Ann.Bull.RILP(1985)No.19, 7-19

CONTEXTUAL VARIATION IN THE JAW POSITION FOR THE VOWELS IN /CVC/ UTTERANCES

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

We have been conducting obsevations of the movements of the jaw during speech and an analysis of its dynamic characteristics. In our previous study, we reported characteristics of the contextual variation in the position of the jaw for the consonant in VCV sequences[1]. It was shown that, using a linear second order system model, the pattern of the contextual variation could be summarized in a simple form as 1) the target positions for the consonants /s/ and /t/ are essentially constant regardless of the vowel context; and 2) the remaining contextual variation in the consonants /p/ and /k/ are mainly due to assimilatory phenomena to the vowel context. In the present study, the characteristics of the contextual variation in the position of the jaw for the vowels in CVC sequences are investigated using a similar method of analysis.

Data Recording

Test words used in the present study were as follows.

$/aC_1VC_2a/$ $\quad C_1, C_2$ = p, t, k, s.
$/iC_1VC_2i/$ $\quad V$ = u, i, e, a.

For each of the four vowels, 16 consonant contexts were provided in the two types of test word. The test words were uttered in a carrier phrase /---desu/, and for each test word, 5 utterance tokens were recorded.

The movements of the jaw were observed by a special device using a PSD (position sensitive detector). As shown in Fig.1, two infrared light emitting diodes were attached to a solid steel wire which was attached to the lower front teeth, coming out of the mouth. Two additional LEDs were attached to another solid wire which was fixed to the frame of the glasses of the subject to monitor the possible movement of the head during speech. The images of the LEDs were formed on the PSD within a camera body. Coordinate signals of the diodes from the PSD were fed into a computer.

In order to faciliate the recording of a large number of utterances, input signals were continuously sampled into a cyclic buffer in the computer memory. The subject read a list of test words and at the end of each test word, he pushed a control button. Then, the computer transferred the data in the cyclic

* School of Engineering, Waseda University

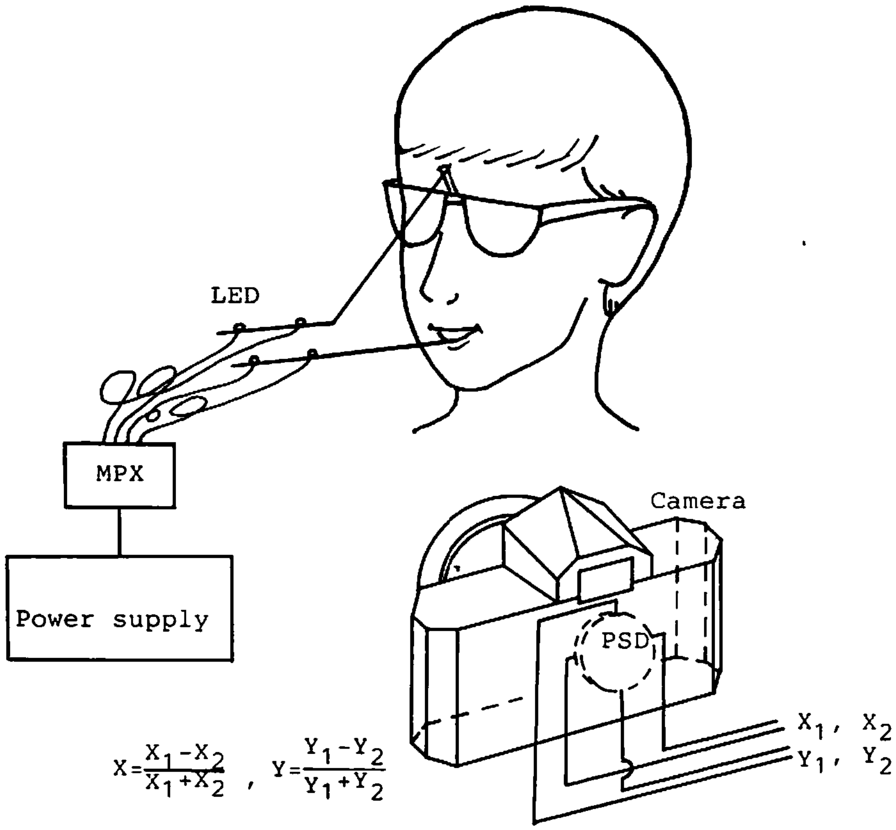

Fig. 1 A method for recording the movement of the jaw using a PSD (Position sensitive detector).

buffer onto a disk file. The cyclic buffer contained data of two seconds in length.

Context Variation in the Jaw Position

Fig.2 shows the variation in the positions of the jaw for the midde vowels in /VCVCV/. The jaw positions for stationary vowels are also shown for reference.

The degree of jaw opening for stationary vowels increased in order for /u/, /i/, /e/ and /a/. Differences between pairs of vowels were always significant. The significance level was 1% for all the pairs except the /e/-/a/ of subject 1, for which the significance level was 5%.

For the vowels in the /VCVCV/ words, the jaw openings were

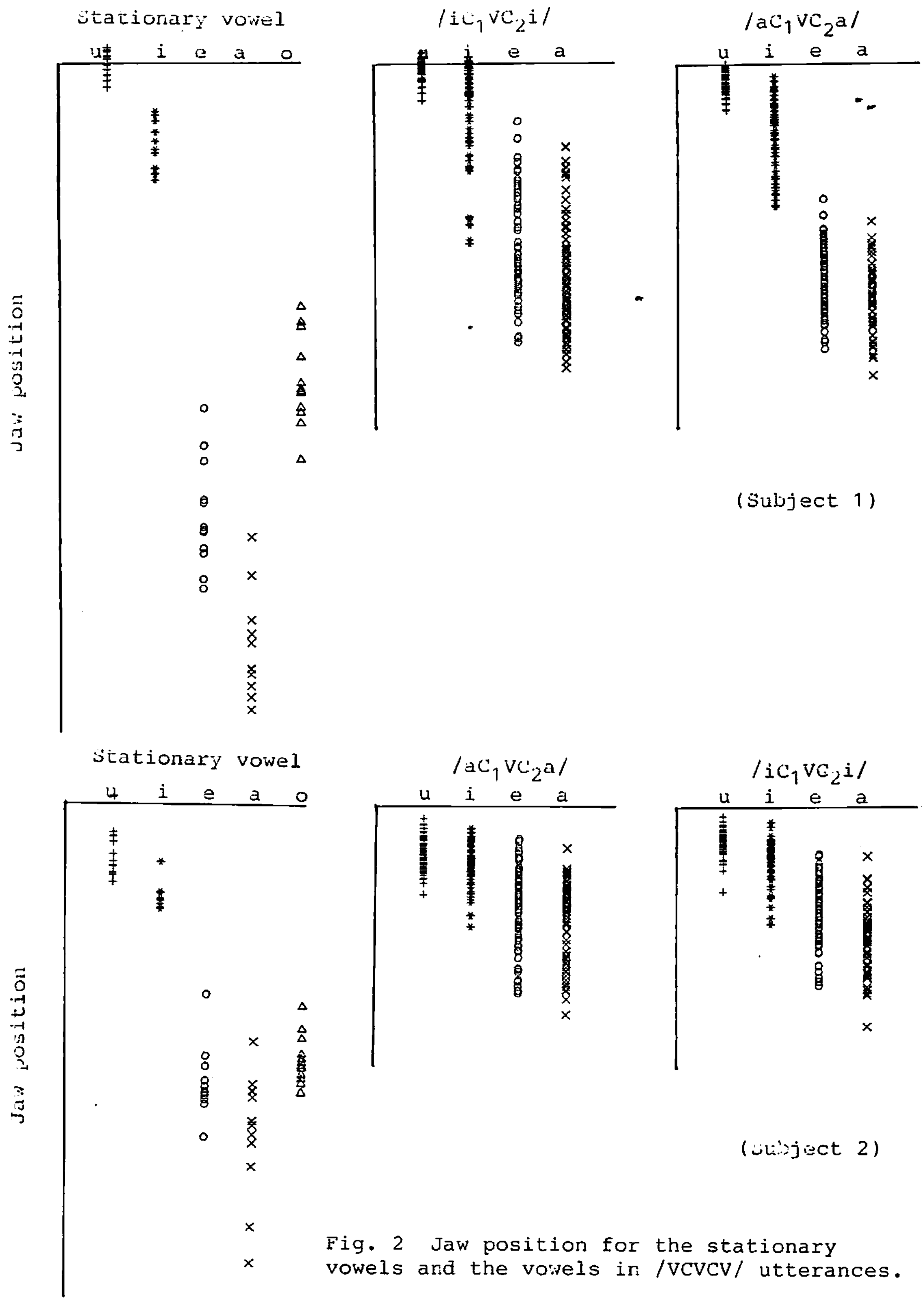

Fig. 2 Jaw position for the stationary vowels and the vowels in /VCVCV/ utterances.

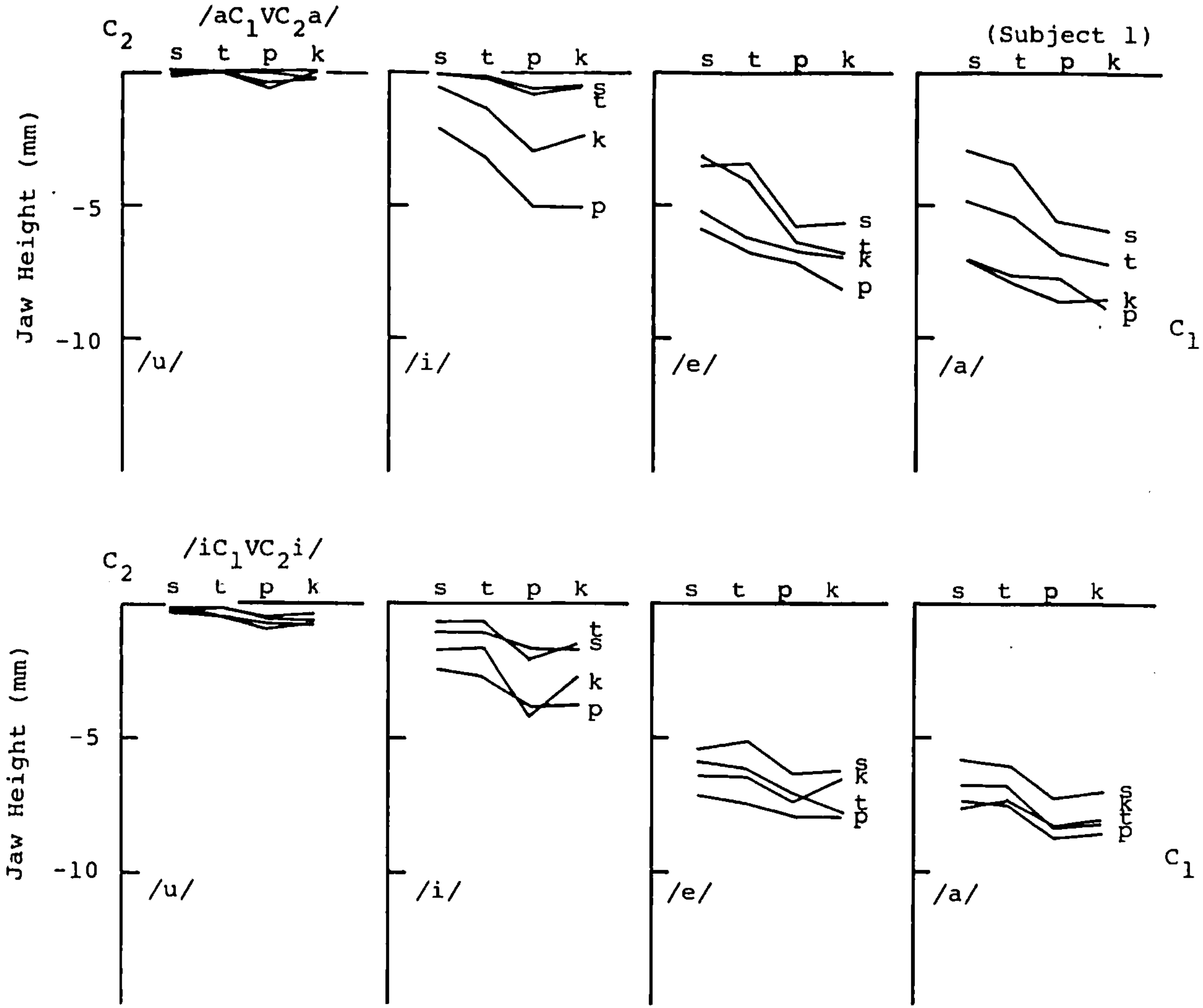

Fig. 3 Jaw position in the individual consonant context in /VCVCV/ utterances.

generally smaller than for the corresponding stationary vowels. The difference was particulary large for the open vowels /e/ and /a/. For these vowels, the jaw openings in the /VCVCV/ words were about 40% of those for the stationary vowels.

It also can be seen in the figure that there are considerable overlaps between the ranges of the jaw positions of the vowels. However, statistical tests revealed that there still existed significant differences among the vowels. Furthermore, when the jaw positions for the four vowels in the individual consonant contexts were compared, it was found that, in almost all the consonant contexts, the degree of thejaw opening was in the order of /u/, /i/, /e/ and /a/, which was the same as that for staionary vowels. This order was observed in 30 context conditions for subject 1 and in 28 context conditions for subject 2, out of 32 (16 consonant x 2 vowel) conditions.

Fig.3 shows the jaw positions of each vowel in individual consonant contexts. Contextual variation in the jaw position was greatest for the vowel /a/ and next greatest for the vowel /e/. When the utterances with the same first consonant C_1 were compared, the jaw opening for the vowel was smaller when the second consonant was /s/ or /t/ than /p/ and /k/. Similary, when the utterances with the same second consonant were compared, jaw opening for the vowel was are smaller when the first consonant was /s/ or /t/ than /p/ and /k/. A similar tendency was also observed for the vowel /i/, although the contextual variation for this vowel was smaller than for /e/ and /a/. As for the vowel /u/, the contextual variation was the smallest among the four vowels.

As for the effect of the consonants /s/ and /t/, the effect of /s/ was larger than that of /t/ for the vowels /e/ and /a/, but for the vowels /i/ and /u/, a clear difference was not observed (in this case /s/ and /ts/).

Comparison of the utterances /iCVCi/ and /aCVCa/ showed that, for the vowels /i/ and /u/, the magnitude of the contextual variation was nearly the same. However, for the vowels /a/ and /e/, the contextual variation was clearly larger in the utterance /aCVCa/ than in the utterance /iCVCi/. This is mainly due to the fact that, when C_1 was /s/ or /t/, the jaw position in /aCVCa/ became higher than that in /iCVCi/. In the utterance /asVCa/ or /atVCa/, the displacement of the jaw from the first vowel to C1 was particulary large, and the velocity of the movement was also great. It may be concluded that this caused a great undershoot in the movement of the jaw for the middle vowel.

Analysis based on the Linear Second Order System Model

In order to analyze the characteristics of the contextual variations in the jaw position of the vowels, an analysis based on a linear second order system model was performed. The movements of the jaw were regarded as the response of the second order system to the hypothetical input step commands. The input step command was assumed to be the step function which represented a sequence of target positions for the successive phonemes.

Method of Analysis

For the present utterances, 6 steps were assumed as shown in Fig. 4, which corresponded to V_1, C_1, V_2, C_2, V_3 and /d/ respectively, and the amplitude of the steps and the timing of the step changes which gave the best approximation of the actual jaw movements were estimated. The approximation error was evaluated over the time interval from the begining of the utterance to the implosion of /d/. To determine the best values of the parameters, a method of dynamic programming was used.

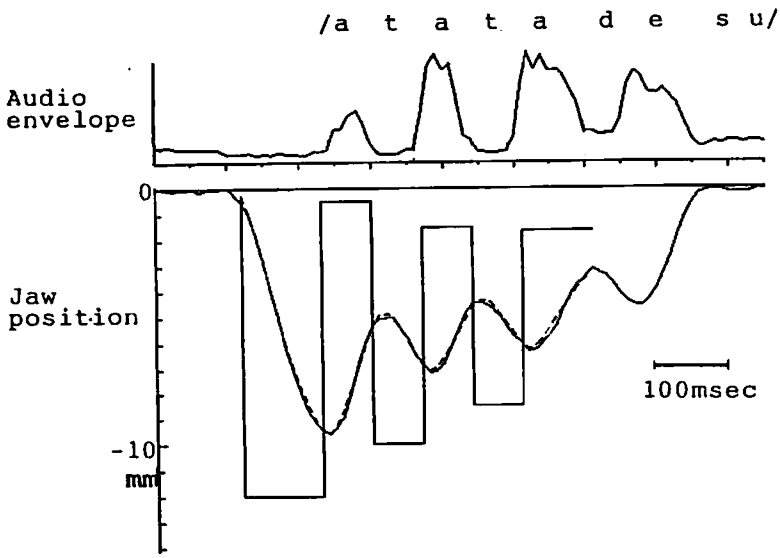

Fig. 4 Approximation of the jaw movement by the step function response of the linear second order system.

The range for the search of the timing of the step change corresponding to each phoneme-to-phoneme transition was specified with reference to the moment of the characteristic acoustic event on the speech envelope. The appropriate limit for this range was determined through preliminary examination as follows.

First, the moments of voice onset and voice offset were marked on the speech envelope as the begining and end of the pertinent phonemes. At the same time, the moment of the peak or bottom in the time curve of the jaw movement were determined, and their locations relative to voice onset or offset were measured.

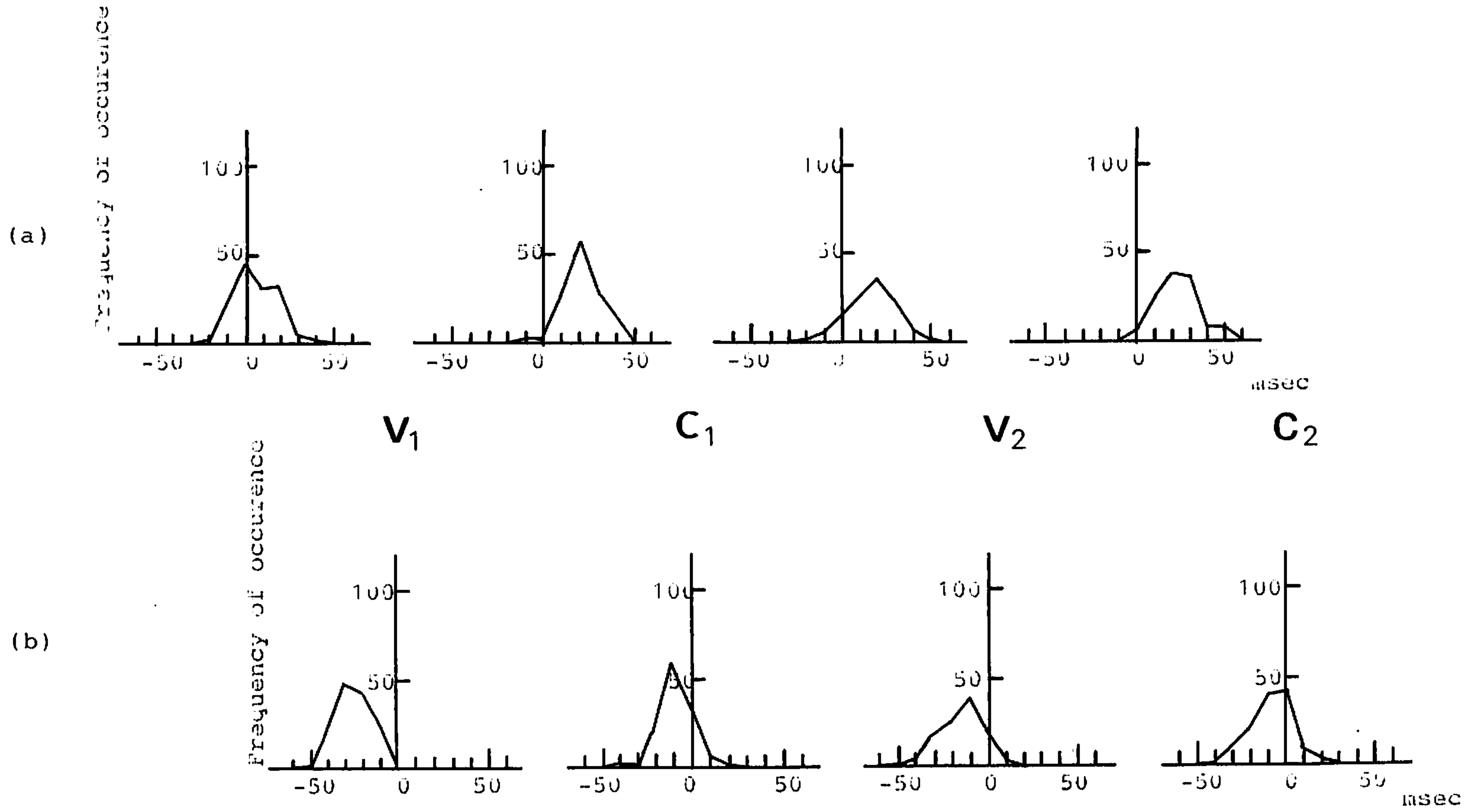

Fig. 5 Distribution of the moment of the peak or bottom in the jaw movement
(a) relative to the beginning of each segment (voice onset or offset)
(b) relative to the mid-point of each segment.

These turning points in the movement of the jaw were considered close to the begining of the movement towards the following phoneme.

Fig. 5 shows that these turning points were generally located after the begining (voice onset or voice offset) of the current phoneme. The range of the distribution of the turning points was from -20msec to +50msec relative to the begining of the current phoneme.

If the locations of the turning points were examined relative to the midpoint of each phoneme, the turning points were generally located before the midpoint and lie between -50msec and +10msec. It was noted that duration of the middle vowel was generally longer in the utterance /aCVCa/ than in the utterance /iCVCi/. Corresponding to this duration difference, the start of the jaw movement towards the following vowel relative to the voice onset of the vowel was later in /aCVCa/ than in /iCVCi/.

Based on these observations, the range for the search of the timing of the step change towards the following phoneme was determined as from 10msec before the begining to 10msec after the mid-point of the current phoneme.

Analysis with a Critically Damped Second Order System

In the first step of the analysis, approximations were performed under the assumption that the second order system was critically damped. Parameter values for the best approximation were eatimated for each value of the time constants 20, 30, 40, and 50msec.

Fig. 6 shows the magnitude of the approximation error averaged over the set of utterances as a function of the time constant value. When the time constant was too small, the approximation error was large. This was naturally due to the fact that the slope of the caluculated curve was too steep to fit the actual curve. When the value of the time constant was larger, the slower response could be compensated for by setting the step amplitude greater than the actual excursion of the jaw movement. Thus, the approximation error did not show a marked increase with the increase in the value of the time constant. Still, there was a slight increase in error for larger time constants, and the point of the minimum approximation error was at around 40msec for the time constant value.

For the open vowels /a/ and /e/, the contextual variation in the estimated target position of the jaw was smaller than the contextual variation in the observed jaw position. For an appropriate value of the time constant, the output from the second order system showed an undershoot effect. The contextual variation in the observed jaw position was (in part) explained by this effect and, consequently, the contextual variation in the estimated target position was smaller. When the value of the

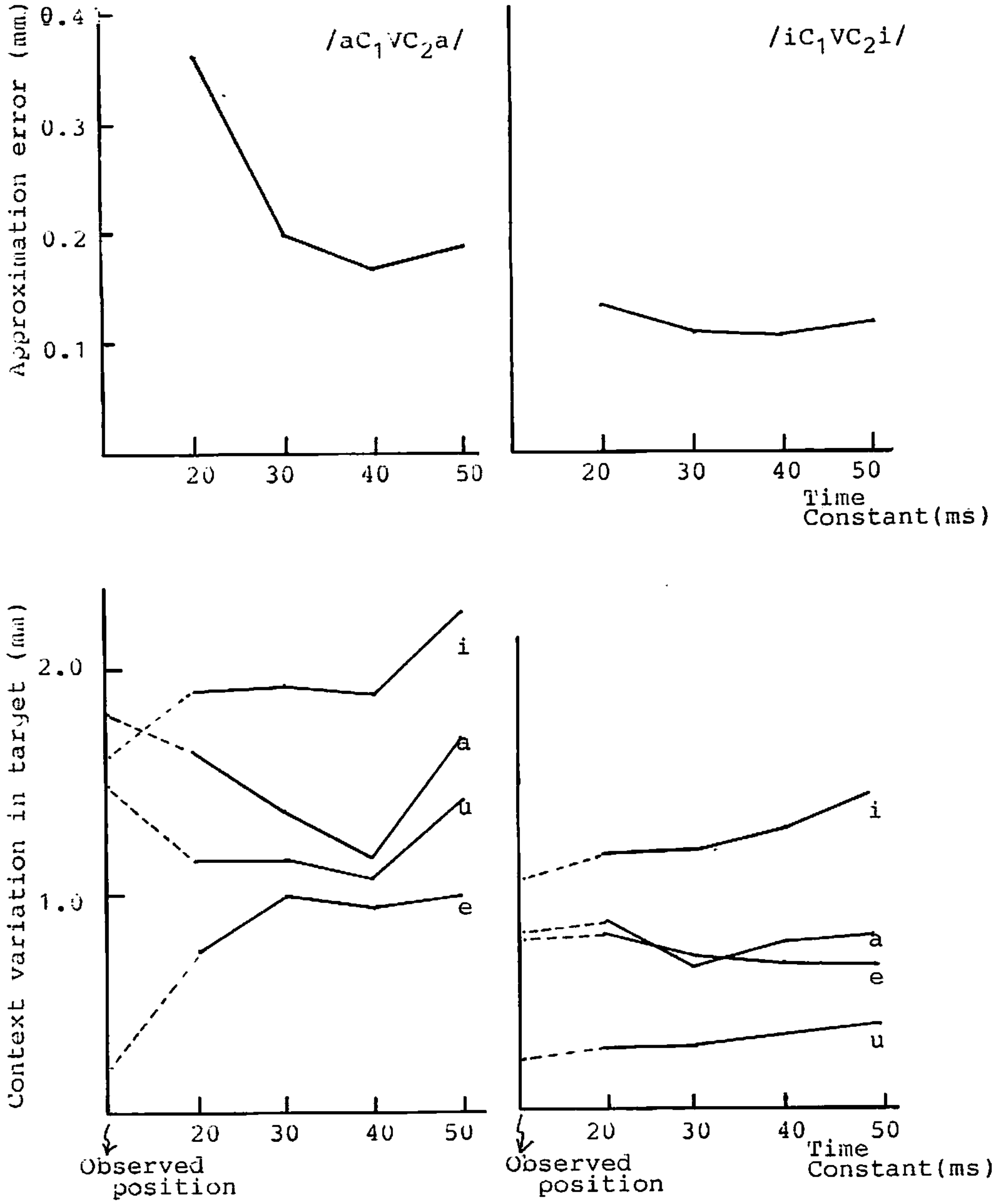

Fig. 6 Approximation error and the contextual variation in the target position as a function of the time constant of the critically damped second order system.

time constant was too large, the variation in the target position due to the consonant context became larger.

Thus, for the utterance /aCVCa/, a time constant of about 40msec gave the minimum value both for the average approximation error and for the magnitude of the contextual variation in the target position. For the utterance /iCVCi/, the variations in these values, as a function of the time constant value, were small.

Fig.7 shows the target positions of the vowel in 16 consonant contexts which were estimated for the time constant of 40msec. It can be seen in the figure that, for the open vowels /e/ and /a/, the contextual variation in the utterances /iCVCi/ and /aCVCa/ is clearly reduced compared to that in the actual jaw position (in Fig. 4). This result is apparently due to the undershoot effect in the second order model.

It was noted, however, that the average positions of the vowels /e/ and /a/ in the utterance /aCVCa/ were lower than that in the utterance /iCVCi/. The difference between the two types of utterance was larger than that in the actual jaw positions. It was also noted that, in the calculated target position, the contextual variation for /u/ was larger than in the actual jaw position. These two phenomena suggest that the approximation by the critically damped second order system has some limitations, and that to obtain a better approximation, it is necessary to introduce some kind of modification in the model.

Analysis with an Underdamped Second Order System

In order to get more insight into the characteristics of the contextual variation in the jaw position, further analysis of the approximation error was performed for the second order system with a wider range of system constants. The approximation error for the underdamped second order system was examined. The possibility of a difference in the time constant between the opening movement and the closing movement was also taken into consideration.

Fig. 8 shows the magnitude of the average approximation error and the contextual variation in the target position. It can be seen that the approximation error is the smallest for the critically damped system with the time constant of 40msec. However, the approximation error does not increase significantly when the damping factor becomes 0.7 (under damped system).

As for the contextual variation, the minimum value was observed for the condition where the opening time constant was 40msec, the closing time constant 30msec and the damping factor 0.7. This is mainly due to the fact that the two types of characteristic variation observed for the critically damped system (mentioned at the end of the preceding section) were reduced by adopting the underdamped system and the smaller time

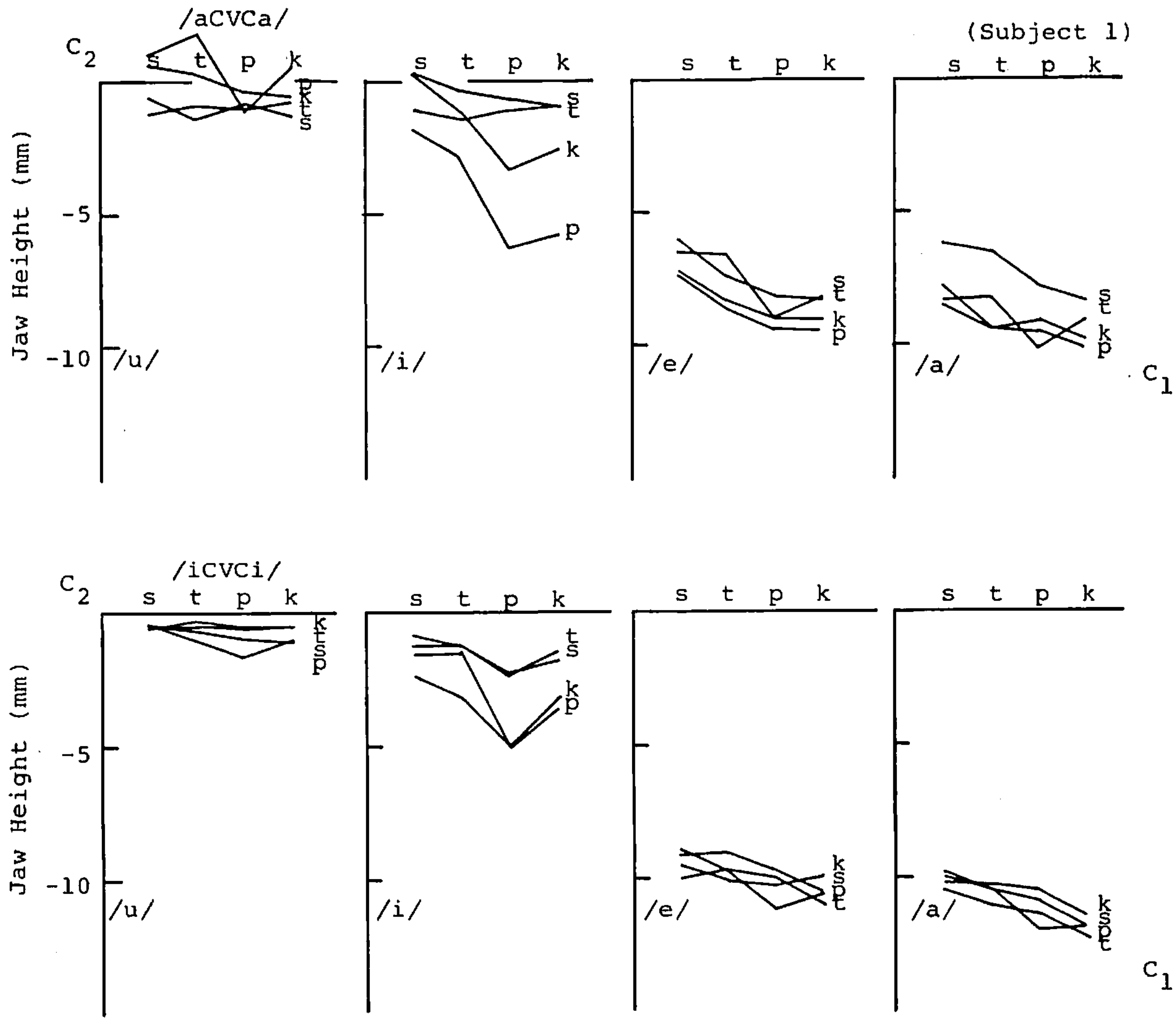

Fig. 7 Target position in each consonant context estimated for the critically damped second order system with a time constant of 40msec.

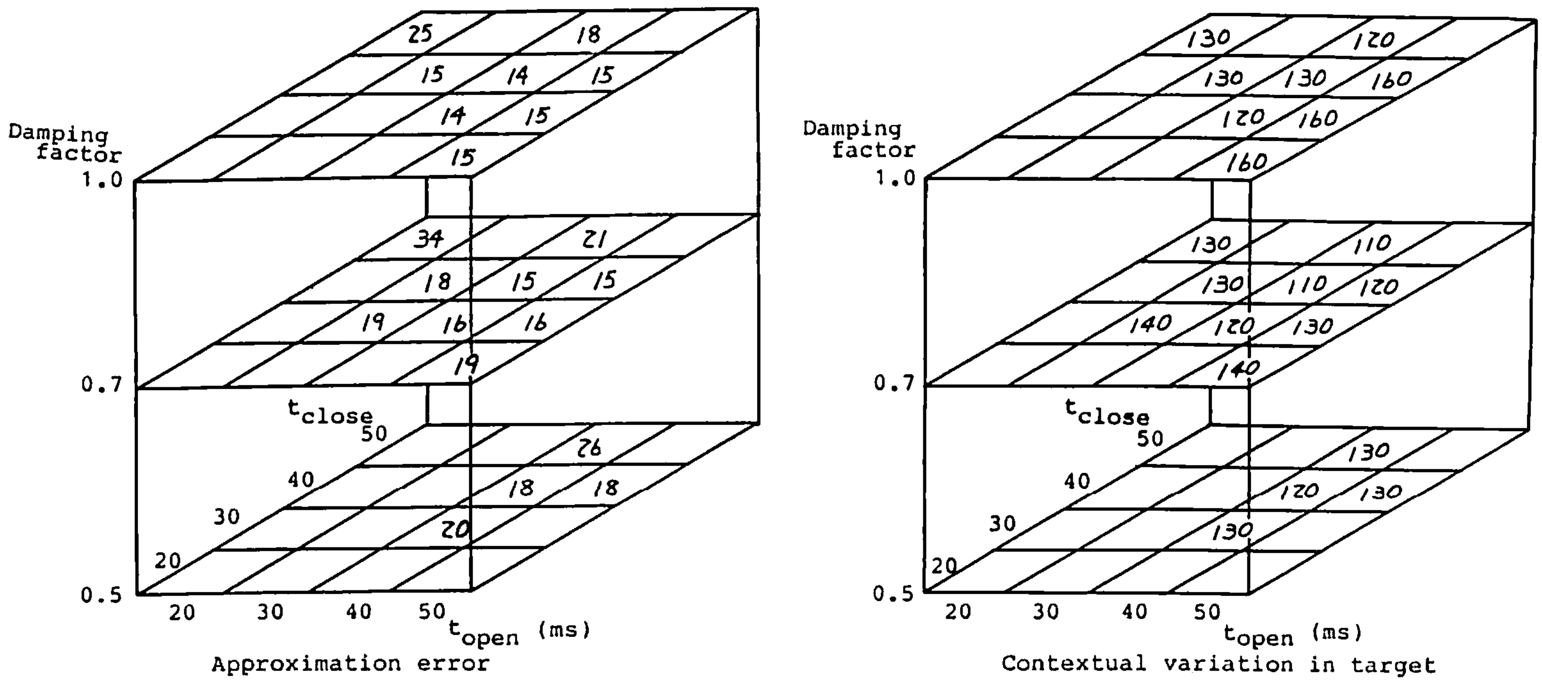

Fig. 8 Approximation error and contextual variation in the target position for selected conditions of the time constant and the damping factor .

constant for the closing movement. When the system was underdamped, the average target positions for /a/ and /e/ in /iCVCi/ became higher and came close to the target position in /aCVCa/. For the vowel /u/, the target positions in /apuCa/ and /akuCa/ became lower. In addition to this, when the time constant of the closing movement was made smaller, the target positions in /asuCa/ and /atsuCa/ became lower. As a result, the variation in the target position for /u/ became smaller than in the case of the critically damped system.

Summary

In the present study, the movements of the jaw in the production of CVC sequences were observed and the contextual variations in the position of the jaw for vowels were analyzed. In the case of CVC utterances, the degree of the jaw opening for the vowel became smaller than that for the stationary vowels, and there were considerable overlaps in the range of the jaw opening among the vowels. However, the differences among the vowels were still statistically significant, and for vowels in the same context the degree of the jaw opening was in the order of /u/, /i/, /e/ and /a/ in almost all the consonant contexts, which is the same order as that for the stationary vowels.

An analysis based on the approximation of the jaw movement by the response of the linear second order system model showed that the minimum contextual variation in the target position of the vowel could be obtained for the underdamped second order system. This result agrees with the result of our previous analysis of the jaw position of the consonant in /VCV/ utterances. It appears that the dynamic characteristics of the movement of the jaw during speech are better represented by the underdamped system than by the critically damped system. The present analysis also suggests that, for some vowels, a smaller time constant for the closing movement may give a better approximation. The characteristics of this effect should be further examined.

Reference

1. Kiritani, S. and T. Tanaka, K. Hashimoto, S. Masaki and K. Shirai(1983); Contextual Variation of the Jaw Movement for the Intervocalic Cconsonant in VCV Utterances, Ann. Bull. Rilp, 17,45-53.