ARTICULATORY POSITIONS OF JAPANESE VOWELS AS A FUNCTION OF DURATION COMPUTED FROM A LARGE-SCALE SPONTANEOUS SPEECH CORPUS

Takayuki Kagomiya

The National Institute for Japanese Language and Linguistics
t-kagomiya@ninjal.ac.jp

ABSTRACT

The purpose of this study is to illustrate certain features of the place of articulation of Japanese vowels on the basis of formant frequency. For this purpose, two types of relationship between vowel duration and formant frequency were examined. The first of these relationships was the correlation between speaking rate and formant frequency. The result of this analysis showed that as speaking rate slows, articulations get more distinct: open vowels increase in openness, close vowels increase in closeness, front vowels increase in frontness, and back vowels increase in backness. The second relationship was between formant frequencies and vowel length. The result of this analysis showed that place of articulation is more distinct in long vowels than in short vowels. The exception was the back vowel /u/, which did not show greater backness in long vowels. This result supports previous arguments that that Japanese /u/ is not a typical back vowel.

Keywords: Japanese vowels, articulatory position, speaking rate, corpus analysis, formant

1. INTRODUCTION

The purpose of this study is to illustrate certain features of the place of articulation of Japanese vowels. To fully describe the Japanese phonetic system, it is important to comprehend this attribute of Japanese vowels. Standard (Tokyo) Japanese has five vowels: /i/, /e/, /a/, /o/, and /u/. The traditional understanding of the place of articulation of these five vowels is seen in Table 1. However, articulation is not a static state of affairs but a dynamic movement, continually changing during speech. Thus, static schemata like the one in the table are not sufficient to describe vowel features.

Moreover, articulatory positions vary by speaker, emotion, speaking rate, and many other factors. For instance, it is widely observed that vowel articulatory movement is reduced as speaking rate increases [7, 3, 5], while in contrast, in slow speech, articulatory position is closer to its canonical target [3, 11]. In this article, this relationship between articulatory position and speaking rate is investigated in greater depth.

Table 1: Traditional description of the five Japanese vowels

	frontness	openness
i	front	close
e	front	mid
a	mid	open
o	back	mid
u	back	close

As mentioned above, it has been assumed that the articulatory position of a vowel becomes closer to its target as speaking rate becomes slower and duration becomes longer. Thus, it has been supposed that the relationship between speaking rate and direction of articulatory position of a vowel reveals its articulatory target position.

However, direct observation of tongue movement in running speech is quite difficult. In this study, therefore, articulatory positiowns were estimated on the basis of formant frequencies. Formant frequency is related to vowel articulatory position: the first formant (F1) increases with jaw openness, while the second formant (F2) increases with tongue frontness [1, 3, 9].

Hirata and Tsukada [4] also examined the relationship between speaking rate and formant frequency in Japanese. However, their study was conducted under laboratory conditions and using read speech, in which articulation is more careful and tends to be closer to target position regardless of speech rate [9]. Thus, spontaneous speech data is preferable to comprehend articulation status in relation to a target.

In the present study, data were extracted from a large-scale spontaneous speech corpus, the Corpus of Spontaneous Japanese (CSJ) [8]. This corpus contains speech by various people, on various topics, spoken at various rates. Thus, it is suitable for the investigation of the relationship between speaking rate and formant frequency.

Furthermore, the Japanese language makes a phonemic distinction between short and long vowels (that is, between /i/, /e/, /a/, /o/, and /u/ on the one hand and /i:/, /e:/, /a:/, /o:/, and /u:/ on the other). For instance, /obasaN/ means 'aunt', while /oba:saN/ means 'grandma'. Because long vowels have longer duration than short vowels do, it can be supposed that the articulatory positions of long vowels are closer to target position than those of short vowels. To test this, the difference in formant frequency between long and short vowels was also examined.

2. DATA EXTRACTION

2.1. Speech materials

The speech materials were extracted from the "Core" data subset of CSJ [8]. The Core dataset contains 177 spontaneous speech passages spoken by 137 speakers (79 male, 58 female), for a total of about 500,000 words or about 45 hours. All of the speech has been segmentally labeled for phoneme and timescale by trained human labelers. These segmental labels were referred to in order to extract vowel sections from the corpus.

2.2. Formant estimation

To extract formants, 12th-order LPC analyses were conducted on the results of FFT spectral analyses with an 0.049 s window length, by 0.01 s steps. These operations were carried out using the Snack Sound Toolkit software (`http://www.speech.kth.se/snack/`). The average estimated formant values across the duration of a segment were regarded as the formant values for the segment.

To deal with the inevitable errors in auto-estimation, data was screened using the following procedure:

1. Mean formant values and standard deviations (SDs) were estimated by vowel type by speaker.
2. Segments which had greater or smaller formant values than the mean $\pm$ 2 SD were regarded as outliers and removed from the data list.

The retained tokens after the data screening are shown in Table 2. Hereafter, /a/, /e/, /i/, /o/, and /u/ represent short vowels, while /aH/, /eH/, /iH/, /oH/, and /uH/ indicate long vowels.

Table 2: Number of vowels after data screening

male

a	e	i	o	u
98142	54631	53881	77361	32017
aH	eH	iH	oH	uH
3962	14580	2487	19269	8238

female

a	e	i	o	u
86231	42535	46574	63964	25956
aH	eH	iH	oH	uH
1811	7355	1540	15394	6277

Table 3: Correlation coefficients between formant frequencies and duration

male	a	e	i	o	u
F1	0.227	0.115	-0.281	0.058	-0.192
F2	-0.087	0.193	0.325	-0.203	-0.054
female	a	e	i	o	u
F1	0.254	0.233	-0.239	0.085	-0.120
F2	-0.053	0.135	0.169	-0.294	-0.081

3. ANALYSES AND RESULTS

3.1. Relationship between articulatory position and vowel duration

First, shift of F1 and F2 values was examined according to change of speaking rate. It has been widely observed that the segments of slow speech, with their long duration, become closer to the target in terms of articulatory position, while fast speech has short-duration segment sound and their articulation becomes ambiguous [7, 3, 11, 5]. Thus, it can be supposed that the direction of formant shift as a function of speaking rate shows the target articulation.

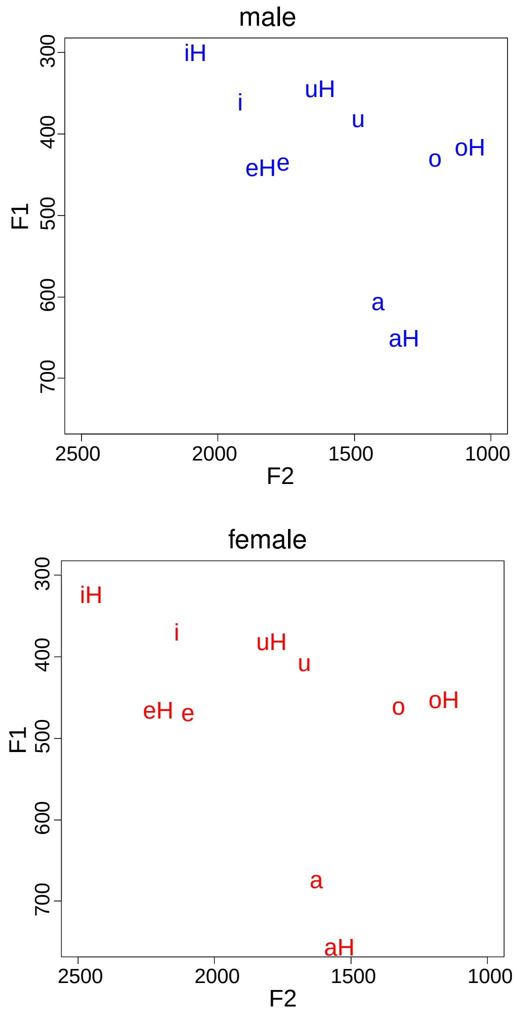
To test this assumption, correlation coefficients between formants and vowels of different duration were examined for short vowels /a/, /e/, /i/, /o/, and /u/.

Table 3 shows correlation coefficients between each formant and speaking rate by gender. All correlation coefficients were significant at the 5% level.

The analysis revealed that:

1. F1 increased as speaking rate slowed in the cases of /a/, and /e/, while for /i/ and /u/, the formant decreased as speech slowed.
2. In /e/ and /i/, F2 value increased as a function

Figure 1: Difference in formant frequencies between long and short vowels



of segment duration, while in /o/ the value decreased.

The results indicate that slow speech shows greater jaw openness in /a/ and /e/ and less openness in /i/ and /u/, and that tongue frontness increases in /e/ and /i/ and decreased in /o/ as a function of duration.

3.2. Difference in articulatory position between long and short vowels

The Japanese language has a phonemic long/short distinction. Since long vowels have longer duration, it was assumed that their articulatory positions will be closer to the targets than that of short vowels. To test this, the difference in formant frequencies between long and short vowels was examined.

Table 4 shows average F1 and F2 values of each vowels, and Figure 1 shows scatter diagrams for

Table 4: Average F1 and F2 values of each vowels and results of t-tests between short and long vowels

male F1	short	long	*p*
a	612.705	655.864	$p < 0.001$
e	441.547	446.956	0.373
i	366.562	305.362	$p < 0.001$
o	436.360	422.300	0.024
u	388.399	349.625	$p < 0.001$
male F2	**short**	**long**	***p***
a	1425.653	1344.985	$p < 0.001$
e	1771.912	1870.688	$p < 0.001$
i	1922.230	2101.465	$p < 0.001$
o	1216.104	1102.421	$p < 0.001$
u	1498.385	1653.117	$p < 0.001$
female F1	**short**	**long**	***p***
a	680.833	762.520	$p < 0.001$
e	476.031	471.548	0.594
i	375.423	329.116	$p < 0.001$
o	467.588	458.201	0.314
u	414.401	387.534	$p < 0.001$
female F2	**short**	**long**	***p***
a	1639.177	1567.687	$p < 0.001$
e	2107.474	2231.998	$p < 0.001$
i	2143.303	2471.657	$p < 0.001$
o	1336.980	1187.085	$p < 0.001$
u	1683.164	1816.875	$p < 0.001$

these values. In general, the tendency of formant shift is the same as that according to speaking rate change: F1 decreased in /iH/ and /uH/ and increased in /aH/, while F2 increased in /eH/, /iH/, and /uH/ and decreased in /aH/ and /oH/. A series of t tests revealed that these tendencies were significant at the 5% level (Table 4).

4. DISCUSSION

In most cases, the results of the analyses indicated that both slowly spoken and phonemically long vowels increase the peripherality (distinction, extremity) of their articulation. As shown in Table 3, slow speech has higher F1 values for /a/ and /e/ and lower F1 for /i/ and /u/. Also, F2 increases in slow /e/ and /i/ and decreases in slow /o/. These results indicate that as speaking rate decreases, open vowels get more open, close ones get more close, front ones get more front, and back ones get more back.

Similar tendencies were observed in the difference between long and short vowels. As shown in Figure 1 and Table 4, /aH/ increased in jaw openness, /eH/ and /iH/ in tongue frontness, and /oH/ in backness.

These results are consistent with findings for English vowels [3, 2, 9]. It thus seems likely that on the whole, crosslinguistically, the positional features of vowels articulated carefully in slow speech tend to become enhanced in this way.

However, the close back vowel /u/ showed a different tendency. Correlation coefficients between F2 and speaking rate were small (male: -0.054, female: -0.081): This result indicated that tongue backness did not necessarily increase according to decrease in speaking rate. Moreover, long vowel /uH/ had a greater F2 value than short vowel /u/ did, indicating that tongue backness was weakened in this long vowel.

Some previous descriptions of Japanese vowels have pointed out that Japanese /u/ is not a typical back vowel [10, 12, 6]. That is, as compared to more canonical back vowels like French /u/, the articulatory position of Japanese /u/ is more centered [10, 12, 6]. Further, the /u/ of standard Tokyo Japanese lacks strong lip rounding, and thus the vowel is commonly given not as [u] but as [ɯ] [10, 12]. Kubozono noted that the vowel weakens in backness according to decrease in lip rounding [6]. The results of the present study support these observations from a quantitative angle.

However, the results of this study were not consistent in one way with the study of Hirata and Tsukada [4]; in their research, the articulatory position of long vowel /uH/ did not shift to front. This difference can be accounted for by the different speech materials used between the two studies. Hirata and Tsukada's study was conducted under laboratory conditions, and speech materials were obtained by reading nonsense words embedded in carrier sentences. In this situation, it can be suspected that speakers might unconsciously be more careful and their articulations become more canonical or closer to target position.

In the present study, in contrast, speech materials were obtained from a corpus of spontaneous speech. It was assumed on this basis that speakers were not concentrating on articulation and as a result that the tokens produced were looser, that is, departed more from target position. This result indicates one advantage of the analysis of spontaneous speech for the exploration of speech variation.

5. CONCLUSION

To illustrate the articulation of Japanese vowels, this study examined the distribution of formant frequencies as a function of speaking rate and long/short contrast, using a large-scale spontaneous speech corpus. The results showed that most vowels enhanced their features according to length (as a result of either slow speech rate or phonemic length difference) but that the back vowel /u/ decreased its backness, showing that Japanese /u/ is not a typical back vowel. This tendency has not been clearly observed in laboratory phonetics, meaning that the results of the present study also show one advantage of analysis of a corpus of spontaneous speech over lab-elicited speech for our understanding of speech variation.

6. ACKNOWLEDGMENTS

This research was supported by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (24500675, 25280063, 2628130 and 26560320).

7. REFERENCES

[1] Chiba, T., Kajiyama, M. 1942. *The vowel: It's nature and structure*. Tokyo: Tokyo-Kaiseikan.

[2] Fourakis, M. 1990. Tempo, stress, and vowel reduction in American English. *Journal of Acoustical Society of America* 90(4), 1816–1827.

[3] Gay, T. 1978. Effect of speaking rate on vowel formant movement. *Journal of Acoustical Society of America* 63(1), 223–230.

[4] Hirata, Y., Tsukada, K. 2009. Effects of speaking rate and vowel length on formant frequency displacement in Japanese. *Phonetica* 66, 129–149.

[5] Imaizumi, S., Kiritani, S., Hirose, H., ichi Togami, S., Shirai, K. 1987. Preliminary report on the effects of speaking rate upon formant tranjectories. *Annual Bulletin Research Institute of Lagopedics and Phoniatrics* (21), 147–151.

[6] Kubozono, H. 1999. *Nihongo-no Onsei (Phonetics of Japanese)*. Tokyo: Iwanami-Shoten.

[7] Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of Acoustical Society of America* 35(11), 1773–1781.

[8] Maekawa, K. 2003. Corpus of Spontaneous Japanese: Its design and evaluation. *Proceedings of ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition* Tokyo. 7–12.

[9] Moon, S.-J., Lindblom, B. 1994. Interaction between duration, context, and speaking style in english stressed vowels. *Journal of Acoustical Society of America* 96(1), 40–55.

[10] Sakuma, K. 1929. *Nihon Onseigaku (Japanese Phonetics)*. Tokyo: Kyobun-Sha.

[11] Sonoda, Y., Nakakido, K. 1986. Effect of speaking rate on jaw monvements in vowel sequences. *The Journal of The Acoustical Society of Japan (E)* 7(1), 5–12.

[12] Vance, T. J. 1987. *An Introduction to Japanese Phonology*. Albany, New York: State University of New York Press.