Ann. Bull. RILP (1983) No. 17, 37–44

AN ANALYSIS OF THE DURATION OF THE JAW MOVEMENTS IN THE VOWEL SEQUENCE WORDS

Shinobu Masaki, Katsuhiko Shirai,* Shigeru Kiritani and Hiroshi Imagawa*

1. Introduction

In the study of the speech synthesis by rule, adequate control of the segmental durations is considered one of the important factors affecting the naturalness of the synthetic speech. There have been several studies in which variation of the segmental durations due to the type of the segment and/or the phonetic context were measured acoustically and formalated into rules for the speech synthesis. The pattern of these variation is influenced, at least in part, by the physical and physiological characteristics of the articulatory movements. Therefore, an analysis of the dynamic characteristics of the articulatory movements would be useful for clarifying the characteristics of the segmental duration in the connected speech.

In this study, movements of the jaw in the production of meaningless words consisting of vowel sequences were observed, and the duration of the jaw movements for individual vowels was measured. The influence of word length, word-internal position and the type of vowel on the duration of individual vowels was analyzed.

2. Data Recording

Figure 1 shows the experimental method used for observing the jaw movements. The details of the method are described elsewhere.[6]

The speech materials were 88 meaningless words consisting of the 4 Japanese vowels /a/, /i/, /u/ and /e/. (The Japanese vowel /o/ was omitted because of an occasional interference of the solid wire bearing the LEDs with the lip protrusion.) The set of words used were as follows;

a) 4 one-mora words /a/, /i/, /u/ and /e/.
b) 12 two-mora words in the form of /V_1 V_2/ ($V_1 \neq V_2$), such as /ai/ and /ia/.
c) 36 three-mora words in the form of /V_1 V_2 V_3/ ($V_1 \neq V_2$ and $V_2 \neq V_3$), such as /aia/ and /aiu/.
d) 24 four-mora words in the form of /V_1 V_2 V_3 V_4/ ($V_i \neq V_j$ where i,j = 1,2,3 and 4, i ≠ j), such as /aiue/ and /auie/.
e) 12 four-mora words in the form of /V_1 V_2 V_1 V_2/ ($V_1 \neq V_2$), such as /aiai/ and /auau/.

* Department of Electrical Engineering, Waseda University

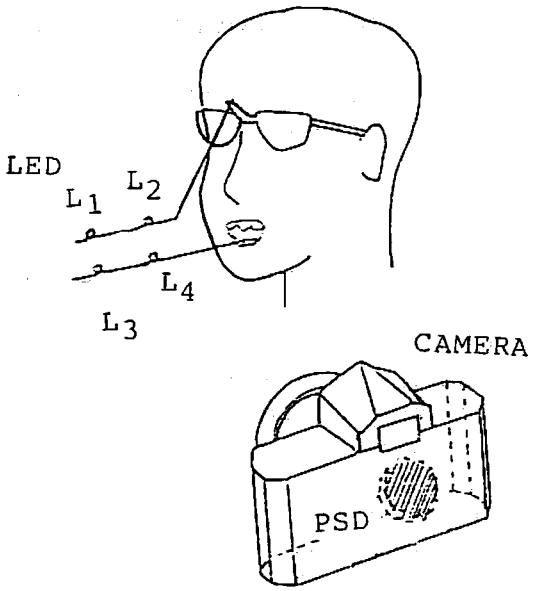

Fig. 1 *Experimental method for observing the movements of the jaw.*
PSD (Position Sensitive Detector)
LED (Infrared Light Emitting Diode)

Table 1 *List of Test Words*

a *	aia *	uai *	aiue	eaiu
i *	aiu	uau *	aieu *	eaui
u *	aie *	uae	auie	eiau *
e *	aua *	uia	auei *	eiua
	aui	uiu	aeiu	euai *
ai *	aue *	uie	aeui	euia
au *	aea	uea	iaue *	aiai *
ae	aei	uei *	iaeu	auau *
ia *	aeu	ueu *	iuae	aeae
iu	iai *	eai	iuea	iaia *
ie *	iau *	eau	ieau	iuiu
ua *	iae	eae	ieua *	ieie *
ui	iua	eia *	uaie *	uaua *
ue *	iui	eiu	uaei	uiui
ea	iue	eie *	uiae	ueue *
ei *	iea	eua *	uiea	eaea
eu *	iei *	eui	ueai	eiei *
	ieu *	eue *	ueia *	eueu *

* Words analyzed in the present study.

The test words were uttered within the carrier sentence / desu/. Each sentence was repeated 3 times in succession. For the test word /aia/, for example, the subject uttered [aia desu, aia desu, aia desu]. The data for the second sentence in the utterance was stored in the computer. The subject read the list of 88 words 3 times. Thus, 3 utterances for each test word per subject were stored in the computer. The subjects were two adult males.

In this preliminary study, the data of 44 out of the 88 words were analyzed in which open vowels and close vowels alternated with each other. These 44 words are marked with an asterisk in Table 1.

3. Method of Analysis

The vertical movement of the LED marked as L4 in Fig. 1 was analyzed. Figure 2 shows an example of the time function of the audio emvelope (top) and the vertical movement of the LED (bottom) for the utterance /auei desu/.

In the time function of the jaw movement, the time moments of maximum displacement were determined, and the time intervals between these peaks were measured. In the present study, this time interval was defined as the "duration" of the jaw movement for each vowel.

The duration of the word-initial vowel was defined as the period between the beginning of the utterance and the moment of the peak of the jaw movement for the word-initial vowel. When the word-initial vowel was a close vowel, the starting moment of the transition movement of the jaw for the second vowel was taken as the moment of the peak for the word-initial vowel.

When the word-final vowel was a close vowel, it was difficult to determine the moment of the peak of the jaw movement for this vowel because of the closing movement for the following /d/ of the next word in the sentence frame. Thus, the duration of the word-final close vowel could not be measured. The duration of the word-final vowel + /d/ was defined as the period between the peak moment for the penultimate vowel and that for /d/.

In addition, the period between the beginning of the utterance and the moment of the implosion of /d/ was taken as the duration of the test word.

In Fig. 2, t1 and t2 are the beginning of the utterance and the moment of the implosion of /d/, respectively; and p1, p2 and p3 are the peaks for the word-initial vowel (V_1), the second vowel (V_2), and the third vowel (V_3), respectively. In this case, the peak for the word-final vowel (V_4) could not be determined. Finally, p_d is the peak for /d/. The duration of each vowel and word were as follows.

The duration of V_1	/a/	: t_1 -p_1
The duration of V_2	/u/	: p_1 -p_2
The duration of V_3	/e/	: p_2 -p_3
The duration of V_4 + /d/	/i/ + /d/	: p_3 -p_d
The duration of the word		: t_1 -t_2

4. Results

4.1 The mean durations of the one-, two-, three- and four-mora words, and the mean durations of the vowels at each word-internal position

Figure 3 shows the mean durations of the one-mora, two-mora, three-mora and four-mora words, and the mean durations of the vowels for each word-internal position.

The duration of the words linearly increased as the number of moras within them increased.

The beginning of the utterance was almost the same as the starting moment of the transition movement from the word-initial vowel to the second vowel. Thus, the duration of the word-initial vowel, as defined in this study, was shorter than that of the word-medial vowels in three and four-mora words.

As for the word-medial vowels, the duration of the second vowel in three- and four-mora words were approximately the same. The duration of the third vowel in the four-mora words was slightly shorter than that of the second vowel in the three- and four-mora words. The difference was 10% at most.

As for the word-final open vowels, their duration was shorter than that of the word-medial vowels in the three- and four-mora words. However, the duration of the word-final open vowel in the two-mora words was longer than in the three- and four-mora words, and was almost the same as that of the word-medial vowels in the three- and four-mora words.

4.2 Variations in the duration of the individual vowels

Figure 4 shows the duration of each vowel /a/, /i/, /u/ and /e/. The duration for each vowel was measured separately at different positions within the words. In the figure, each data point represents the average duration over the test words containing a given vowel at a given position within the word. For example, the duration of the second vowel /a/ in the three-mora words was the average duration of /a/ in the test words /iai/, /iau/, /uai/ and /uau/, each test word having three utterance tokens.

It can be seen in the figure that, in the case of the word-medial vowels in the three- and four-mora words, the duration of the close vowels /i/ and /u/ was longer than that of the open vowels /a/ and /e/. This tendency was observed clearly for both subjects.

As for the word-initial vowels produced by Subject 1, the mean duration of /i/ was longer than that of the open vowels or /u/. For Subject 2, the duration of the open vowels /a/ and /e/ was nearly the same as in the case of Subject 1, and the duration of /i/ was shorter and the duration of /u/ was longer than that of the open vowels. For both subjects, the difference of duration among the vowels became smaller as the number of moras within the test words increased.

As for the duration of the word-final vowels + /d/, the duration of /a/ + /d/ and /e/ + /d/ was longer than that of /i/ + /d/ and /u/ + /d/. This difference became smaller as the number of moras within the words increased. In the case of the three- and four-mora words of Subject 2 in particular, this difference was not apparent. In the case of Subject 2, the duration of /u/ + /d/ was clearly longer than that of /i/ + /d/. This tendency for the duration of /u/ to be longer than that of /i/ was

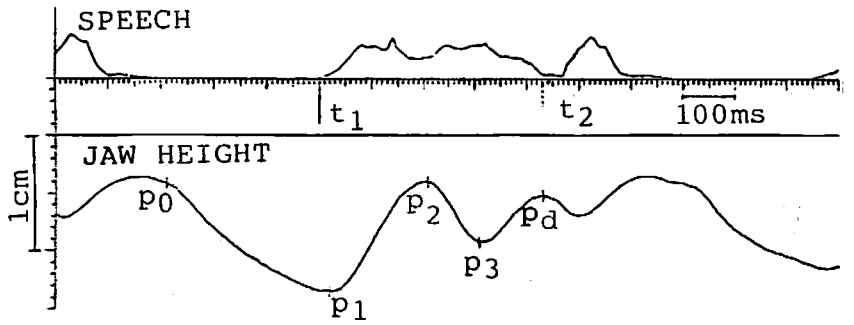

Fig. 2 *An example of the time function of the audio emvelope (top) and vertical movement of the jaw (bottom) for the test sentence /auei desu/.*

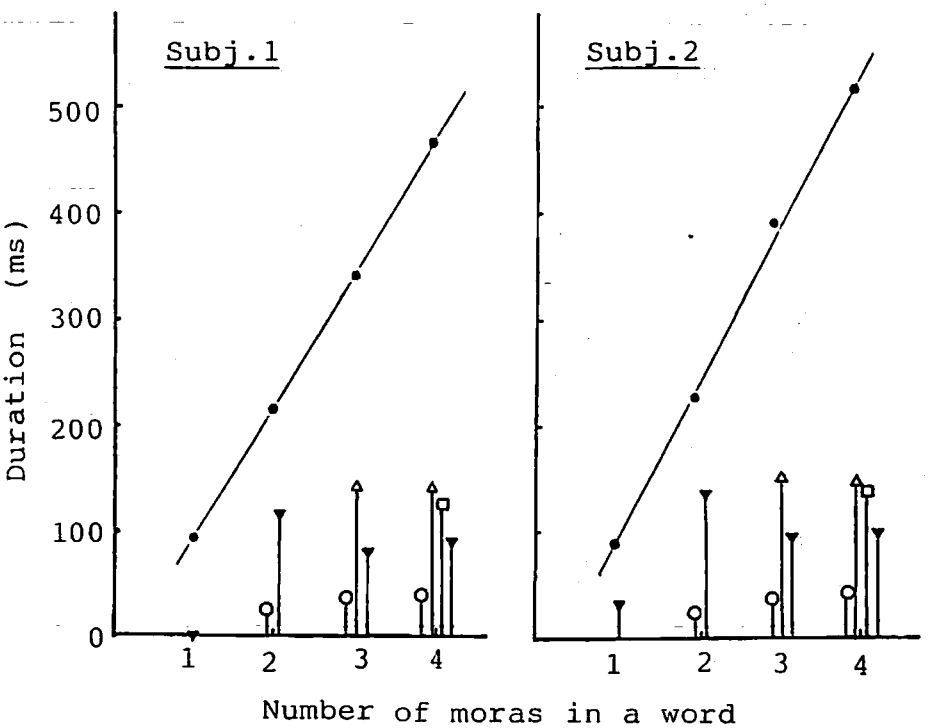

Fig. 3 *Duration of words and vowels.*

●: *word*
○: *word-initial vowel (V_1)*
△: *second vowel (V_2)*
□: *third vowel (V_3)*
▼: *word-final vowel (V_f) (measured only for open vowels)*

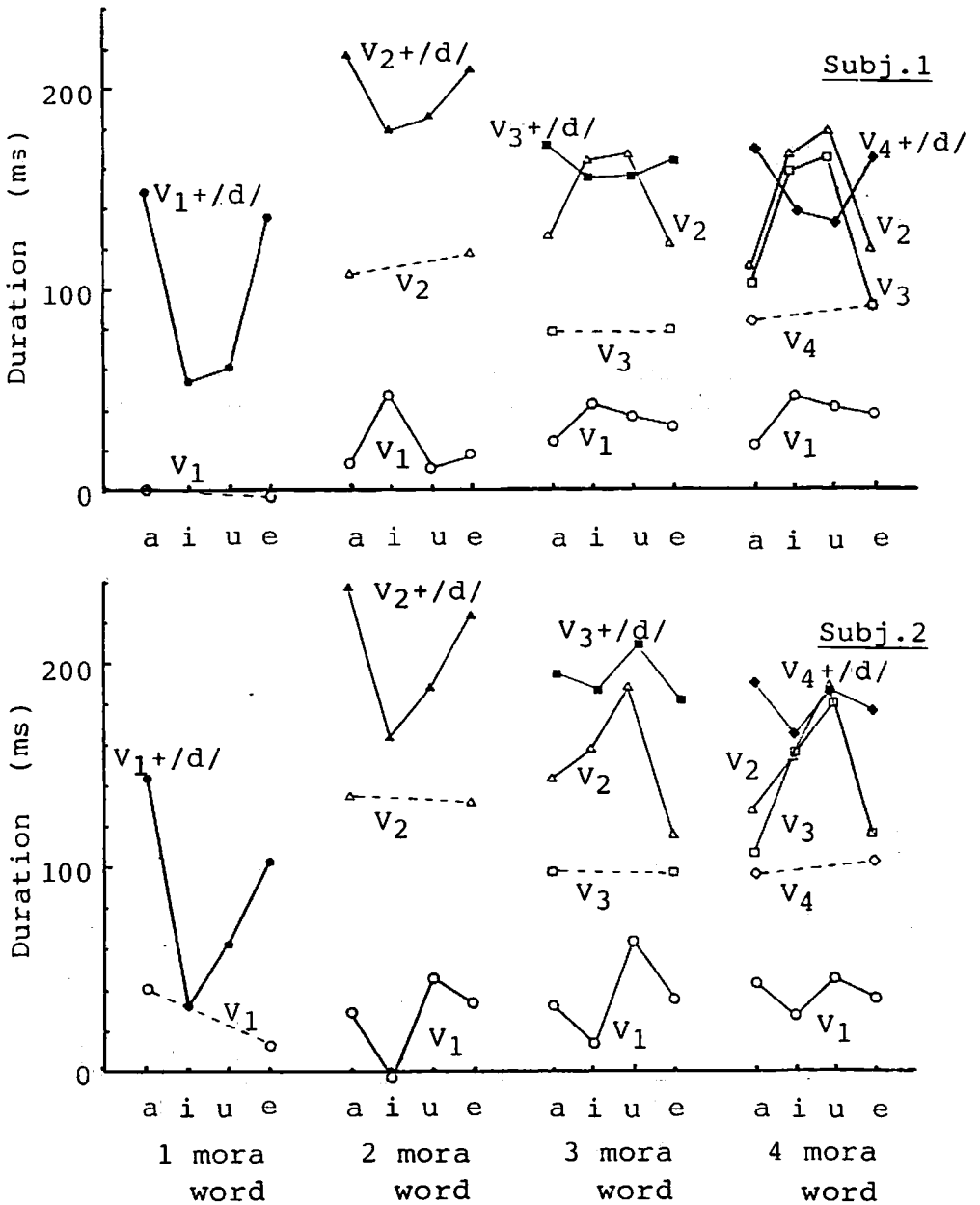

Fig. 4 *Duration of each vowel /a/, /i/, /u/ and /e/ at different positions in the test words.*

observed consistently for this subject at every position within the words.

5. Comment

In the present study, the movement of the jaw for the test words consisting of sequences of vowels were observed, and the duration of the jaw movement for each vowel was analyzed. It was observed that for the second and third vowel in word-medial position the transitional movement from open vowels to close vowels was longer than that from close vowels to open vowels. This phenomenon has been reported in several previous studies.[7-10] The analysis in this study confirms this phenomenon for a larger set of test words. More detailed analysis of the characteristics of this difference between the opening and closing movements of the jaw are being performed in another study.[11]

There was a difference between the two subjects in terms of the relative duration of the two close vowels. In Subject 2, the duration of the jaw movement for /i/ was shorter than the duration for /u/. In Subject 1, on the other hand, the difference between /i/ and /u/ was small. Thus, for Subject 2, the difference between the duration of /i/ and the duration for the open vowels was smaller than in the case of Subject 1.

For the word-final vowels, the time interval between the moments of the two peak displacements of the penultimate vowel and /d/ (in /desu/) was longer in the case of the open vowels than for the close vowels. This difference can be considered to be due to differences in the required amount of displacement from different word-final vowels to /d/.

Further analysis is being performed on the temporal pattern of the jaw movements in test words containing the semivowels /j/ and /w/ and long vowels.

References

1. Klatt, D.H. (1975); Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence. J. Acoust. Soc. Amer., 59, 1208-21.
2. Sato, T. (1977); Segmental Duration and Timing Location in Speech. Trans. Commit. Speech Research, Acoust. Soc. Japan, S77-31.
3. Higuchi, H. and H. Fujisaki (1978); Temporal Organization of Segmental Features in Two-mora Words Consisting of Vowels. Trans. Commit. Speech Research., Acoust. Soc. Japan, S78-37, 275-290.
4. Higuchi, N. and H. Fujisaki (1980); Durational Control of Segmental Features in Connected Speech. Trans. Commit. Speech Research, Acoust. Soc. Japan, S80-40, 315-321.
5. Sagisaka, Y. and Y. Tohkura (1981); Rule of Segmental Durations Using Statistical Features of Segment. Trans. Commit. Speech Research, Acoust. Soc. Japan, S80-72, 561-568.
6. Kiritani, S., T. Tanaka, K. Hashimoto, S. Masaki and K. Shirai (1983); The Movement of the Jaw in the Production of the VCV Sequence. Trans. Commit. Speech Research., Acoust. Soc. Japan, S82-105, 835-842.
7. Kim, B. and H. Fujisaki (1974); Analysis of Jaw Control Characteristics in Connected Vowels. Report of the Autumn Meeting of Acous. Soc. Japan, 317-318.
8. Sonoda, Y. (1979); Articulatory Characteristics of Tongue and Jaw Point Movements in Connected Sounds of Japanese. Trans. IECE Japan, J62-A, No. 9, 555-562.

9. Kiritani, S., H. Imagawa, T. Takahashi, S. Masaki and K. Shirai (1982); Temporal Characteristics of the Jaw Movements in the Production of Connected Vowels, Ann. Bull. RILP, 16, 1-10.
10. Sekimoto, S., H. Imagawa and S. Kiritani (1978): Dynamic Characteristics of Tongue Movement in the Production of Connected Vowels. Ann. Bull. RILP, 12, 11-20.
11. Imagawa, H., S. Kiritani, S. Masaki and K. Shirai (1983): Comparison of velocity and duration between open to close vowel transitions and close to open vowel transtion. Ann. Bull. RILP, No. 17, 33-36