Ann. Bull. RILP, No. 11, 31-37 (1977)

PARAMETER DESCRIPTION OF THE TONGUE MOVEMENTS FOR VOWELS

S. Kiritani, S. Sekimoto, and H. Imagawa

In the present study, the movements of the tongue in the production of the Japanese vowels observed by the x-ray microbeam system [1] [2] were analyzed by a statistical method and a set of articulatory components for the specification of vowel articulations were derived.

As is shown in Figure 1 (a), several metal pellets were attached to the surface of the tongue, lower lip and lower incisor and their movements were tracked by the x-ray microbeam. Four pellets were attached to the tongue. The fourth pellet was placed about one cm back from the tongue tip. The other pellets were attached with a distance of about 3 cm between them. Figure 1(b) shows the xy-trajectories of the pellets during the pronunciation of a sequence of five Japanese vowels /ieaou/.

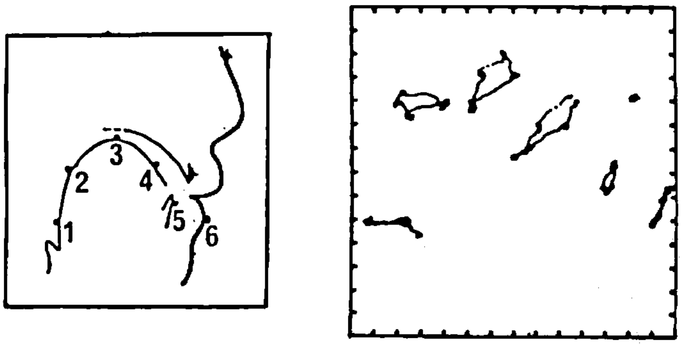

Fig. 1. Locations of pellets on the tongue, the lower incisor and the lower lip (a), and pellet trajectories for vowel sequence /ieaou/ (b).

The third pellet (No. 3) was placed at about the highest point of the tongue for vowel /i/. Starting from the high position for vowel /i/, it moved down for vowel /e/ and went backwards for vowels /a/ and /o/. The backmost pellet moved essentially in a horizontal direction, mainly reflecting front-back movement of the tonuge. The coordinates of the pellets at each time-frame were stored in a computer and subjected to a later analysis.

The method of analysis for deriving the articulatory componetns was as follows:

It was assumed that the movements of the pellets could be approximated by a linear superposition of a small number of basic articulatory components. This assumption is expressed as follows:

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

In the above equation, (x) is a twelve dimensional vector listing the x- and y-coordinates of 6 pellets. (X^J), (X^L) and (X^{T1}) are the basic vectors representing the pellet displacements for each articulatory component. The coefficients J, L and T_1 show the degree of each articulatory component in a given tongue configuration. (x^J) is a component related to jaw movement. (x^L) is lip movement. (x^{T1}) and (x^{T2}) are the components related to change in tongue shape. Pellet displacements for each articulatory component were determined in the following way.

First, the movement of the jaw was approximated as linear. For this purpose, the movements of the jaw pellet during the entire period of the speech sample were plotted on the xy-plane and the linear regression line was determined. At each time frame, the position of the jaw pellet was projected onto the regression line. Displacement of the projected point on the line defined the value of parameter J. Then, by a linear regression analysis, the component of the movements of the other pellets, which was linearly dependent on the jaw position was determined. These displacements were represented as the jaw-dependent component (x^J). Next, the jaw-dependent displacements were subtracted from the coordinates of the tongue pellets. The residual movements were subjected to the principal component analysis and the articulatory components (x^{T1}) and (x^{T2}) were derived as the principal eigen vectors. The movement of the lip pellet independent of the parameter J was also approximated as linear.

First, the method was applied to the analysis of the continuous sequence of stationary vowels. The utterance consisted of vowel sequences /ieaou/, /iuoae/, /iaueo/ and /ioeua/, containing about 1000 time samples of pellet-coordinates. Figure 2 shows the displacement of the pellets for the three articulatory components.

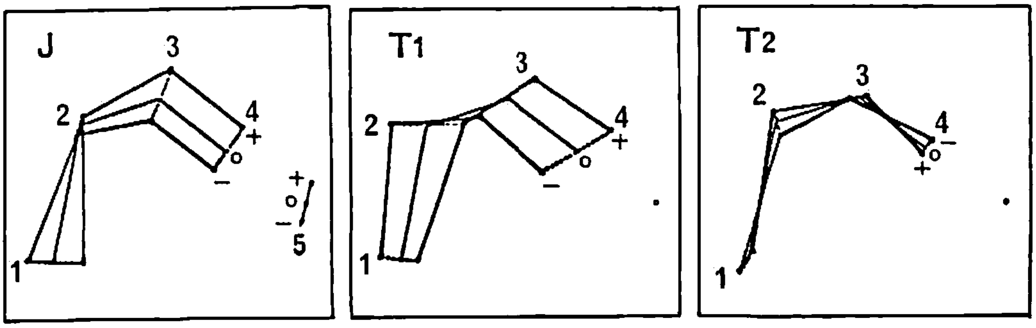

Fig. 2. Pellet displacements for each articulatory component.

The J-component at the top consists of the jaw movement and the movements of other pellets associated with the jaw movement. In the figure, the pellet positions are connected by the straight lines. In the three curves, the middle curve shows the mean positions of the pellets over the entire utterance. The other two curves show approximately the maximum range of displacement observed in the sample utterances. This component is clearly related to the so-called high-low dimension of the tongue movement. It is to be noted that the back-most pellet (pellet No. 1) moves forward with the closing of the jaw. The T_1 component in the middle corresponds to the front-back movement of the tongue.

The bottom figure shows the T_2 component. This component can be interpreted as a kind of a bulging of the tongue dorsum. It is seen that the displacement of the T_2 component is considerably smaller than that of T_1.

The accuracy in describing pellet positions by using these main components is summarized in Table 1. As for the four pellets on the tongue, when the movements of the pellets were approximated as the linear sum of the three components (J-component, T_1-component and T_2-component), the approximation error was 1.3 mm in standard deviation, on the average.

PARAMETER SET	ERROR/PELLET FRAME		
	TONGUE	JAW	LIP
J, L	6.0	0.8	0.9
J, L, T_1	2.1	"	"
J, L, T_1, T_2	1.4	"	"

Table 1. Accuracy in the approximation of the pellet coordinates for the stationary vowel sequences (Standard deviation in mm).

Figure 3 shows the trajectories of the pellets for the vowel sequence /ieaou/, approximated by using 3 and 4 main components. It appears that the essential characteristics of the vowels are represented by the J and T_1 components. The T_2 component is necessary for describing the movement of pellet No. 2 on the tongue dorsum.

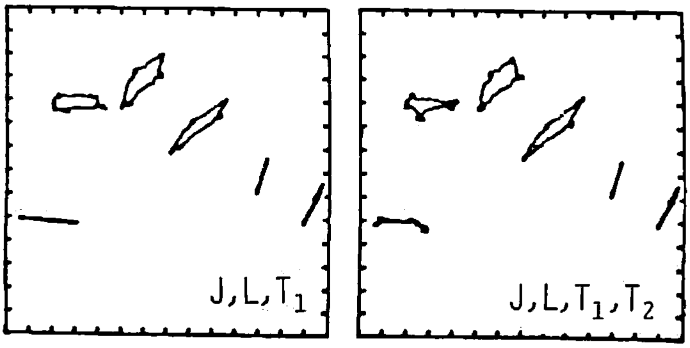

Fig. 3. Pellet trajectories for /ieaou/ approximated by using main articulatory components.

The analysis shown above was mainly concerned with the stationary portion of the vowels. The sample utterances contained the transitions between the stationary vowels, but, in the entire period of the utterances, the transitional period was relatively small. In order to further examine the effectiveless of these parameters, a similar analysis was made of the continuous vowels. One group of the speech samples was a set of nonsense words consisting of $V_1V_2V_1$ sequences. V_1 and V_2 were one of the five Japanese vowels. The number of the test words was 20.

Another set of speech samples was a collection of Japanese words, which consists of the sequence of vowels and semivowels. In the production of these test words, the tongue moves continuously throughout the pronunciation of the word. It was often difficult to define a stationary portion for a given vowel.

Figure 4 compares the accuracy in the parameter representation of the pellet movements for the stationary vowel sequences and the $V_1V_2V_1$ test words. On the horizontal axis is shown the set of articulatory components employed in the approximation. The vertical axis is the error in the approximation of the pellet coordinates. It can be seen that the results for the two sets of data are nearly the same. The description of tongue movements by these articulatory components is useful not only for stationary vowels, but also for the transitional movements of the tongue. It was also confirmed that the articulatory components derived from each set of speech samples were quite identical.

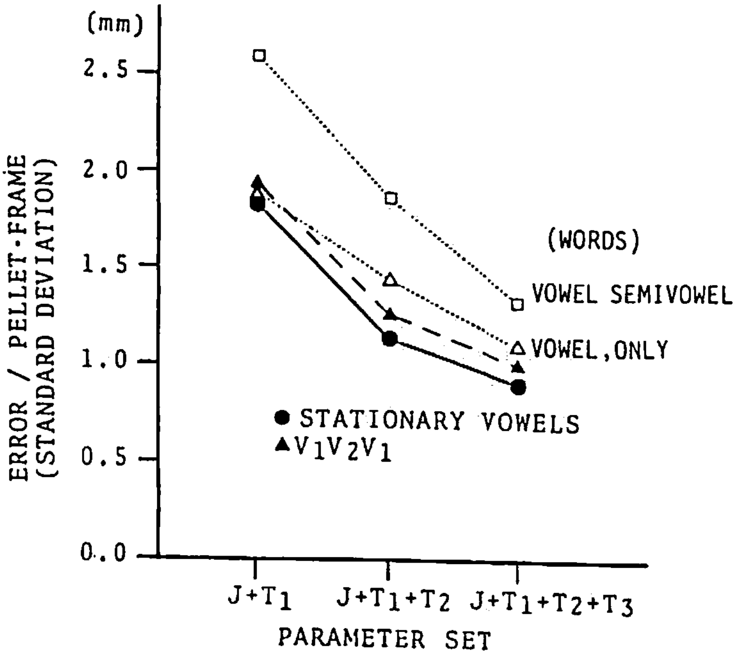

Fig. 4. Accuracy in the approximation of pellet coordinates.

As for the real words, the observed tongue movements were approximated using articulatory components, which were derived from $V_1V_2V_1$ test words. The words were divided into two types of groups. In one group, the test words consisted of vowels only. In the other group, the test words contained semivowels. It can be seen that for the latter group, the error was considerably greater than for the vowel-only words. This result indicates that there is a difference in the articulatory components between vowels and semivowels in Japanese.

The relationship between the articulatory parameters and the formant frequencies was also investigated by a nonlinear regression analysis. Table 2 shows the cumulative contribution rates in the estimation of the formant frequenceis. Either a set of three articulatory parameters (J, L, T_1) or a set of four parameters (J, L, T_1 and T_2) were used in combination with the 1st and 3rd order equations. It was found that the 3rd order equation using four articulatory parametsrs results in a contribution rate as high as 99%. As for the 1st and 2nd formant frequencies, difference in the contribution rate was small whether the parameter T_2 was included in a parameter set or not. However, parameter T_2 appears to be important in the estimation of the 3rd formant frequencies.

	STATIONARY VOWEL				$V_1V_2V_1$ WORD	
	3rd order (J,L,T_1,T_2)	1st order $(J.L.T_1,T_2)$	3rd order (J,L,T_1)	1st order (J,L,T_1)	3rd order (J,L,T_1,T_2)	1st order (J,L,T_1,T_2)
F_1	98.9 %	89.5 %	98.1 %	87.7 %	96.7 %	85.4 %
F_2	99.0 %	96.3 %	98.6 %	94.6 %	98.6 %	96.1 %
F_3	91.5 %	43.6 %	84.2 %	18.6 %	78.6 %	17.0 %

Table 2. Cumulative contribution rate in the regression estimation of formant frequencies.

Figure 5 compares the time functions of the measured formant frequencies and the estimated formant frequencies. It can be seen that in the case of the 3rd order estimation, the estimation is nearly complete. The 1st order estimation of the 2nd formant frequency results in an estimation error of about 200 Hz in some vowels.

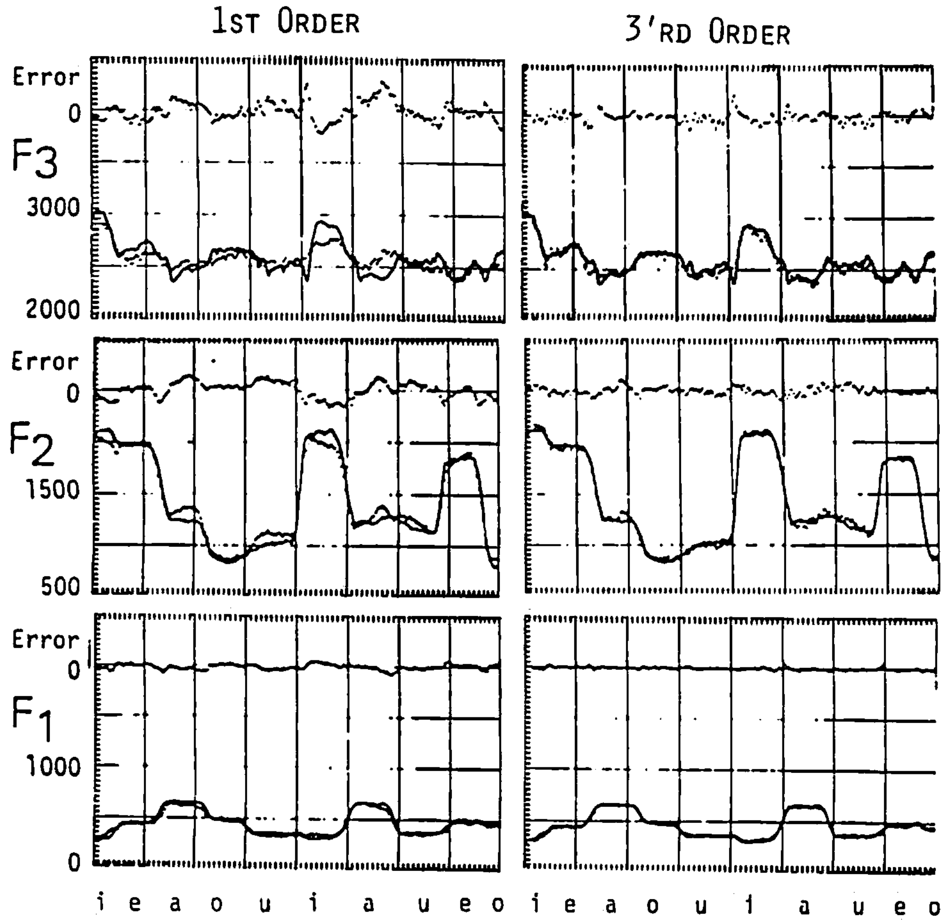

Figure 5. Time functions of the formant frequencies. Original (solid line) and approximated (dotted line).

The positions of the pellets corresponding to a given set of formant frequencies can be calculated using the regression equations derived above. Because we used three formant frequencies and four articulatory parameters a range of pellet positions were obtained for each formant pattern. Figure 6 shows the pellet positions corresponding to the typical formant pattern of each Japanese vowel. The articulatory parameters were assumed to vary independently of each other. The possible range of each parameter was determined as being between the maximum and minimum values observed in the sample utterances. The range of pellet positions for each vowel forms a connected region, and the pattern of the pellet positions is clearly distinct for each vowel. For vowel /u/, the range is somewhat greater than that for the other vowels.

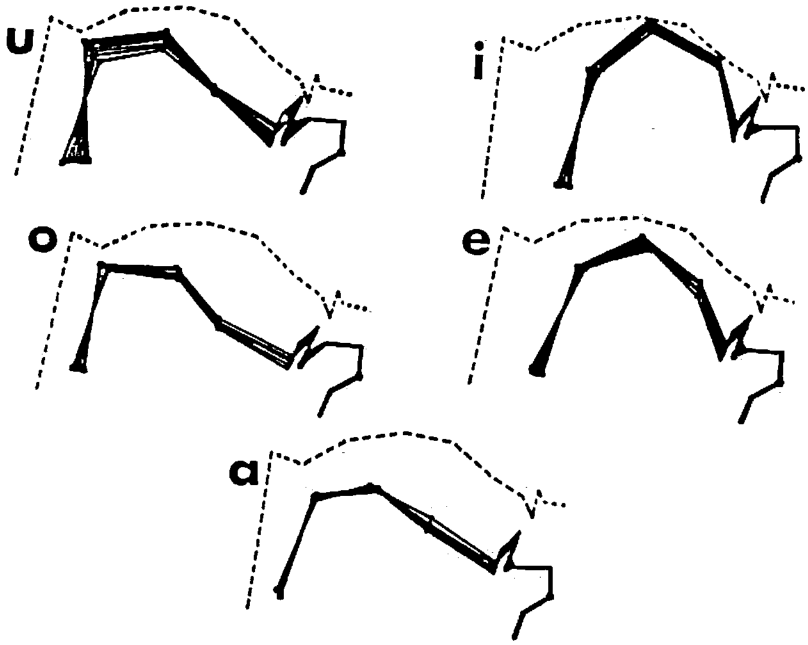

Fig. 6. Pellet positions corresponding to formant patterns of five Japanese vowels.

The tongue movements in the production of vowels in American English were analyzed in a similar way. Test words were monosyllabic with word initial /p/ and word final /p /. Ten different vowels were used as the word medial vowel. The tongue movement in the middle half, between word initial /p/ and word final /p/ was analyzed. The pattern of pellet movement for each articulatory component were essentially the same as that obtained for the Japanese vowels. The J-component corresponds to the high-low dimension, and the T_1-component to the front-back dimension. However, the displacement associated with the T_2 component is much greater for American English than for Japanese.

Figure 7 compares the accuracy in the parameter representation of the pellet positions for the Japanese and English vowels. When three parameters (J, T_1 and T_2) are used, the accuracy is nearly the same for Japanese vowels and English vowels. But, the contribution of the T_2-component is much greater for American English. This component was necessary for describing the tongue shapes for the tense back vowels [u] and [o] (Figure 8).

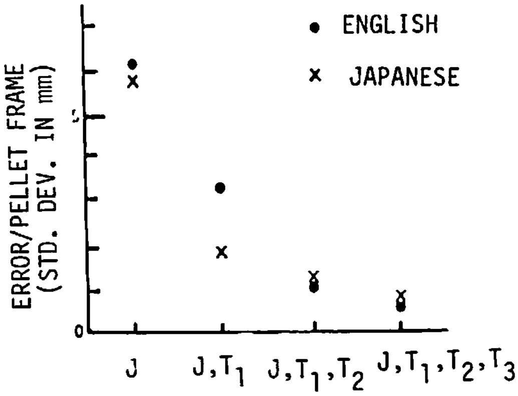

Fig. 7. Comparison of approximation error in English and Japanese vowels.

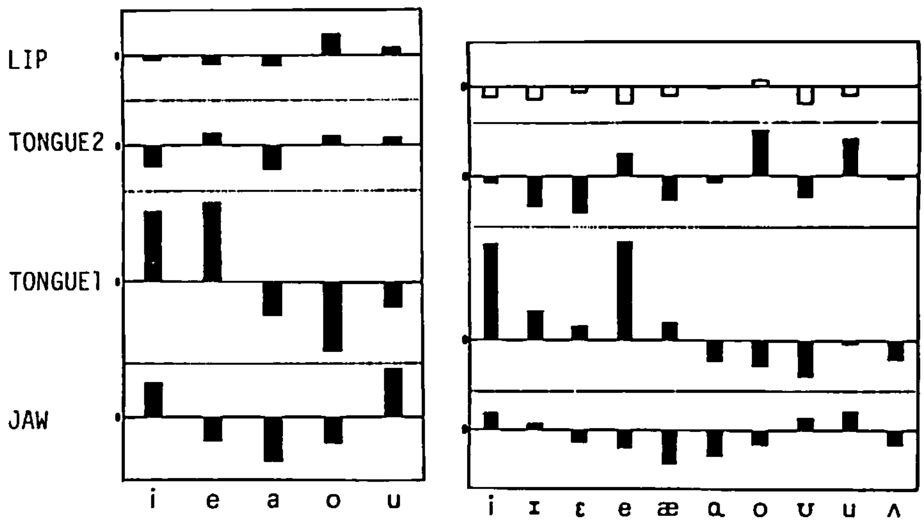

Fig. 8. Contribution of each component in English and Japanese vowels.

This paper was presented at the 9th ICA meeting in Madrid, Spain, 1977. This work was supported in part by Grant in Aid for Scientific Research (No. 185079), Ministry of Education.

References

1. Fujimura, O., S. Kiritani, and H. Ishida (1973); Computer-Controlled Radiography for Observation of Movements of Articulatory and Other Human Organs, Comput. Med. Biol., 3: 371-384.
2. 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.