Ann.Bull.RILP(1985)No.19, 29-46

DIFFERENCES IN JAW OPENINGS FOR VOWELS DUE TO SPEAKING RATE AND WORD-INTERNAL POSITION IN THE PRODUCTION OF VOWEL SEQUENCE WORDS

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

1. Introduction

There have been some studies on the dynamic characteristics of jaw movements in speech [1)-4)]. In the production of vowel sequences, a difference in the jaw opening for the same vowel is observed for different speaking rates and different word-internal vowel position [5)]. Possible reasons for the different jaw openings are supposed to be differences in the target of the jaw opening and/or the duration for each vowel.

In this study, the reasons for the difference in the jaw openings for vowels are discussed. The jaw movements in the production of vowel sequences at fast and slow speaking rates were observed. At first, as a basis of analysis for vowel sequence words, the jaw opening for each Japanese vowel and the time constant for vowel transitions in the production of sequences of stationary vowels were measured. Secondly, the jaw opening and the duration for each vowel were observed. Thirdly, the target of the jaw opening for each vowel were estimated. In this estimation, the jaw movements were approximated by the response of the second order system driven by a sequence of the step functions, each of whose levels was estimated as the target of the jaw opening for each vowel.

2. Data collection

The jaw movements analyzed in the following sections were recorded using a system previously described, which was composed of LEDs and an optical spot position sensitive detector [6)].

Table 1 shows the speech samples analyzed in this study. The speech samples were 36 meaningless three-mora words consisting of the 4 Japanese vowels /a/, /e/, /i/ and /u/. The Japanese vowel /o/ was omitted because of occasional interference from the solid wire bearing the LEDs with the lip protrusion.

Each of the test words was uttered within the carrier sentence /_ desu/. Each sentence was repeated three times in succession. For the test word /aia/, for example, the subject uttered /aia desu, aia desu, aia desu/. Data for the second sentence in the utterance was stored in the computer.

----

* Tokyo Metropolitan Institute of Gerontology
** Waseda University

The vowel sequences shown in Table 1 were also uttered as sequences of three stationary vowels within the carrier sentence /_ desu/. In this case, in order to maintain a constant speaking rate, a sequence of click sounds was presented before every 18 speech samples. The interval between the click sounds was 0.625 s, corresponding to the duration of each vowel.

Four utterances were sampled and stored for each speech sample. The subjects were two adult males.

Figure 1 shows a superposition of the observed movements of LEDs L_3 and L_4 on the X-Y plane in the production of 36 speech samples uttered as three-mora vowel sequence words. The movements of LEDs L_3 and L_4 do not show any significant curvature. The movement on the Y coordinate of LED L_4 was analyzed to be the jaw movement.

3. Jaw opening for each vowel and the time constant for each vowel transition in the production of sequences of three stationary vowels

3-1. Method of analysis

As a basis for the analysis of the vowel sequence words, the jaw opening for each vowel and the time constant for each vowel transition in the production of three stationary vowels were measured.

Figure 2 shows an example of the time function of the jaw opening in the production of sequences of stationary vowels. The curve shown at the top is the audio envelope, and that at the bottom is the observed jaw movement (solid curve).

The jaw opening for each vowel was determined as follows. In the time function of the jaw movement, the stationary portion for each vowel was clearly vesible. At the middle of this portion, the jaw openings within 50 ms were averaged as the jaw opening for each vowel. In Fig. 2, t_i (i=1, 2 and 3) represents the midpoint of the period in which the jaw opening for the stationary vowel was measured.

The time constant for each transition in the production of the sequences of three stationary vowels was also measured. There were two vowel transitions in the time function of the jaw opening in the production of one speech sample; one was the transition from the first to the second vowel, and the other from the second to the third vowel. The jaw movements in these transitions were approximated by the response of a critically damped linear second order system driven by the step function, and the time constant of the transition which gave the best approximation for each vowel transition was estimated.

Table 1 Speech samples

aia	iai	uai	eai
aiu	iau	uau	eau
aie	iae	uae	eae
aua	iua	uia	eia
aui	iui	uiu	eiu
aue	iue	uie	eie
aea	iea	uea	eua
aei	iei	uei	eui
aeu	ieu	ueu	eue

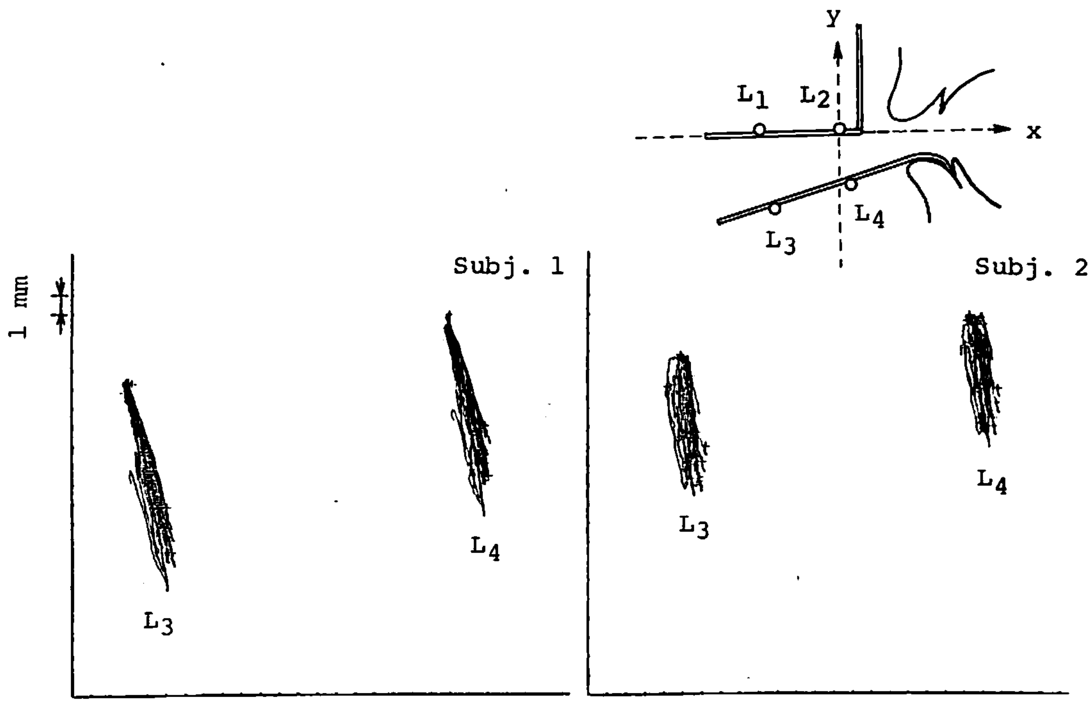

Fig. 1 Trajectories of LEDs for the lower jaw relative to the upper jaw in the production of three-mora vowel sequence words.

The time constant τ and the moment of step T_i (i=1 and 2) were determined so as to minimize error $E(\tau,T_i)$ given by

$$E(\tau,T_i)=\sum_{t=t_i}^{T_i}\{J_i-F(t)\}^2+\sum_{t=T_i}^{t_{i+1}}\{f(J_i,V_i,J_{i+1},\tau,t-T_i)-F(t)\}^2. \qquad (1)$$

In Eq. (1), $F(t)$ is the observed jaw movement, and $f(J_i,V_i,J_{i+1},\tau,t)$ is determined from Eq. (2),

$$f(J_i,V_i,J_{i+1},\tau,t)$$
$$=[\{V_i-\frac{1}{\tau}(J_{i+1}-J_i)\}t-(J_{i+1}-J_i)]e^{-\frac{t}{\tau}}+J_{i+1}, \qquad (2)$$

which represents the response of the critically damped second order system, whose time constant is τ, driven by the step function for the case where the initial level, the initial velocity and the target level are J_i, V_i, and J_{i+1}, respectively. Here J_i (i=1, 2 and 3) is the jaw opening for the i-th vowel measured at the moment t_i and $V_i=0$ (i=1 and 2).

Actually, the moment of the step was first determined by visual inspection. In order to get the best approximation of the time function of the jaw movement, the moment of the step was automatically searched for in a 50 ms range before and after the determined moment in 10 ms steps, and the time constant was sought between 10 ms and 500 ms in 10 ms steps.

For some vowel transitions, since the difference between the jaw openings for vowels preceded by the transition and those followed by the transition was small, the optimum time constant of the second order system was not found in the range between 10 ms and 500 ms. This was the case for the open to open vowel transitions (/ae/ and /ea/) and for the close to close vowel transitions (/iu/ and /ui/). The time constant for such cases will not be discussed in this study.

At the bottom of the example in Fig. 2, the broken curve shows the approximated jaw movement, and each vertical broken line indicates the moment of each step T_i (i=1 and 2).

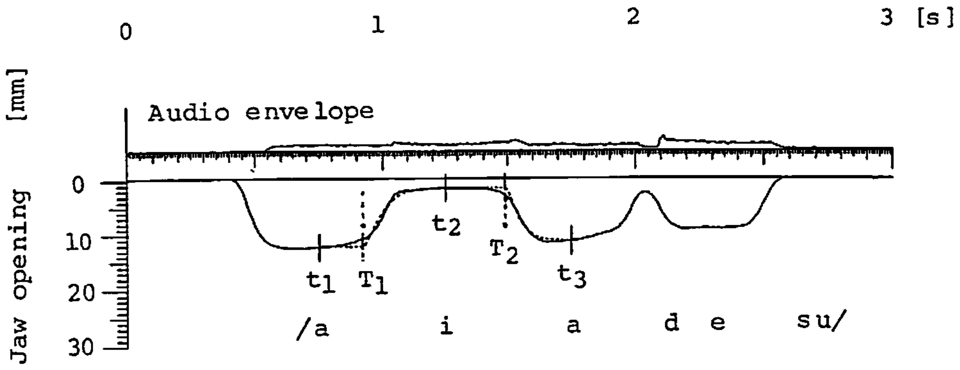

Fig. 2 An example of the time functions of the audio envelope and the jaw opening in the production of a sequence of stationary vowels.
At the bottom of the example, the observed jaw movement (solid curve) and the approximated jaw movement (brokem curve) are shown.

t_i: moment at which the jaw opening for each vowel was measured (i=1, 2 and 3)
T_i: starting moment of the step function for the next vowel (i=1 and 2)

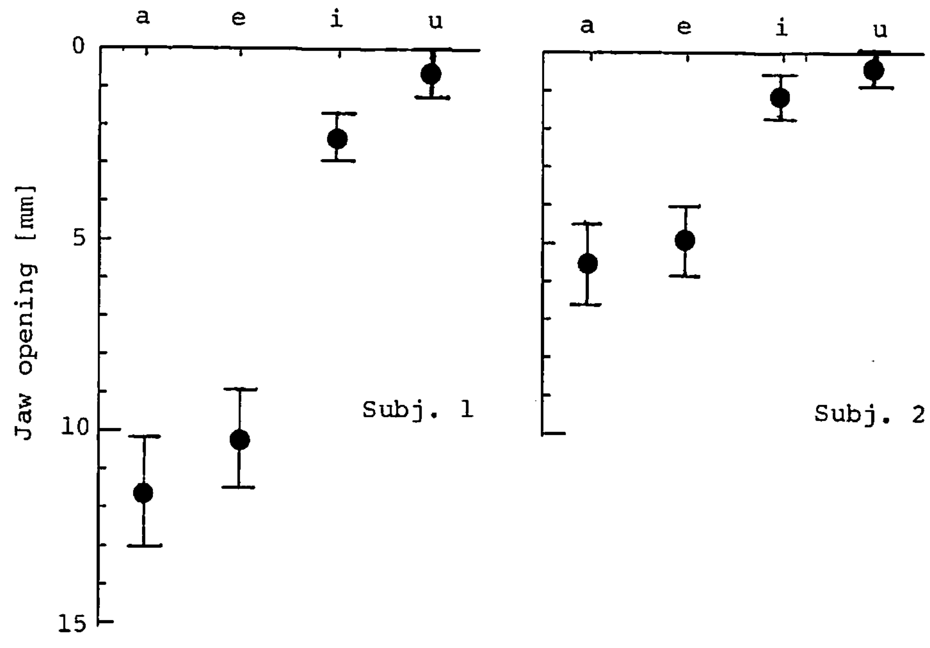

Fig. 3 Jaw opening for each vowel in the production of sequences of stationary vowels.

3-2. Results

3-2-1. Jaw opening for each vowel

Figure 3 shows the jaw opening for each vowel in the production of sequences of stationary vowels. The solid circle and the vertical line represent the average and the standard deviation of the jaw opening for each vowel, respectively.

The jaw opening for /a/ was the greatest, with those for /e/, /i/ and /u/ following in order. Part of the data for /a/ and /e/, and for /i/ and /u/, shows a common jaw opening.

Figure 4 shows the influence of word-internal position on the jaw opening for each vowel. The average jaw openings for word-initial, -medial and -final vowels are shown by the solid circle, triangle and square, respectively.

For /a/, /e/ and /i/, the jaw opening for word-final vowels was less than that for word-initial and -medial vowels. For /u/, a difference in jaw opening caused by the different word-internal positions was not observed.

3-2-2. Time constant for each vowel transition

Figure 5 shows the average values of the time constant for the open to close vowel transitions and the close to open vowel transitions.

For both subjects, the time constant for the open to close vowel transitions (/ai/ and /eu/, etc.) was greater than that for the close to open vowel transitions (/ia/ and /ue/, etc.) both from the first to the second vowel and from the second to the third vowel. For subject 1, the average time constant for the open to close vowel transitions was 36.0 ms and that for the close to open vowel transitions was 30.6 ms. For subject 2, the respective average time constants were 35.4 ms and 21.7 ms.

4. Jaw opening and duration for each vowel in the production of three-mora vowel sequence words

4-1. Method of analysis

Figure 6 shows examples of the time function of the jaw opening in the production of three-mora vowel sequence words. In each example, the curve shown at the top is the audio envelope, and that at the bottom is the observed jaw movement (solid curve). In the figure, t_0 and t_d, determined by visual inspection of the audio envelope, indicate the moment of the voice onset and the explosion of /d/ in /desu/, respectively.

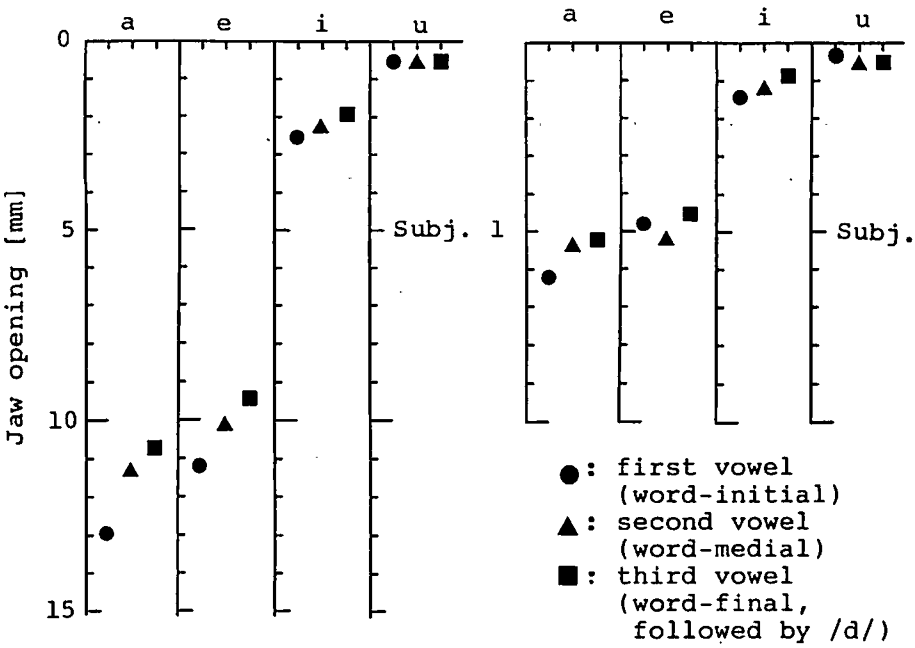

Fig. 4 Jaw opening for each vowel at each position in the production of sequences of stationary vowels.

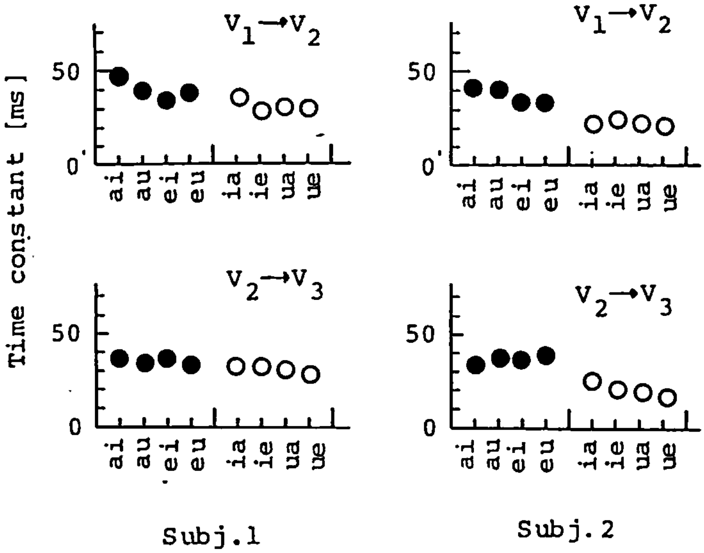

Fig. 5 Estimated time constants for vowel transitions in the production of sequences of stationary vowels.

In these time functions, the moment for measuring the jaw opening for each vowel was determined as follows. If a peak for a vowel was seen in the time function of the jaw opening, the moment of the peak was taken as the moment for measuring the jaw opening for the vowel. Otherwise, the moment at which the angle of the slope changed was selected as the moment for measuring the jaw opening for the vowel.

In Fig. 6, the moment at which the i-th vowel in a word was measured is shown as t_i (i=1, 2 and 3). The duration of the i-th vowel was determined as the period from t_i to t_{i-1} (i=1, 2 and 3). In Fig. 6, t_1, t_2 and t_3 in (a), for example, were determined by peaks, and those in (c) were determined by the moments at which the slope changes took place.

4-2. Results

4-2-1. Jaw opening for each vowel

Figure 7 shows the jaw opening for each vowel. The solid circle and the vertical line represent the average and standard deviation of the jaw opening for each vowel, respectively.

The jaw opening for /a/ was the greatest, with those for /e/, /i/ and /u/ following in order. Part of the data for /a/ and /e/, and for /i/ and /u/, shows a common jaw opening.

Figure 8 shows the influence of word-internal position on the jaw opening for each vowel. The average jaw openings for word-initial, -medial and -final vowels are shown by the solid circle, triangle and square, respectively.

For /a/ and /e/, the jaw opening for word-final vowels was less than that for word-initial and -medial vowels. For /i/ and /u/, variations due to different word-internal positions were not seen.

4-2-2. Duration for each vowel

Figure 9 shows the measured duration for each vowel at each word-internal position. The solid circle and vertical line represent the average and standard deviation of the duration for the vowel at each word-internal position, respectively.

The duration for word-final vowels was shorter than that for word-medial vowels. As shown in Fig. 6, the jaw movement for word-initial vowels begins prior to the onset of voicing. However, in this study, since the duration for a word-initial vowel was determined as the period from the onset of voicing to the moment at which the jaw movement starts for the next vowel, the measured duration for word-initial vowels was much shorter than for word-internal vowels.

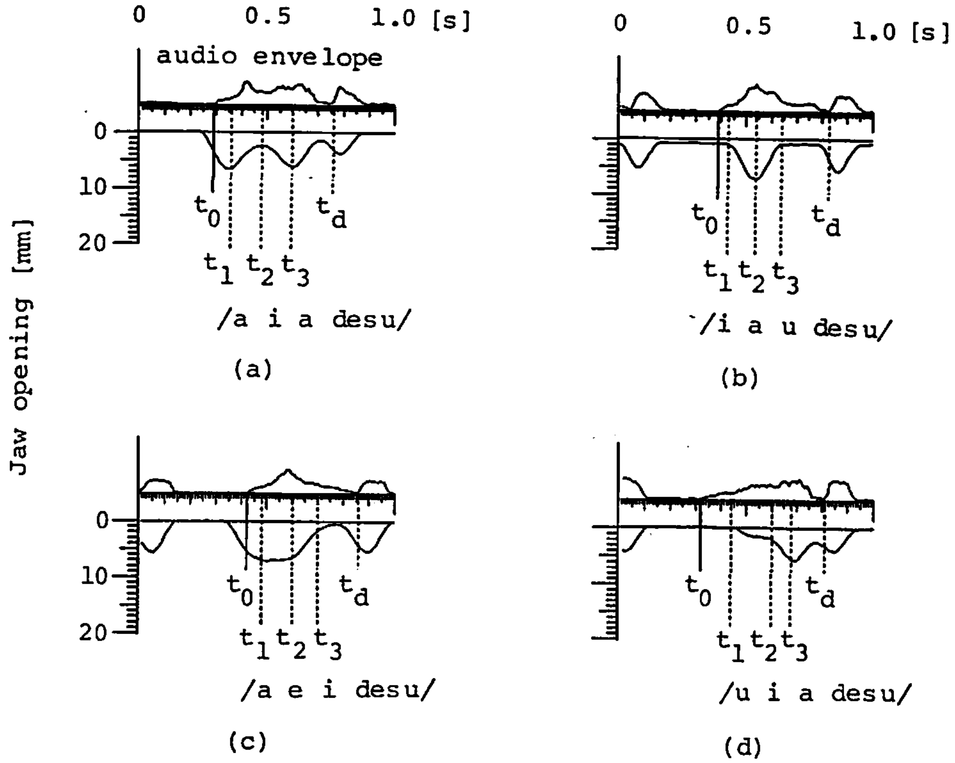

Fig. 6 Examples of the time function of the audio envelope and the jaw opening in the production of vowel sequence words.

t_0: moment of the voice onset
t_i: moment at which the jaw opening for i-th vowel is measured (i=1, 2 and 3)
t_d: moment of /d/ explosion in /desu/

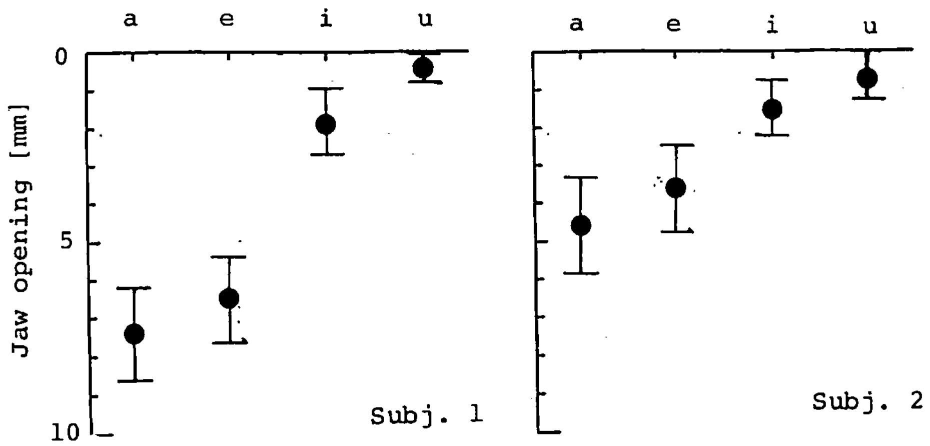

Fig. 7 Jaw opening for each vowel in the production of vowel sequence words.

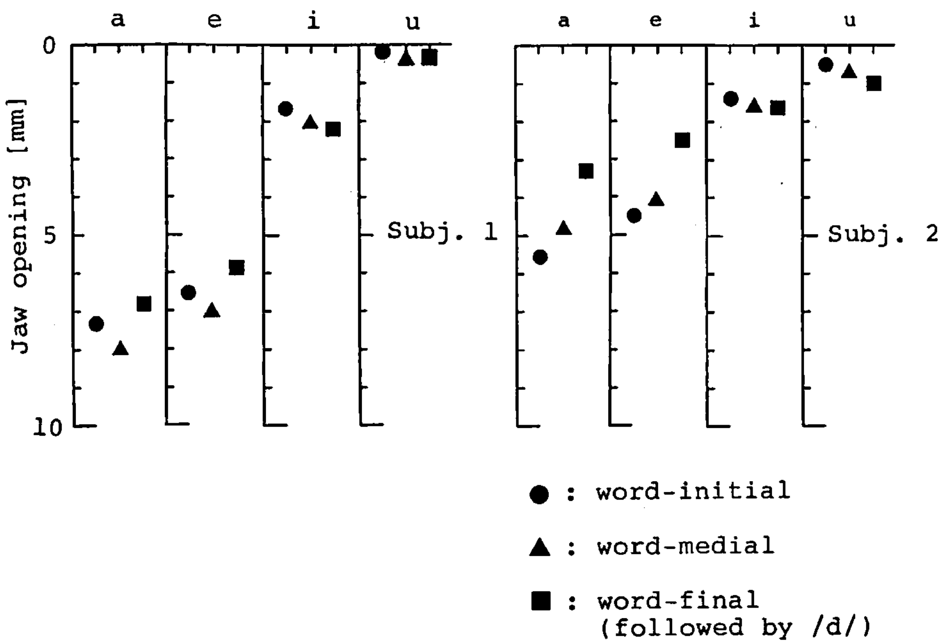

Fig. 8 Jaw opening for each vowel at each word-internal position in the production of vowel sequence words.

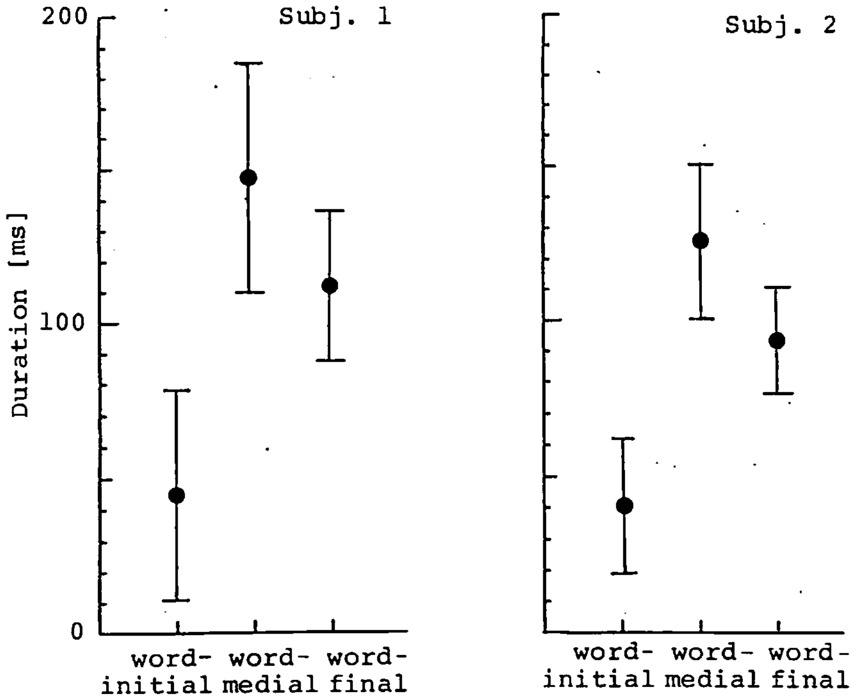

Fig. 9 Measured duration of each vowel in the production of vowel sequence words.

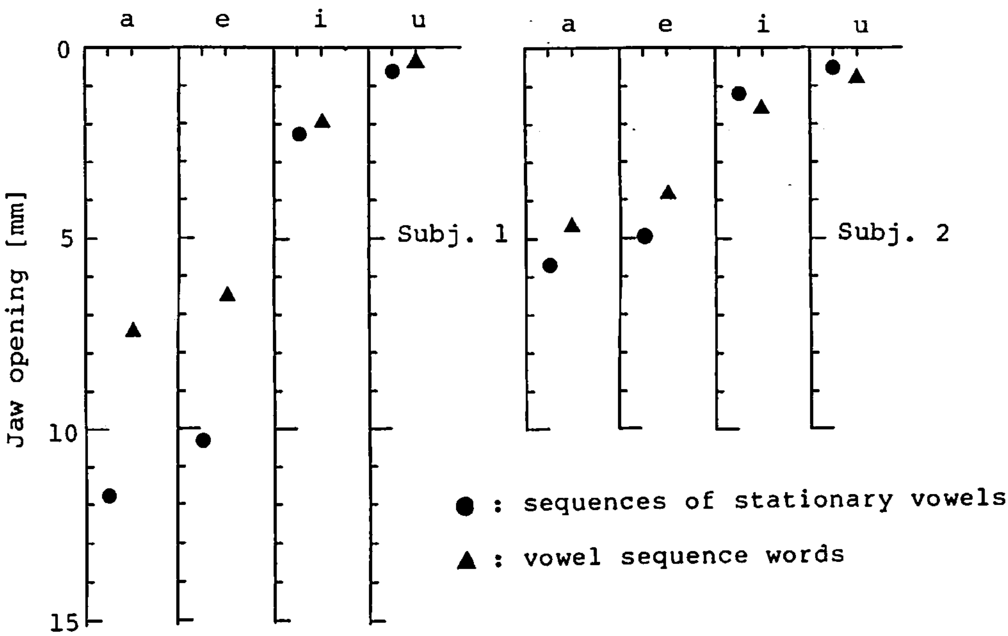

Fig. 10 Comparison of the jaw opening for each vowel between sequences of stationary vowels and vowel sequence words.

5. Target of jaw opening in the production of three-mora vowel sequence words

As shown in the previous section, the duration of the jaw movement for word-final vowels was shorter, and, at the same time, the jaw opening for word-final open vowels was less than for word-initial or -medial open vowels. Thus, the possible reason for the difference in jaw openings between the different positions is the under-shoot caused by the shortness of the duration.

Figure 10 shows a comparison of the jaw openings for each vowel between sequences of stationary vowels and vowel sequence words. The solid circle and triangle represent the average jaw opening in each vowel for sequences of stationary vowels and vowel sequence words, respectively.

For /a/ and /e/, the jaw opening for each vowel in the production of vowel sequence words was less than that in the production of sequences of stationary vowels. However, for /i/ and /u/, the differences in jaw opening between vowel sequence words and sequences of stationary vowels were small.

The duration for each vowel in the production of vowel sequence words was shorter than that in the production of sequences of stationary vowels. Thus, a possible reason for the variations in jaw opening in different speaking rates is also the under-shoot caused by the shortness of the duration.

However, there remains the possibility that the difference in the target of the jaw opening is an additional reason for the variation in jaw opening in different speaking rates and different word-internal vowel positions. In order to clarify these possibilities, examination of the target of the jaw opening for each vowel in the production of vowel sequence words is described in the following sections.

5-1. Method of analysis

The time function of the jaw opening in the production of vowel sequence words was approximated by the response function of a critically damped linear second order system driven by input step function. The target of the jaw opening for each vowel was determined by the input step level for each vowel, and the duration was determined by the period between the steps.

The target of the jaw opening and the duration were estimated by means of dynamic programing. The time moment and the level of each step for each vowel T_{i-1} and L_i (i=1, 2 and 3) and those for /d/ following the word-final vowels T3 and L4 were determined so as to minimize the error $E(T_0,T_1,T_2,T_3,L_1,L_2,L_3,L_4)$ which was determined as follows:

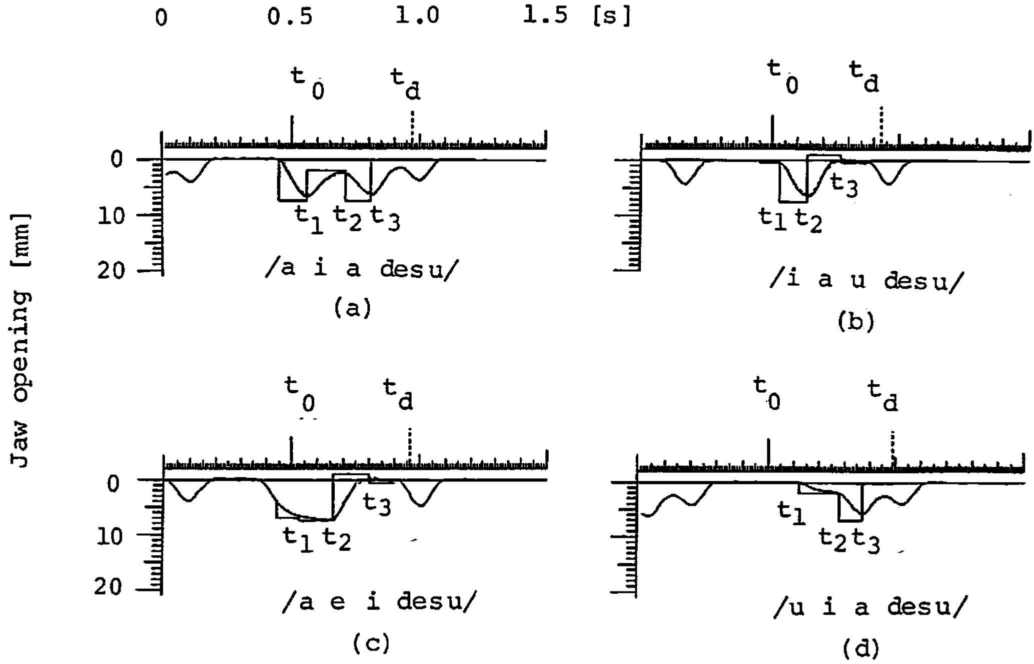

Fig. 11 Examples of the time functions of the jaw opening in the production of vowel sequence words.
In each example, the observed jaw movement (solid curve), the approximated jaw movement (broken curve) and the step function (rectangular solid line) are shown.

t_0: moment of the voice onset
t_i: moment of step for the i+1 -th vowel
t_d: moment of /d/ explosion in /desu/

$$E(T_0, T_1, T_2, T_3, L_1, L_2, L_3, L_4) = \sum_{i=0}^{3} \left[\sum_{t=T_i}^{T_{i+1}} \{f(J_i, V_i, L_{i+1}, \tau, t-T_i) - F(t)\}^2 \right]. \quad (3)$$

In Eq. (3), $F(t)$ is the observed jaw movement, and $f(J_i, V_i, L_{i+1}, \tau, t)$ is determined from Eq. (2); where J_0 and V_0 are the jaw opening and velocity at T_0 in the observed jaw movement, and J_i and V_i (i=1, 2 and 3) are the jaw opening and velocity at T_i in the approximated jaw movement. J_i and V_i (i=1, 2 and 3) are determined from the following equations:

$$J_i = f(J_{i-1}, V_{i-1}, L_i, \tau, T_i - T_{i-1}),$$
$$V_i = \frac{d}{dt} f(J_{i-1}, V_{i-1}, L_i, \tau, t - T_{i-1}) \Big|_{t=T_i}. \quad (4)$$

In Eq. (3), T_4 is 6 ms before the explosion for /d/, and the time constant τ of the second order system is determined as the average value over all open to close vowel transitions and all close to open vowel transitions. The time constants are 33.3 ms and 28.6 ms for subjects 1 and 2, respectively.

Actually, the moment of the step for each phoneme was first determined by visual inspection. In order to obtain the best approximation of the time function of the jaw movement, the moment of each step was searched for in a range 30 ms before and after the determined starting moment in 10 ms steps, and the level of the step was also searched for between 50 mm below and above the closed jaw position in 0.5 mm steps.

The duration for the i-th vowel was determined as the period from t_i to t_{i-1} (i=1, 2 and 3), where $t_i (=T_i)$ (i=1, 2 and 3) was the moment of the step for the i-th vowel and t_0 was the moment of the voice onset as determined by a visual inspection of the audio envelope.

Figure 11 shows examples of the observed jaw movement (solid curve), the approximated jaw movement (broken curve) and the estimated step function (solid rectangular lines).

5-2. Results

Figure 12 shows the target of the jaw opening for each vowel. The solid circle and vertical line represent the average and standard deviation of the jaw opening for each vowel, respectively.

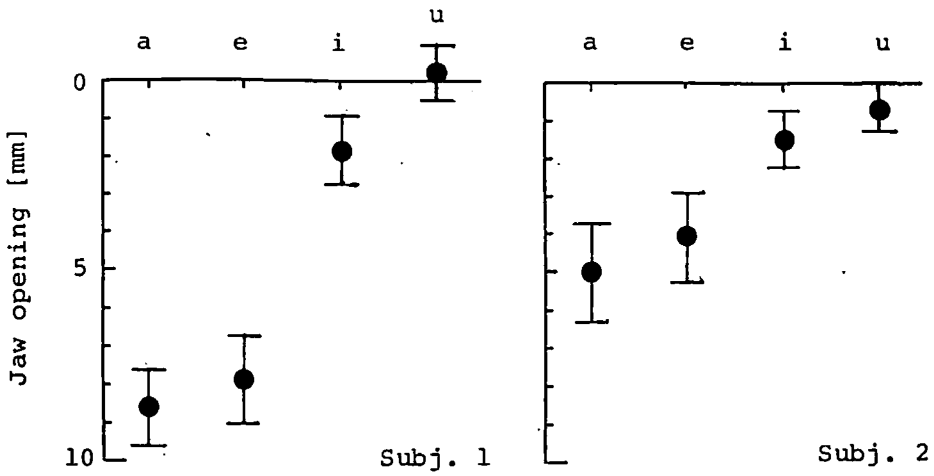

Fig. 12 Target of the jaw opening for each vowel in the production of vowel sequence words.

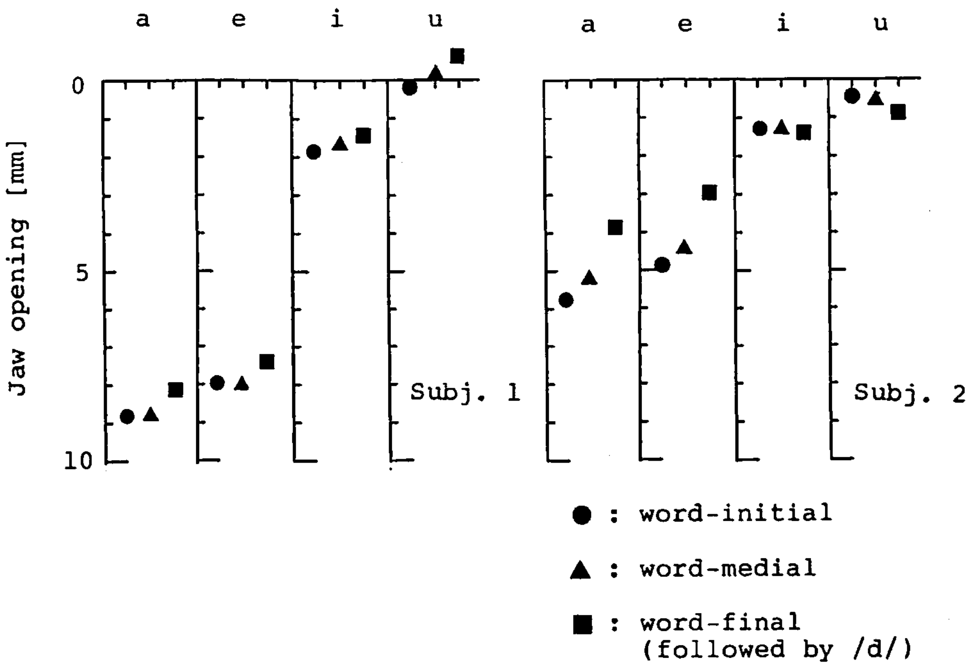

Fig. 13 Target of the jaw opening for each vowel at each word-internal position in the production of vowel sequence words.

The target of the jaw opening for /a/ was the lowest, while those for /e/, /i/ and /u/ followed in order. Part of the data for /a/ and /e/, and for /i/ and /u/, shows a common target for the jaw opening.

Figure 13 shows the target of the jaw opening for each vowel at each word-internal position. The solid circle, triangle and square represent the average values for word-initial, -medial and -final vowels, respectively.

For /a/ and /e/, the target of the jaw opening for word-final vowels was higher than that for word-initial or -medial vowels.

For /i/ and /u/, variations caused by the different word-internal positions were not seen except for /u/ in the case of subject 1. In this case, the target of the jaw opening for word-final /u/ was estimated as a position higher than the closed jaw position. This phenomenon may have had the following cause. For subject 1, the observed jaw opening for /u/ and /d/ of /desu/ was at the closed jaw position. Since the targets of the jaw opening for /u/ and /d/ were combined into one target level, the estimated duration for /d/ was lengthened; thus, the duration of the word-final /u/ became shorter. In order for the jaw opening to reach the closed jaw position in this short interval, the target of the jaw opening for the word-final /u/ had to be higher than the closed jaw position.

6. Discussion

As was seen in the previous section, for /a/ and /e/, the target of the jaw opening for word-final vowels, followed by the /d/ of /desu/, was higher than that for word-initial or -medial vowels. This suggests that the differences in jaw openings between word-final vowel and word-initial or -medial vowels were caused not only by the under-shoot due to the shortness of the durations for the vowels but also by variations in the targets of the jaw opening.

Figure 14 shows the average target of the jaw opening for each vowel in the production of vowel sequence words. The broken line represents the average, with the solid circle and triangle designating the average actual jaw opening for each vowel in the production of sequences of stationary vowels and vowel sequence words, respectively.

In a comparison between the target and the actual level of the jaw opening for open vowels (/a/ and /e/) in the production of vowel sequence words, the former was lower than the latter. In addition, comparing the target of the jaw opening for these vowels in the production of vowel sequence words and the observed jaw opening for these vowels in the production of sequences of stationary vowels, the former was higher than the latter. The observed jaw opening for sequences of stationary vowels is

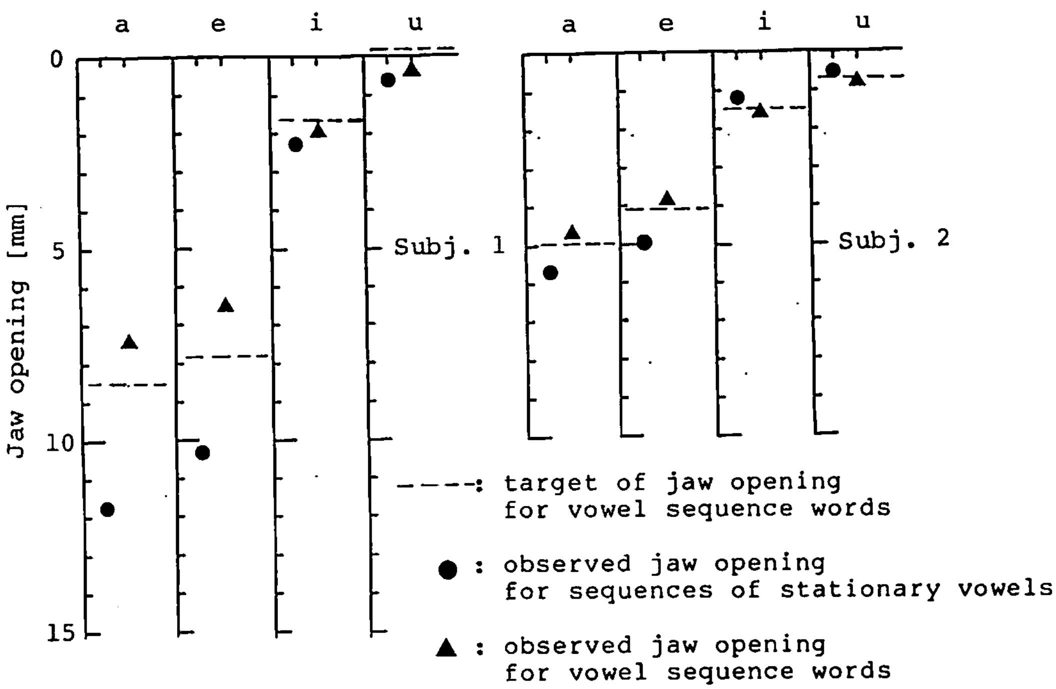

Fig. 14 Comparison of the target of the jaw opening and the observed jaw opening.

probably identical to the target of the jaw opening, since the duration of such vowels is long enough to reach the target. Thus, these results suggest that the differences in jaw opening between the two speaking rates were caused not only by the under-shoot due to the shortness of the duration of the vowel but also by variations in the target of the jaw opening.

References

1. Fujisaki, H. and B. Kim (1974); Observation of Jaw Control in the Articulation of Vowels. Transactions of the Committee on Speech Research, Acoust. Soc. Jpn., S74-08.
2. Sonoda, Y. (1979); Articulatory Characteristics of Tongue and Jaw Point Movements in Connected Sounds of Japanese. Trans. IECE Jpn., J62-A, 9, 555-562.
3. Masaki, S., K. Shirai, S. Kiritani and H. Imagawa (1983); Analysis of the Duration of the Jaw Movements in the Vowel Sequence Words. Ann. Bull. RILP, 17, 37-44.
4. Sonoda, Y. and K. Nakakido (1984); Effect of Speaking Rate on Jaw Articulation in Vowel Sequences. Transactions of the Committee on Speech Research, Acoust. Soc. Jpn., S83-97.
5. Masaki, S., K. Shirai, H. Imagawa and S. Kiritani (1984); Jaw Opening and the Time Constant of Jaw Movements in the Production of Sequences of Stationary Vowels. Ann. Bull. RILP, 18, 33-44.
6. Kiritani, S., T. Tanaka, K. Hashimoto, S. Masaki, and K. Shirai (1983); Contextual Variation of the Jaw Movement for the Intervocalic Consonant in VCV Utterances. Ann. Bull. RILP, 17, 45-53.