Ann. Bull. RILP, No. 12, 11-20 (1978)

DYNAMIC CHARACTERISTICS OF TONGUE MOVEMENT IN THE PRODUCTION OF CONNECTED VOWELS

S. Sekimoto, H. Imagawa and S. Kiritani

Introduction

In the present study, the dynamic characteristics of tongue movement in the production of $V_1V_2V_1$ words in Japanese were investigated. The movement of the tongue was represented by a linear superposition of two basic articulatory components. The temporal change of the individual components was approximated by a step response of a critically-damped second order linear system, and the time constants of the vowel transitions and the durations of the step command for individual vowels were estimated. Variations of these parameters among the different types of vowels or articulatory components were examined.

Data Recording

The movement of pellets on the tongue in the production of connected vowels were observed by the X-ray microbeam method for three subjects. The location of the pellets for each speaker is illustrated in Figure 1. For subject 1, four pellets were attached to the surface of the tongue. For the

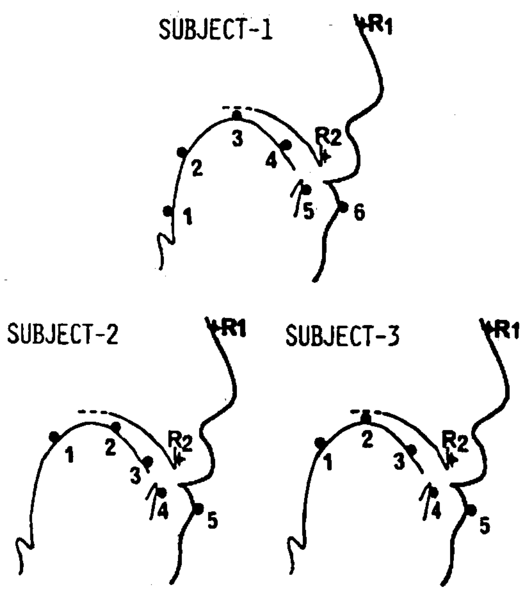


Fig. 1
Location of the pellets for the three subjects.

other two subjects, three pellets were attached to the tongue. A single pellet each was fixed on the lower incisors and on the lower lip. The frame rate of pellet tracking varied depending on the total number of pellets. The time interval between frames was 7.9 msec for subject 1 and 8.5 msec for subjects 2 and 3. The test materials were as follows:

(1) sequences of stationary vowels consisting of the five Japanese vowels /i/, /e/, /a/, /o/ and /u/.
(2) 20 meaningless words of the $V_1V_2V_1$ type, where V_1 and V_2 are one of the Japanese vowels and V_1 is not V_2.

No carrier phrase was used.

Parameter Description of the Tongue Configuration

In order to analyze the dynamic characteristics of tongue movement, a set of articulatory parameters which can effectively represent the resulting tongue configurations were derived. It was assumed that the movement of the pellets on the tongue and jaw could be approximated by a linear superposition of a small number of basic articulatory components. That is, the pellet positions at a given time were expressed as follows:

$$(x) = (x^M) + J\cdot(x^J) + L\cdot(x^L) + T_1\cdot(x^{T1}) + T_2\cdot(x^{T2}) + \cdots$$

where (x) is a vector listing the X- and Y-coordinates of all pellets. (x^M) is a vector listing the mean X- and Y-coordinate of each pellet over the entire time samples. (x^J), (x^L), (x^{T1}), etc. are the basic vectors listing the unit displacements of the pellets corresponding to each articulatory component. J, L, T_1, etc. are the articulatory parameters indicating the degree of each articulatory component at that time. The basic vectors were determined in the following way.

First, the movement of the jaw was approximated as a linear one dimensional movement. For this purpose, the movement of the jaw pellet during the entire period of the test utterances was plotted on the X-Y plane, and the first principal component axis was determined. For each time frame, the position of the jaw pellet was projected onto the axis. The value of the parameter J was defined by the displacement of the projected point from the mean position on this axis. Next, a linear regression analysis of the relationships between the X- and Y-coordinate of each tongue pellet and the value of the parameter J were performed and the regression coefficients were defined as the jaw-dependent component (x^J). Finally, the jaw-dependent displacements were subtracted from the coordinate of each tongue pellet for each time frame. The residual displacements were subjected to the principal component analysis, and (x^{T1}), (x^{T2}), etc. were derived as the eigen vectors.

The analysis was perfomed separately for the stationary vowels (i.e. test material (1)) and for the continuous vowels (i.e. test material (2)). Figure 2 shows the patterns of the pellet displacements for each articulatory component derived for the stationary vowels. In the figure, the positions of the tongue pellets are connected by the straight lines. Among the three curves, the middle line shows the mean positions of the pellets. The other two lines show the maximum range of displacement observed during the test utterances. The J-component consists of the jaw movement and the movement of the tongue pellet linearly related to the jaw movement.

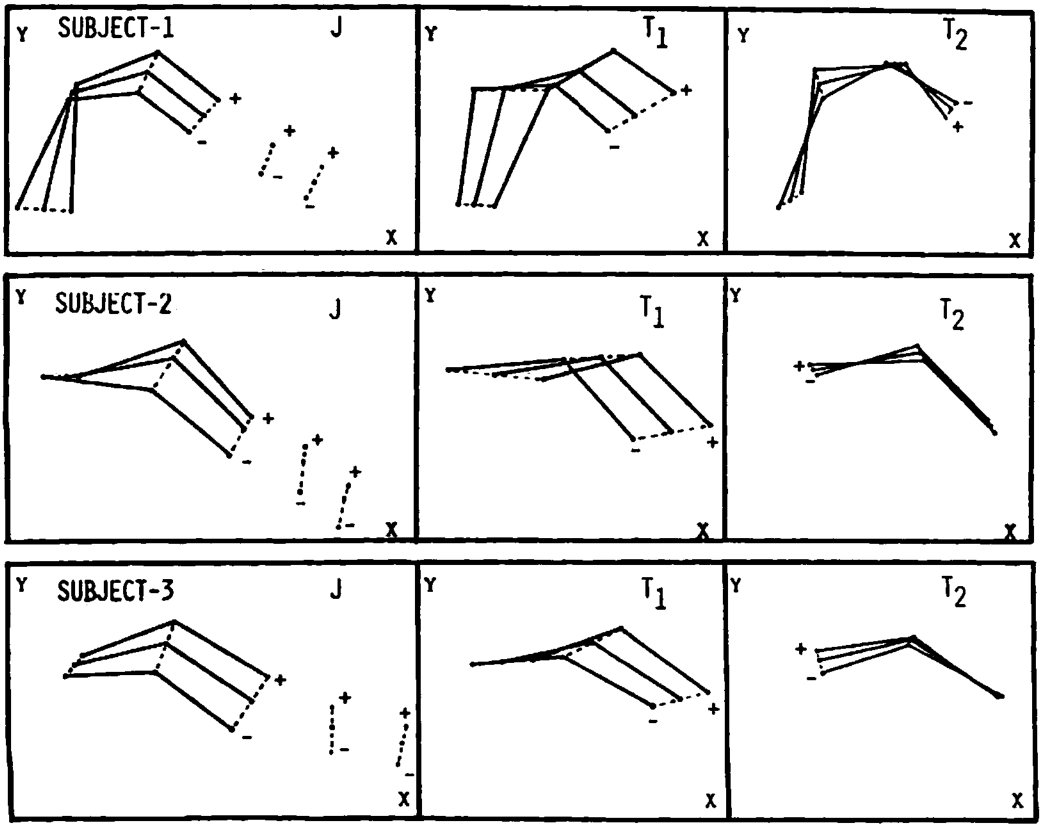


Fig. 2. Patterns of the pellet displacements for the articulatory components J, T_1 and T_2 derived for the stationary vowels.

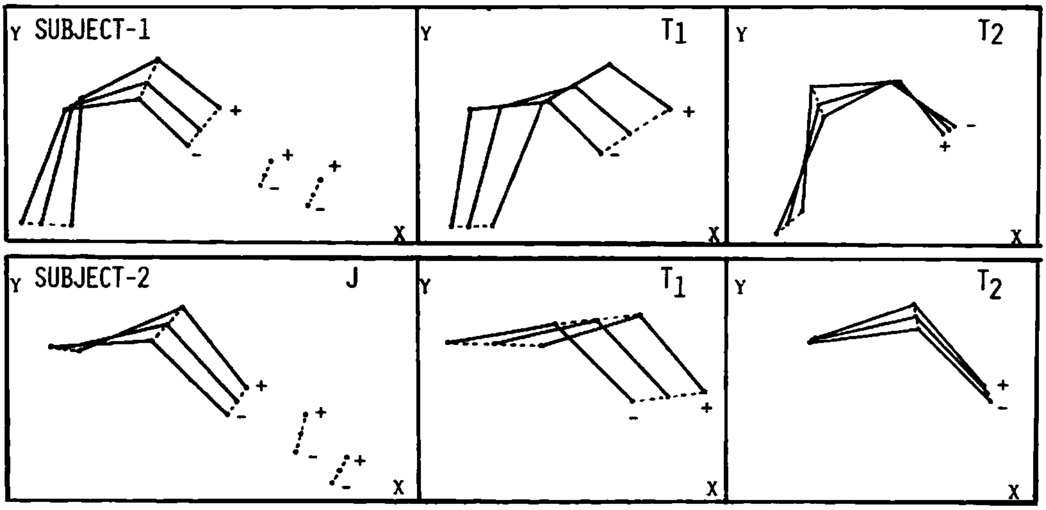


Fig. 3. Patterns of the pellet displacements for the articulatory components J, T_1 and T_2 derived for the continuous vowels.

This component appears to be related to the so-called high-low dimension of the tongue movement. The T_1 component is related to the front-back movement of the tongue. Figure 3 shows the pellet displacement for each articulatory component derived for the continuous vowels. It can be seen that the patterns of pellet displacements for individual components were essentially identical both for the stationary vowels and the continuous vowels. Figure 4 shows the distribution of the five Japanese vowels in the J-T_1 plane. The data for the three subjects were pooled in this figure. The values of the parameter J and T_1 for each subject were normalized so that the variance of the each parameter over the five vowels was the same among the three speakers. The position of each vowel was clearly separated in the plane except for the overlapping between the ranges of the vowels /a/ and /o/ (In the case of Japanese, these two vowels are mainly distinguished by lip gesture).

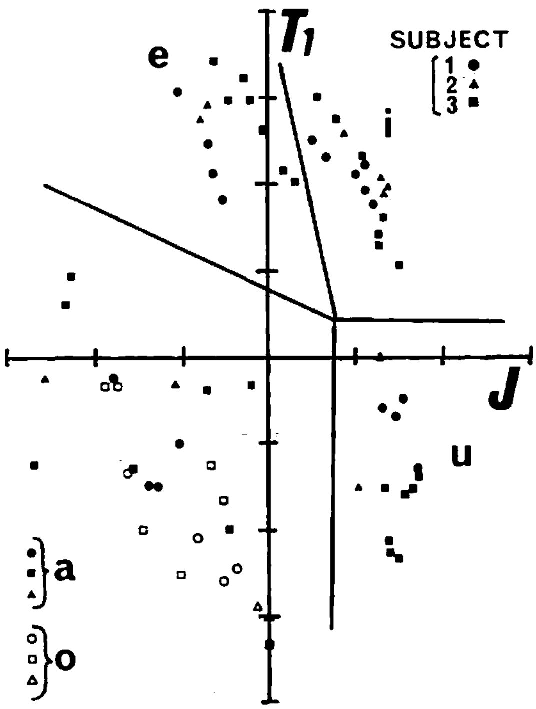


Fig. 4. Distribution of the five Japanese vowels in the J-T_1 plane.

Figure 5 shows the error in the approximation of the coordinates of the tongue pellets estimated by the linear regression estimation using 1 to 4 of the major articulatory components. On the horizontal axis is shown the set of articulatory components employed in the approximation. The standard deviation of the pellet coordinates over the entire test samples is also shown for reference. The vertical axis shows the error in the estimated coordinate values (standard deviation per pellet· frame). The solid lines indicate the results for subject 1 and the broken lines for subject 2. For both subjects, the data for the stationary vowels and continuous vowels are shown. The approximation error using only the J-component was about 6 mm. The error decreased to 1.7 mm by adding the T_1-component. Further addition of the T_2-component, however, merely decreased the error by 0.6 mm. It can be seen that, in the production of the Japanese vowels, the two components J and T_1 make dominant contributions in representing tongue gestures. The result was similar for both subject. The results for the stationary vowels and continuous vowels were also indentical.

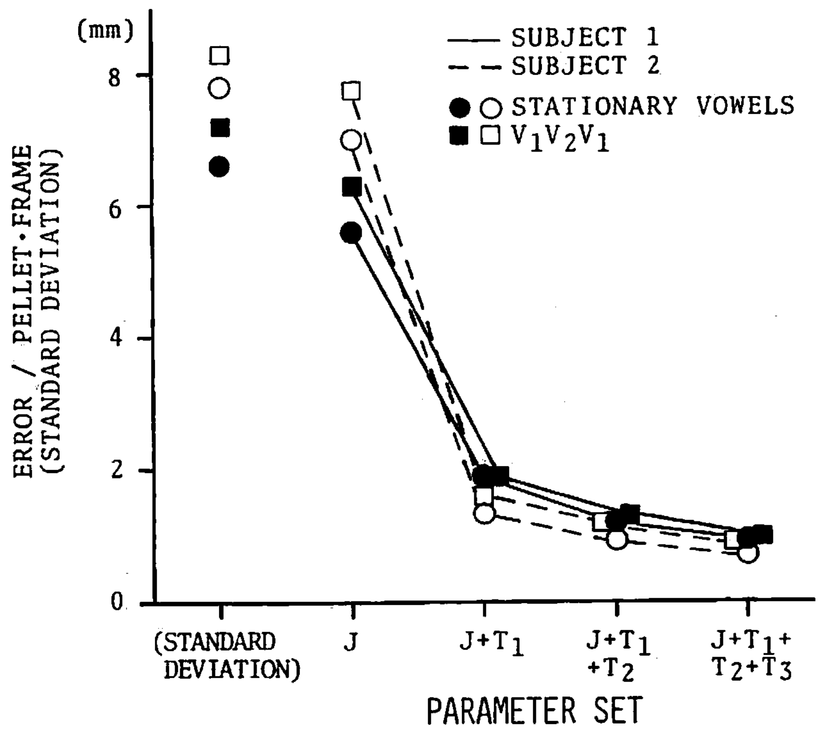


Fig. 5. Error in the approximation of coordinates of tongue pellets estimated by a linear regression estimation using major articulatory components. Leftmost data point shows the variance of the pellet coordinates over the entire test samples.

Dynamic Characteristics of Tongue Movements in $V_1V_2V_1$ words

The time functions of the J- and T_1-parameters for the $V_1V_2V_1$ test words were approximated by the step response function of a critically-damped second order linear system, and the optimum values of the time constants and the timing of the input step function were estimated.

Figure 6 shows an example of the step response approximation of the time function of the articulatory parameter J. The test word in this example is /iai/. The solid line shows the observed curve and the broken line shows the approximated curve. The rectangular function is the estimated input step function. The parameters depicted in the figure are as follows:

A_1, A_2, A_3: target values for the vowel V_1, V_2 and V_3;

t_1 , t_2 : timemoments of the onset of the step function for V_2 and V_3;

τ_1, τ_2: time constants of the V_1-V_2 and V_2-V_3 transition.

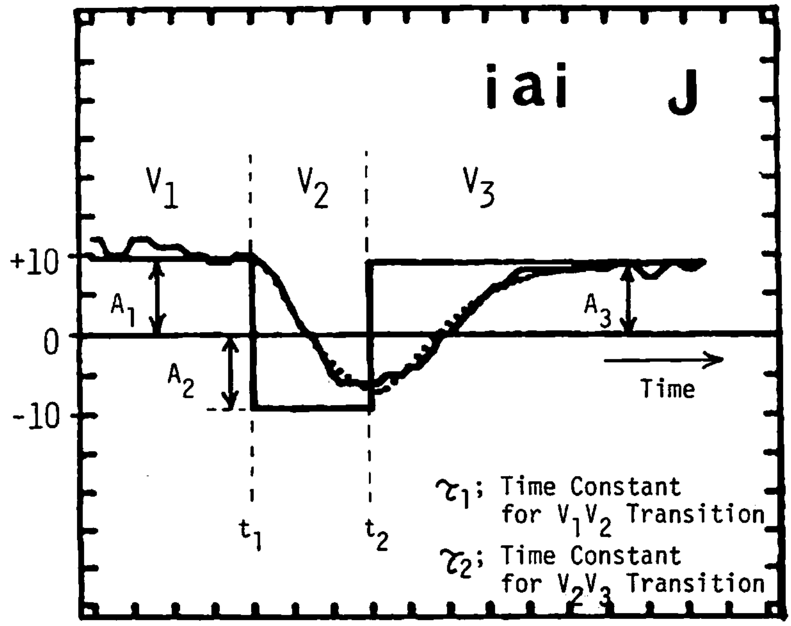


Fig. 6. Example of the step response approximation of the time function of the articulatory parameter J in the test word /iai/.

The time function of the articulatory parameter is approximated by the following equation.

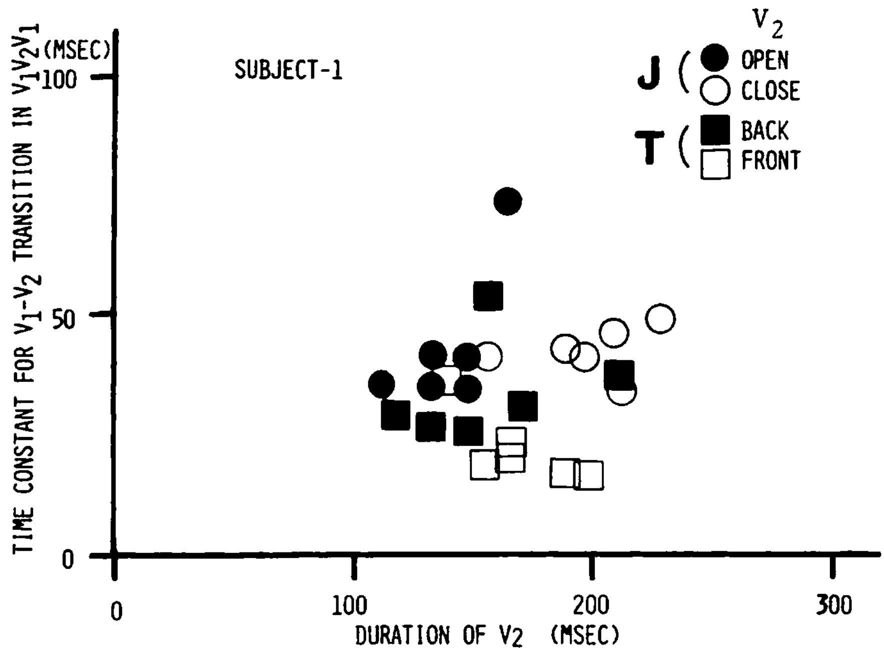
$$x(t)\begin{cases} = A_1 & (t<t_1) \\ = (A_2 - A_1)\left\{1 - (1 + \frac{t - t_1}{\tau_1}) \exp(- \frac{t - t_1}{\tau_1})\right\} + A_1 & (t_1 \leq t < t_2) \\ = (A_3 - A_2)\left\{1 - (1 + \frac{t - t_2}{\tau_2}) \exp(- \frac{t - t_2}{\tau_2})\right\} & \\ \quad - (A_2 - A_1) \cdot (1 + \frac{t - t_1}{\tau_1}) \exp(- \frac{t - t_1}{\tau_1}) + A_2 & (t \geq t_2) \end{cases}$$

In the above equation, during the period of $t<t_1$, the tongue is assumed to be in a stationary state. The value of A_1 and A_3 were determined by a visual inspection of the raw curve. For the J-parameter, only those test words in which either V_1 is high and V_2 is low or V_1 is low and V_2 is high were analyzed. Similary, for the T_1-parameter, those test words in which V_1 and V_2 shows a front-back contrast were analyzed (Here, the vowel /u/ was treated as belonging to the back vowel group). Other test words from test material (2) were excluded from the present analysis because the variation for the parameters J and T_1 was too small to obtain a reasonable estimation of the time constant. Thus, for each parameter, 12 test words were examined. In the following, the parameter T_1 will be referred to as T for the sake of simplicity.

Results

Figure 7 shows the estimated values of the time constant for the V_1-V_2 transition and the duration of V_2 for each test word. The horizontal axis shows the duration of V_2. Here, the duration of V_2 is defined as the time interval between t_1 and t_2. The vertical axis shows the value of the time constant. The data for the J-parameters are shown by circles and the data for the T-parameters are shown by squares.

It was found that the time constant for the V_1-V_2 transition for the J-parameter was about 1. 5 times longer than that for the T-parameter. The mean value of the time constant of the J-parameter was 43 msec for subject 1 and 48 msec for subject 2. The time constant of the T-parameter was 28 msec and 29 msec for each subject. Test words for the J-parameter were divided into two groups. The first group, indicated by the open circles, consisted of words in which V_2 was a closed vowel and V_1 was an open vowel. For example /aia/ belongs to the first group and /iai/ belongs the second group. For the T-parameter, the test words were also divided into two groups. In the first group, indicated by the solid squares, V_2 is a back vowel, and in the second group, indicated by the open squares, V_2 is a front vowel. In the case of subject 2, it can be seen that the duration of V_2 was longer when V_2 was a front or closed vowel. Similar effects were also observed for subject 1, although in the case of the T-parameter, the difference between the two groups of test words was smaller. The difference in the duration of V_2 between the two groups of test words can be observed more consistently in Figure 8. In this figure, the durations for V_2 were compared for each pair of test words, such as /aia/ and /iai/. The solid line represents the duration of V_2 in the test words to the left of the slash and the broken line the test



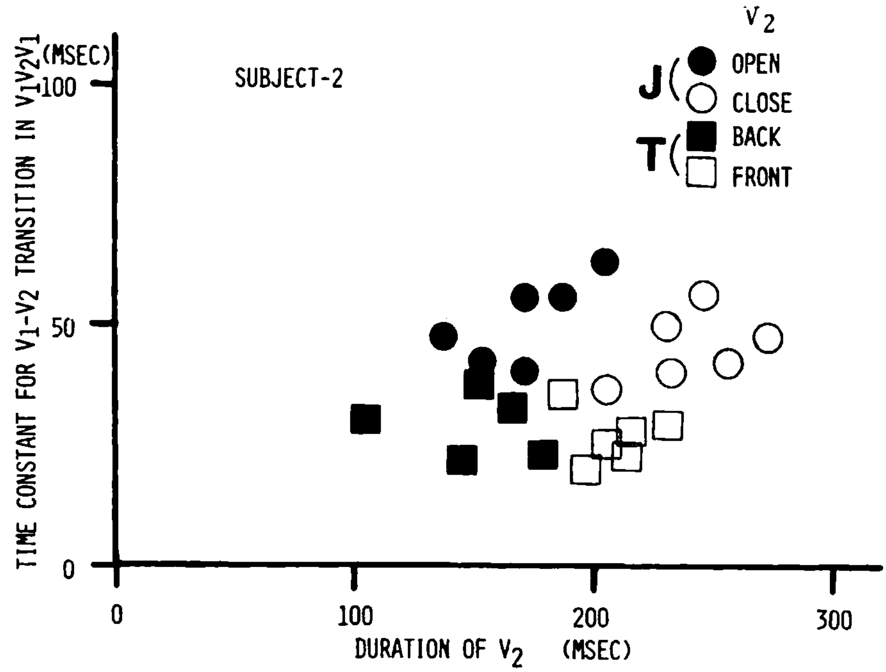


Fig. 7. Estimated values of the time constant for the V_1-V_2 transition vs. the duration of V_2.

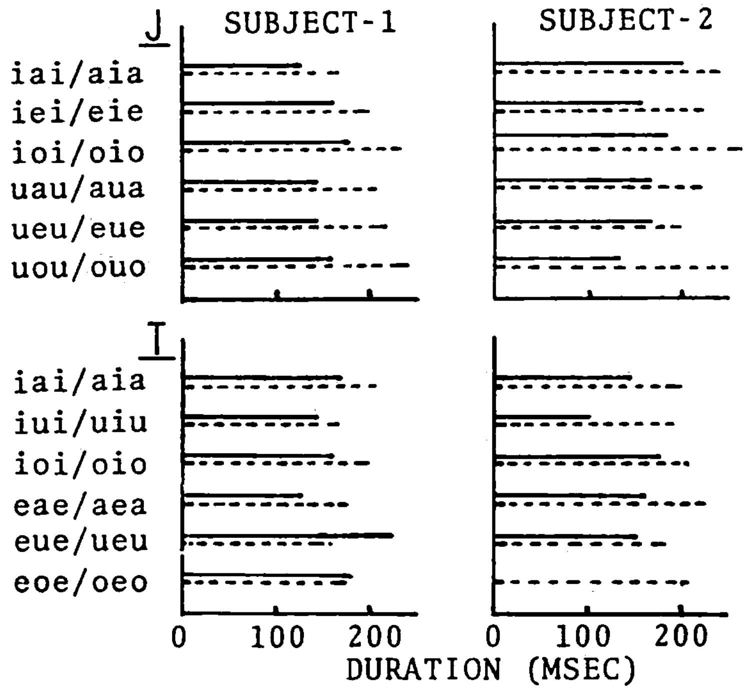


Fig. 8. Comparison of the durations of V_2 in the different pairs of test words.

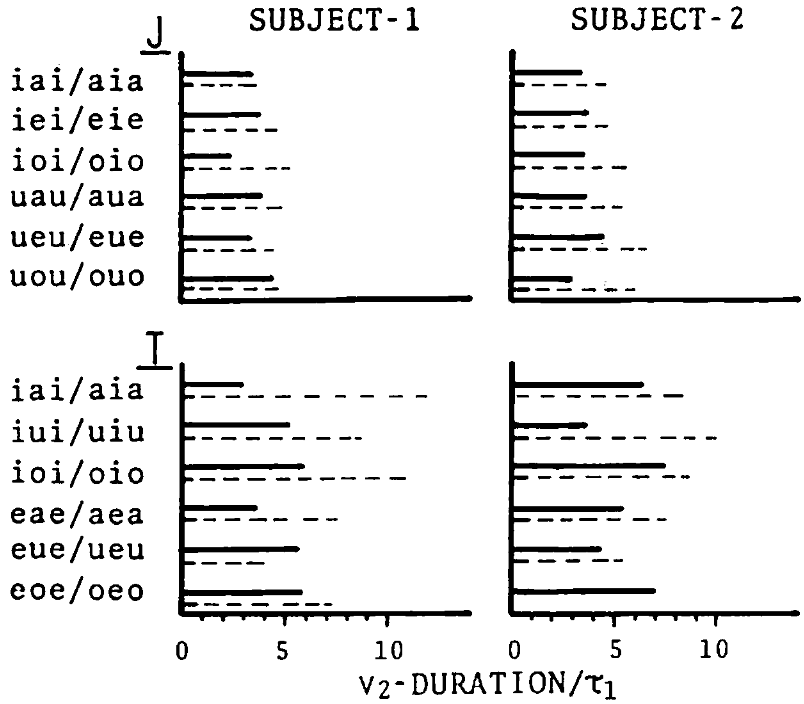


Fig. 9. Differences in the degree of excursion of the $V_1V_2V_1$ transition between the two word groups.

words to the right of the slash. Except in the case of the /ueu/-/eue/ pair of subject 1, the duration of V_2 was longer when V_2 was a front and/or back vowel.

The diffrence in the two groups of test words was also observed if we compare the degree of excursion for the V_1-V_2 transition. Figure 9 compares the value of the V_2-duration $/\tau_1$ for each pair of test words in Figure 8. The use of the solid lines and the broken lines is the same as in Figure 8. It is clear that the 'excursion of transition' is greater for those test words in which when V_2 was a closed and/or front vowel.

In summary, in the production of $V_1V_2V_1$ sequences, it is generally the case that the tongue movement exhibits a quasi-stationary state for V_2 when V_2 is a closed and/or front vowel. On the other hand, when V_2 is a back and/or open vowel, a quasi-stationary state for V_2 was hardly observed. This tendency is more marked for jaw movement than for the front-back movement of the tongue. Thus, for the vowel /a/, jaw movement is generally characterized by an undershoot phenomena.

References

1) Kiritani, S., K. Itoh, and O. Fujimura (1975); Tongue-Pellet Tracking by a Computer-Controlled X-ray Microbeam System, J. Acoust. Soc. Am. 57: 1516-1520.

2) Kiritani, S., S. Sekimoto and H. Imagawa (1977); Parameter Description of the Tongue Movements for Vowels. Ann. Bull. RILP, 11, 31-37.

3) Kiritani, S., S. Sekimoto, H. Imagawa, and H. Fujisaki (1977); Parameter Description of Tongue Point Movements in Vowel Production. Contributed Papers, 9th International Congress on Acoustics, Madrid, I, 419.

Acknowledgement

The work was in part supported by Grant in Aid for Scientific Research, Ministry of Education (No. 239005, No. 349008).