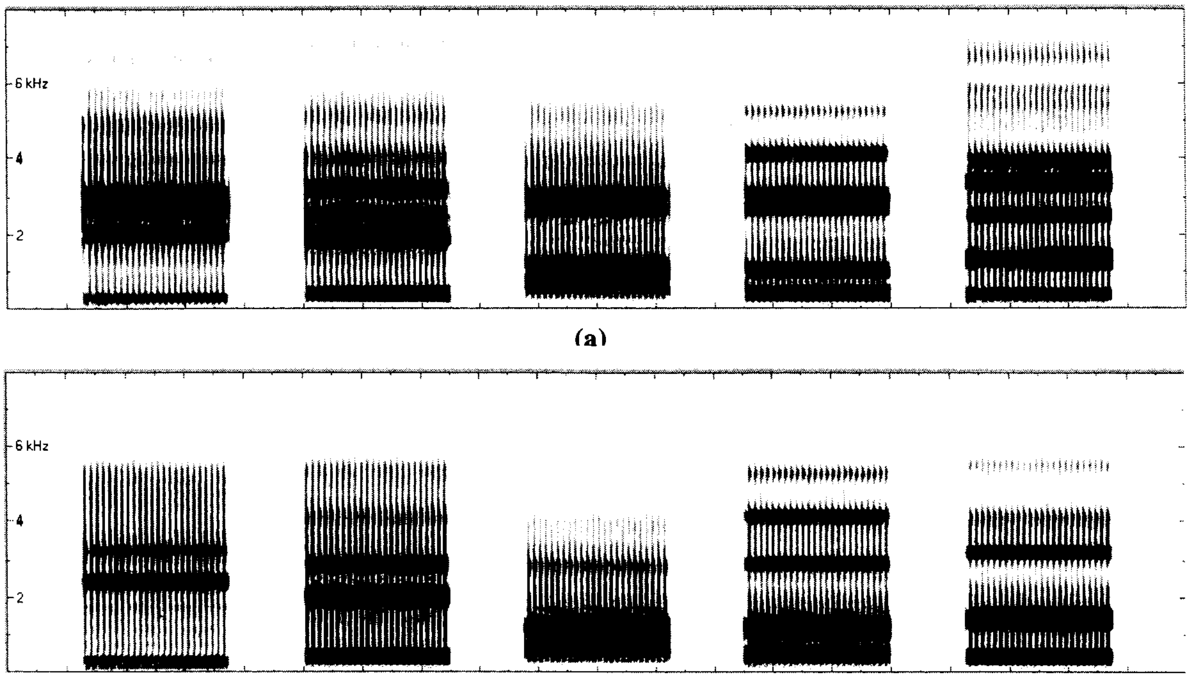
3.1 The sound source

Several prototypes of the human vocal tract have been developed in the past, each using a different sound source. An example of an early vocal-tract model is the mechanical speech synthesizer by Von Kempelem about 1780 (Gold and Morgan, 2000). In this model, vibrating reeds simulate a voiced sound source. Chiba and Kajiyama used a telephone receiver; its diaphragm was made to pulsate magnetically by means of an electric current sent out from a neon-lamp oscillator. A horn driver can be used to feed an arbitrary source signal, as well. We used an electrolarynx of the type often used by patients whose vocal folds have been removed due to cancer of the larynx.

In our model, and indeed in all of these models, the acoustic tube has an opening at each end corresponding to the two ends of the vocal cavity: the mouth and glottis. When the sound source is connected to the glottis end, a vowel-like sound is emitted from the mouth end.

3.2 Spectral analysis

Figure 5 shows the spectrograms of the output signals of the five vowels for the two types of models: precise reproduction and step-wise approximation. This figure confirms that



(b)

Fig. 5: Spectrograms of the produced sounds of five Japanese vowels (from left, [i], [e], [a], [o], and [ɯ]) for each type of model; (a) the precise reproduction model and (b) the step-wise approximation.

Table 2: Measured first and second formant frequencies (F1 and F2 in Hz); (a) for the precise reproduction model and (b) for the step-wise approximation model.

(a)

	[i]	[e]	[a]	[o]	[ɯ]
F1	230	405	654	436	499
F2	1992	1868	1121	1054	1276

(b)

	[i]	[e]	[a]	[o]	[ɯ]
F1	293	436	654	529	423
F2	2383	1992	1152	1183	1370

Table 3: Confusion matrices of a listening test; (a) for the precise reproduction model and (b) for the step-wise approximation model.

(a)

		response				
		i	e	a	o	u
stimulus	i	54	0	0	0	0
	e	7	47	0	0	0
	a	0	0	54	0	0
	o	0	2	1	18	33
	u	1	0	0	0	53

(b)

		response				
		i	e	a	o	u
stimulus	i	54	0	0	0	0
	e	0	53	0	1	0
	a	0	0	54	0	0
	o	0	5	21	18	10
	u	0	14	6	4	30

the sounds produced with both types of models were similar to those in Chiba and Kajiyama (1941). The difference between our two model types was small, especially in the first and second formant frequencies (as shown in Table 2).

3.3 Listening test

An informal listening test was conducted in a classroom with 54 students in a speech and hearing course. Table 3 shows the confusion matrix for each type of model. Using the precise reproductions, the models yielded highly intelligible Japanese vowels, except for [o], which was more often than not mistaken for /u/. This could be due to a higher second formant frequency than average (Table 2). The vowels [a], [i] and [ɯ] sounded particularly natural. The vowel [e] was sometimes confused with /i/.

For the step-wise approximation, similar tendencies were observed; all vowels except [o] were highly intelligible. The vowel [o] sounded slightly open (similar to English [ɔ]), and was again the least intelligible of the vowels. The vowel [ɯ] was sometimes confused with /e/ in this type of model.

4. The replica as an education tool

We predicted that students would more easily understand the abstract concepts involved in the acoustics of speech if they had access to a three-dimensional model such as ours. Specifically, we hoped the tool would help students to grasp the relationship between the configuration of the vocal tract and acoustic output, which is an important topic, and one that is difficult to grasp because the vocal cavity is largely unseen (Arai et al., 2001).

To test our predictions, we used the vocal tract models during a lecture in a speech science classroom to investigate the effects the models had on learning. We mainly used the step-wise approximation models connected to a sound source. We found that after using the models, students were able to experience the out-working of important acoustic theories, especially source-filter theory and perturbation theory.

First, because of the tube's transparency, the location of the constriction is visible to the naked eye, as is the overall shape of the cavity. This design appeared to help observers associate the quality of a vowel heard with the location of constriction on the model.

Second, the relationship between frequency and pitch was illustrated by channeling sound sources with different frequencies through the tube. The audience was able to observe that the pitch of the out coming sound is determined by the fundamental frequency of the input signal.

Third, by changing the order of the plates to simulate constrictions at nodes and antinodes, the shift in formant(s) was heard as well as seen in a spectral analysis of the out coming sounds, which the students had access to as well. This exercise helped the audience to understand how formant frequencies are affected by a change in the constriction point in the vocal tract, which relates directly to an understanding of how vowels change depending on the location of constriction(s) in the vocal tract.

In general, the audience's understanding of vowel production increased after playing with the models. The experimenters received a good deal of positive feedback from participants. The relationships observed between point of constriction and vowel quality as well as between fundamental frequency and pitch are key to understanding source-filter theory. The simulation of constrictions at nodes and antinodes helps students to understand perturbation theory. We conclude from our experience that when used in a classroom environment, our model is particularly effective for increasing student understanding of the acoustic theory of speech production.

5. Summary

In this paper, we reproduced Chiba and Kajiyama's mechanical models of the human vocal cavity. From both subjective and acoustic evaluations of the reproductions we confirmed that the output sounds were very intelligible and had the same characteristics as those in Chiba and Kajiyama (1941). Chiba and Kajiyama's models are simplified but acoustically equivalent to a model that might be more exact to actual vocal cavity configurations. It is more appropriate to use their model in education, because the model makes it clear that it is the overall shape (including the location of constriction) that matters for vowel production, not the exact shape of the cavity.

We hope that models by Chiba and Kajiyama will be more widely recognized as a historical legacy and that they will be used as tools in speech science education in the future.

Acknowledgements:

First, I would like to thank Kikuo Maekawa of the National Language Research Institute, Japan, who gave me this great opportunity. I would also like to thank the members of the Phonetics Lab. and the Arai Lab. at Sophia University (especially Michiko Yoshida, Shingo Sugiyama, Nobuyuki Usuki, and Hasan A. Alwi) for helping me and giving me useful comments. I wish also to thank the students in my class who participated in the informal listening tests. Finally, I would like to express my grateful thanks to Terri Lander, Takashi Arai and Michiko Arai for their helpful support.

References:

Arai, T., Usuki, N. and Murahara, Y. (2001) "Prototype of a vocal-tract model for vowel production designed for education in speech science," *Proc. of Eurospeech,* Aalborg.

Chiba, T. and Kajiyama, M. (1941) *The Vowel: Its Nature and Structure,* Tokyo-Kaiseikan Pub. Co., Ltd., Tokyo.

Fant, G. (1960) *Acoustic Theory of Speech Production,* Mouton & Co., The Hague.

Gold, B. and Morgan, N. (2000) *Speech and Audio Signal Processing,* John Wiley & Sons.

Stevens, K. N. (1998) *Acoustic Phonetics,* MIT Press, Cambridge, MA.

Appendix

Fig. A.1: Drawing of the detailed design for each of the five Japanese vowels; (a) [i], (b) [e], (c) [a], (d) [o], and (e) [ɯ].

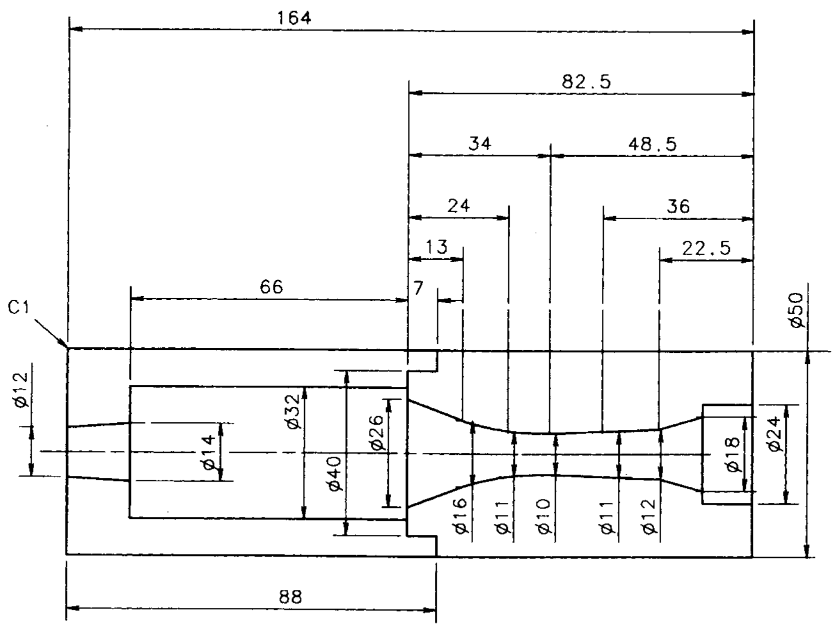


Fig. A.1 (a): Japanese vowel [i]

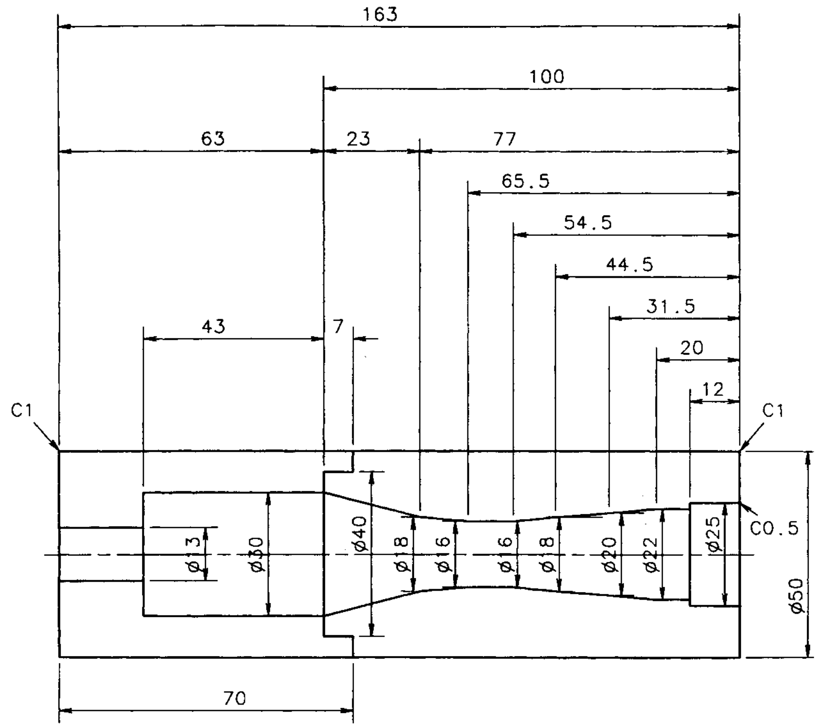


Fig. A.1 (b): Japanese vowel [e]

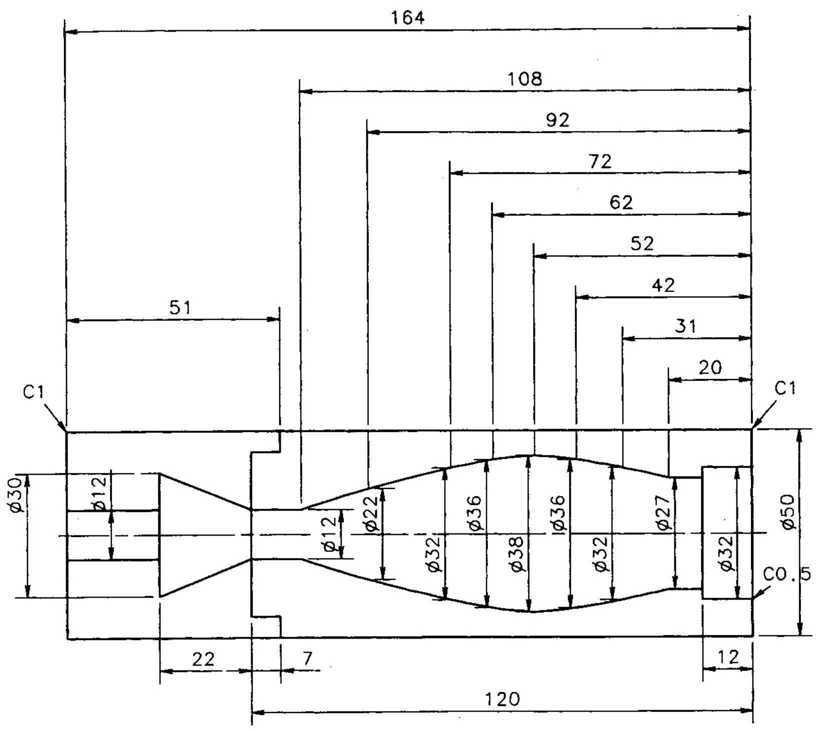


Fig. A.1 (c): Japanese vowel [a]

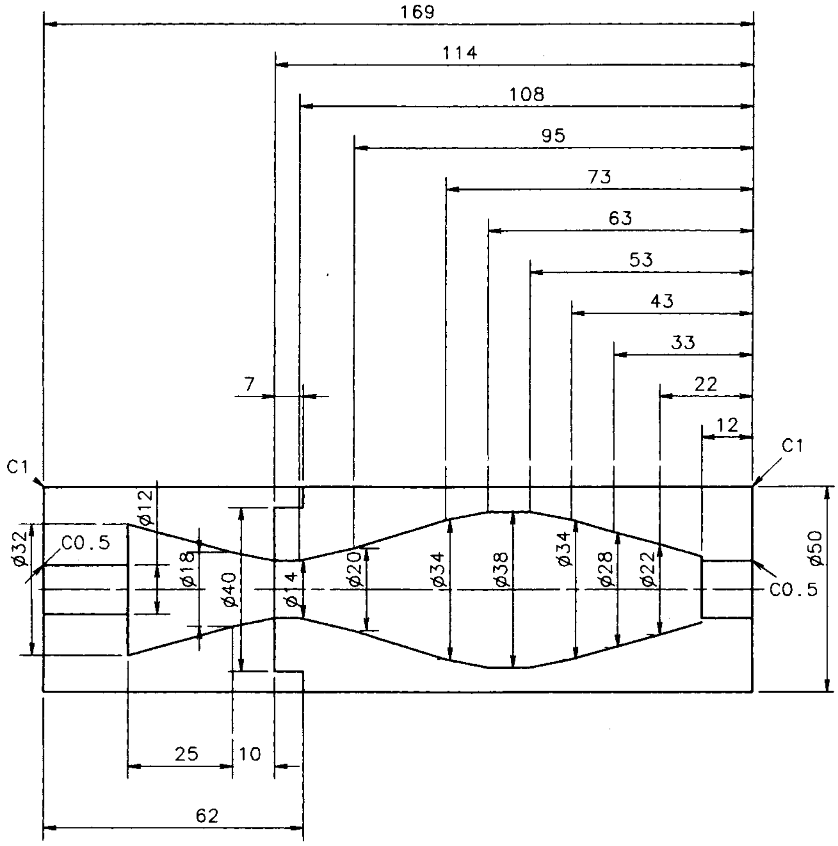


Fig. A.1 (d): Japanese vowel [o]

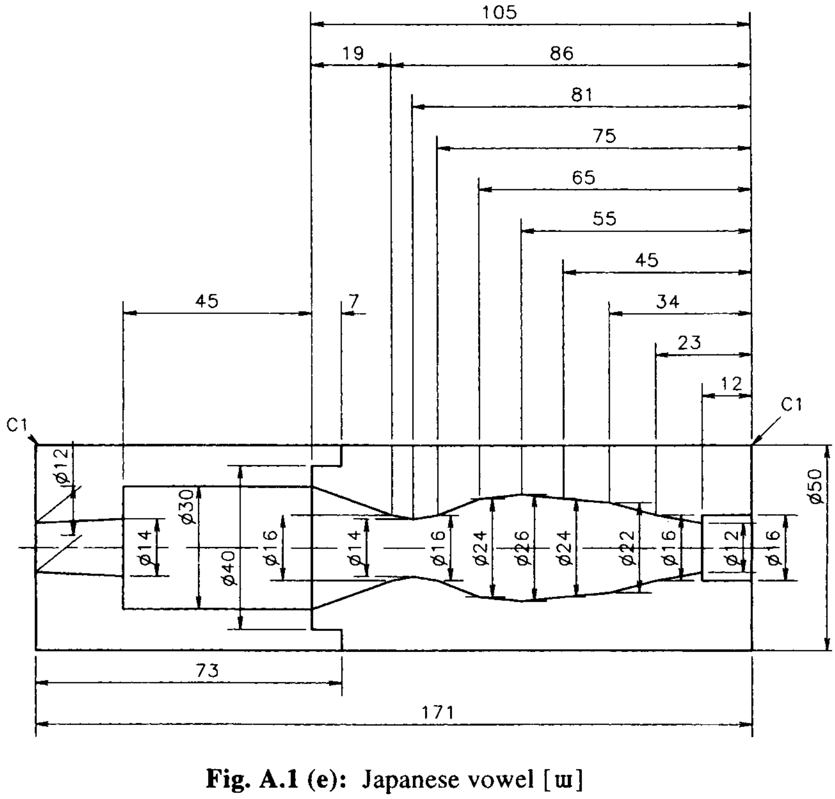


Fig. A.1 (e): Japanese vowel [ɯ]