Ann. Bull. RILP (1983) No. 17, 45-53

CONTEXTUAL VARIATION OF THE JAW MOVEMENT FOR THE INTERVOCALIC CONSONANT IN VCV UTTERANCES

Shigeru Kiritani, Toshiaki Tanaka, Kiyoshi Hashimoto,**
*Shinobu Masaki,** Katsuhiko Shirai***

Introduction

The movements of the jaw in the production of VCV sequences were observed, and the influences of the vowel context on the jaw movements for the consonant were examined. The jaw movement was approximated by the response of a linear second order system to the input step function, which was considered to represent the target positions of the successive phonemes, and the variation in the target positions due to the context were analyzed.

Data recording

The movement of the jaw was recorded using a PSD (optical spot position sensitive detector). Two infrared light emitting diodes were attached to a solid steel wire which was attached to the lower front tooth, 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. These LEDs were used to monitor the possible movement of the head during speech and to calculate the movements of the lower jaw LEDs relative to the upper jaw. A PSD was located within a camera body, and the image of the LED was formed on the PSD through the lens of the camera. The four LEDs were turned on and off one by one in sequence. From the PSD, the current signals x_1, x_2 and y_1, y_2 related to the x-coordinate and y-coordinate of the LED, respectively, were obtained. The coordinate values X and Y of the image of the LED were calculated by the following equations.

$$x = \frac{x_1 - x_2}{x_1 + x_2}, \qquad y = \frac{y_1 - y_2}{y_1 + y_2}$$

In the present device, the above calculation was performed by the analog function circuit. The resulting coordinate signals were sampled by the computer at a rate of 100 frames per second together with the speech envelope signal. It was also possible to monitor the movements of the LEDs on an oscilloscope display by the off-line use of the device.

The speech materials used were meaningless sequences of the form VCV which were uttered within the carrier phrase ' desu.' The first and the second vowel were one of either /i/, /e/, /a/ or /u/, and the inter vocalic consonant was one of

* Electro Communication University
** Department of Electrical Engineering, Waseda University

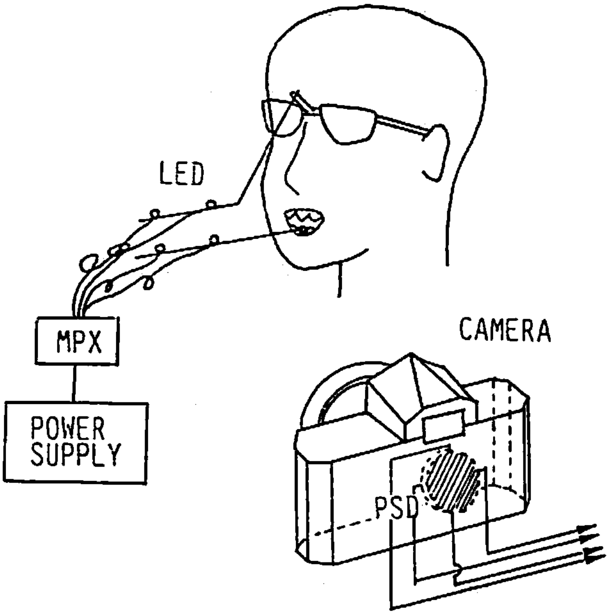

Fig. 1 *Jaw movement recording device using a PSD (optical spot position sensitive detector).*

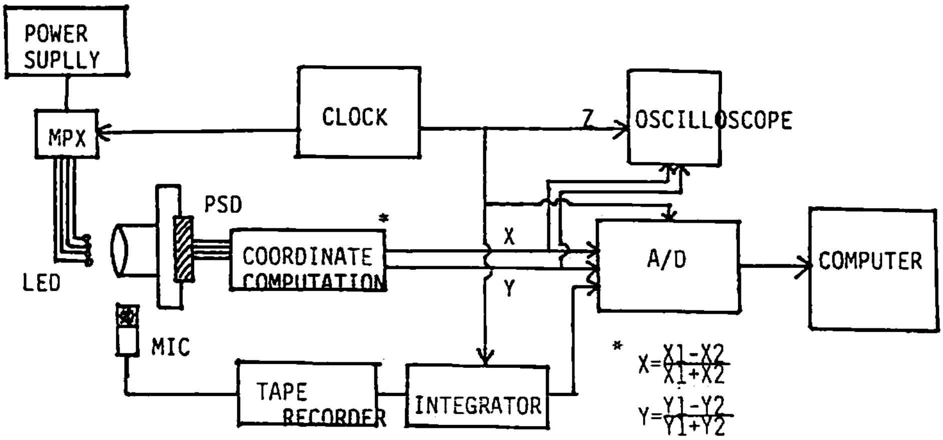

Fig. 2 *Block diagram of the control system of the jaw movement recording device.*

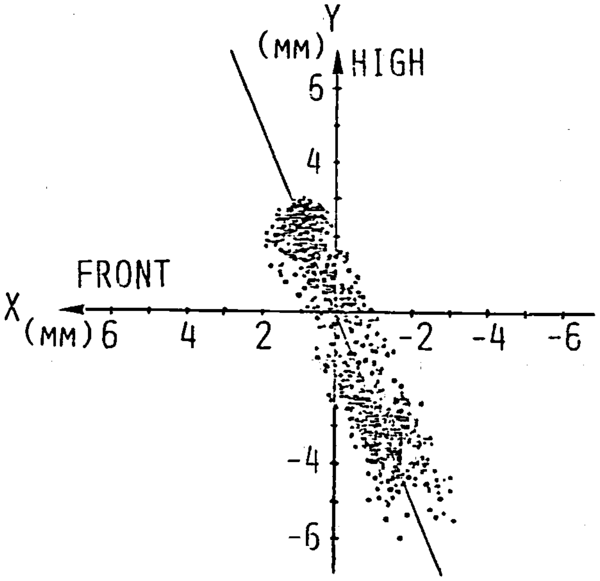

Fig. 3 *Range of the movement of the jaw. Each data point represents the position of the LED for V_1, C or V_2 in the individual utterances.*

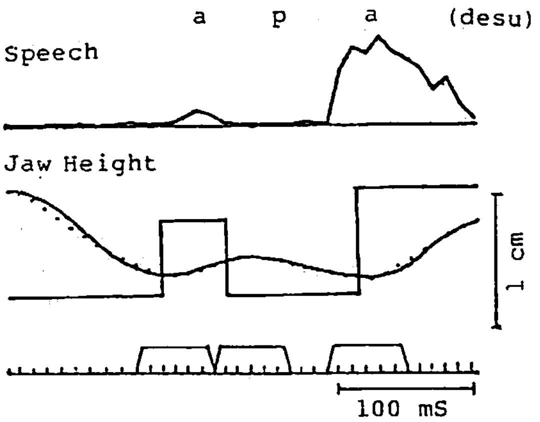

Fig. 4 *An example of the estimated input step function. Top; speech envelope. Middle; step function and the jaw movement (observed = solid line, calculated = dotted line). Bottom; time intervals of the possible step changes.*

either /s/, /t/, /p/ or /k/. Five tokens of the utterances were recorded for each VCV word. The duration of each sentence was 500 msec, approximately.

In the present study, the movement of the jaw was regarded as one dimensional, and the time functions of the y-coordinate of the lower jaw LED were analyzed. Figure 3 shows the range of the movement of the jaw over the entire speech material. For each VCV utterance, the mid point of the first vowel, the consonant and the second vowel were determined by a visual inspection of the speech envelope, and the position of the LED at these time moments were plotted on the xy-plane. It was observed that the range of movement along the regression line for all of the data points was 9.4 mm, and the range of the movement perpendicular to the regression line was 2.1 mm.

Analysis based on the critically damped linear second order system

1. *Method of analysis*

In the present analysis, it was assumed that the motor command for controlling the jaw movement could be represented by the step function which specified the degrees of the jaw opening for the consecutive phonemes. For each VCV utterance, the observed jaw movement was approximated by the response of a linear second order system to the input step function, and the step function which gave the best approximation was estimated.

Figure 4 shows an example of the time function of the jaw movement and the estimated input step function. In the figure, the curve at the top is the speech envelope. In the middle, the observed jaw movement is shown by the solid line together with the estimated step function. The dotted line is the jaw movement calculated as a response of the second order system to the step function. In determining the input step function, the time intervals of the possible change in the step function were determined through a visual inspection of the recorded curve of the jaw movement. At the bottom of the figure, selected time intervals for the transition from V_1 to /p/, /p/ to V_2 and V_2 to /d/ are shown. The search for the best approximation was performed using the technique of dynamic programming. In the following, the level of the step function determined will be refferred as the target level for each pertinent phoneme.

2. *Results*

Figure 5 shows the target levels estimated for the first vowel, the intervocalic consonant and the second vowel in each type of VCV sequence. Analysis was performed using three different values for the time constant of the second order system, i.e., 20, 40 and 60 msec. It was found that for the time constant of 40 msec, the average approximation error over all of the utterances was minimal. The results obtained for this value of the time constant are shown in the figure. Figure 5(a) shows the target levels of the consonants /s/, /t/, /p/ and /k/. The x-axis in the figure represents the first vowel and the y-axis, the second vowel. The z-axis shows the target level of the jaw position. A target level of zero corresponds to the closed position of the jaw.

As for the pattern of the contextual variation in the target levels of the con-

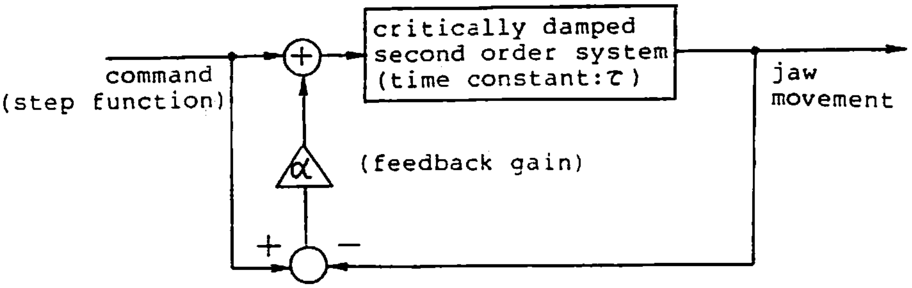

Fig. 6 *Critically damped second order system with compensatory feedback.*

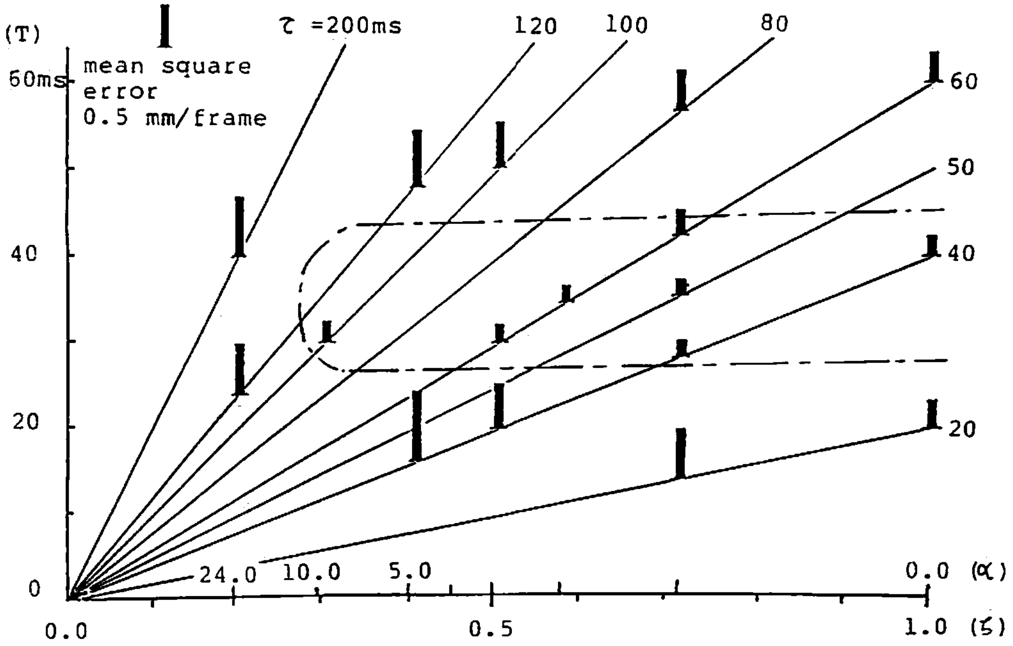

Fig. 7 *Approximation error for various values of the model parameters.*

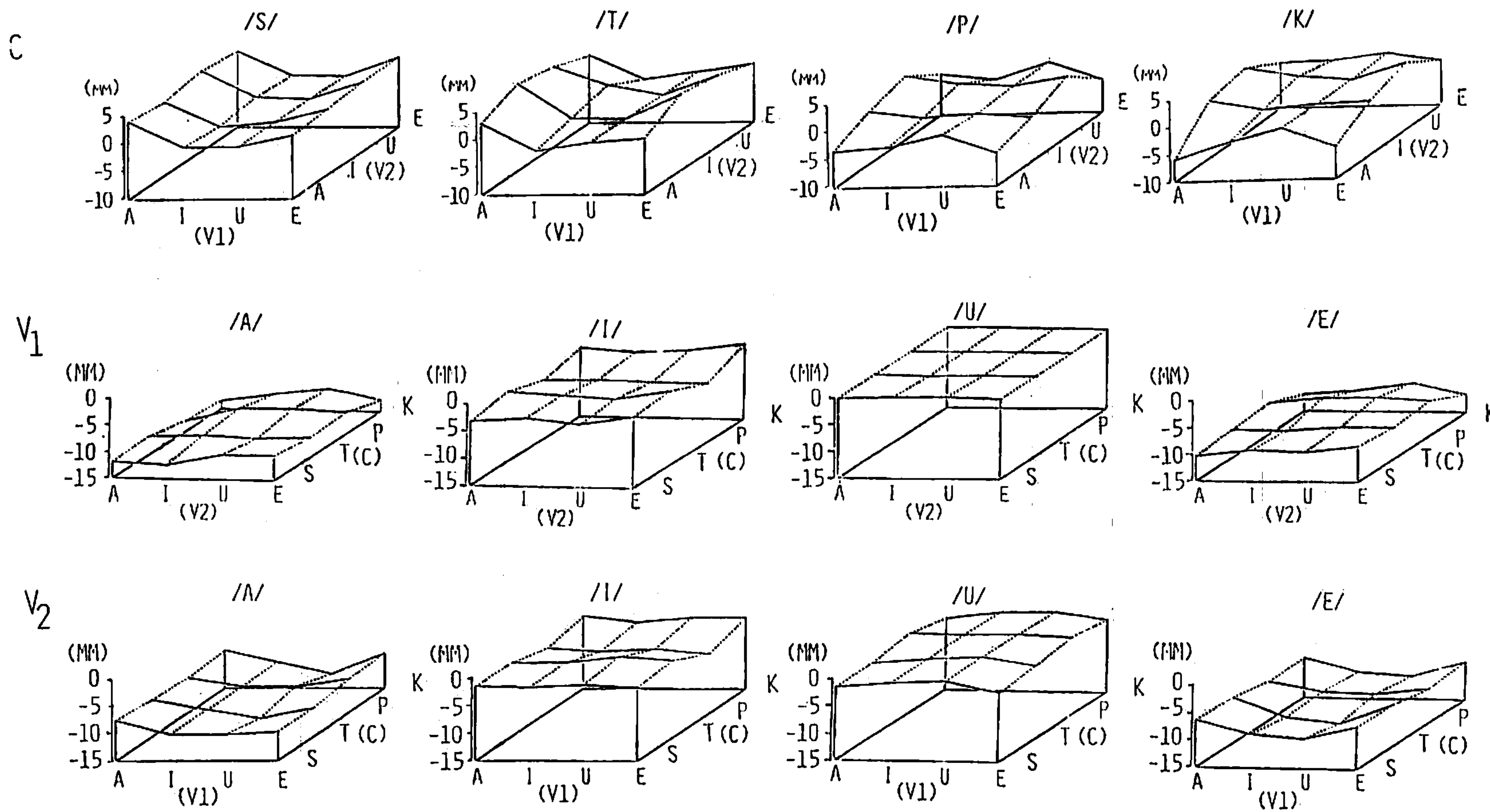

Fig. 5 *Target levels estimated using a critically damped linear second order system (time constant = 40 msec).*

sonants, the following tendencies can be observed in the figure. First, it can clearly be seen that, for the consonants /s/ and /t/, the target levels are higher when the first vowel is /a/ or /e/ than when the first vowel is /i/ or /u/. For example, the target levels of /s/ in /asa/ and /esa/ are higher than in /isa/ or /usa/. Similarly, the target levels of /s/ in /asi/ and /esi/ are higher than in /isi/ or /usi/. Namely, the target levels of the consonants /s/ and /t/ are higher when the first vowel is an open vowel than when it is a closed vowel. This tendency may be termed the "contrast effect.' This contrast effect can be seen also for the consonants /p/ and /k/.

Another phenomenon which can be observed in the figure is that the target levels of /p/ and /k/ become lower when the open vowel /a/ or /e/ appears in the vowel context. This effect may be termed the 'assimilatory effect.'

For the target levels of the first vowel in the VCV sequences, it is hard to determine any systematic effect due to context. As for the second vowel, the target levels of the open vowels /a/ and /e/ are lower when the first vowel is the closed vowel /i/ or /u/. This effect can also be termed the 'contrast effect' of the first vowel on the second vowel.

Analysis based on a critically damped linear second order system with compensatory feedback

1. Method of analysis

The 'contrast effect' observed in the above analysis can be realized effectively within a continuous system model by adopting a critically damped second order system with compensatory feedback (Fig. 6). In this system, extra force is automatically generated when the difference between the level of the input signal and the current level of the jaw opening is large. Using this model, the input step function for each VCV utterance was estimated.

First, the magnitude of the approximation error for the various values of the model parameters, i.e., the time constant τ and the gain of the feedback α, were examined. It can easily be confirmed that the system shown in Fig. 6 is, as a whole, equivalent to an under-damped linear second order system. The time constant T and the damping factor ζ of the equivalent system are given as follows.

$$T = \frac{\tau}{\sqrt{1+\alpha}}, \qquad \zeta = \frac{1}{\sqrt{1+\alpha}}$$

In Fig. 7, the approximation errors for the various conditions of the parameter values are plotted on the T-ζ plane. The approximation error was calculated for a restricted set of utterances (i.e., /ata/, /ati/, /ita/, /aka/, /aki/ and /ika/) to reduce the total amount of computation. It can be seen in the figure that, for the parameter values within the area surrounded by the dashed line, the approximation errors are relatively small. For several conditions of the parameter values within this area, the input step functions were estimated for the entire set of VCV utterances, and the pattern of the contextual variations in the target level of each phoneme was examined.

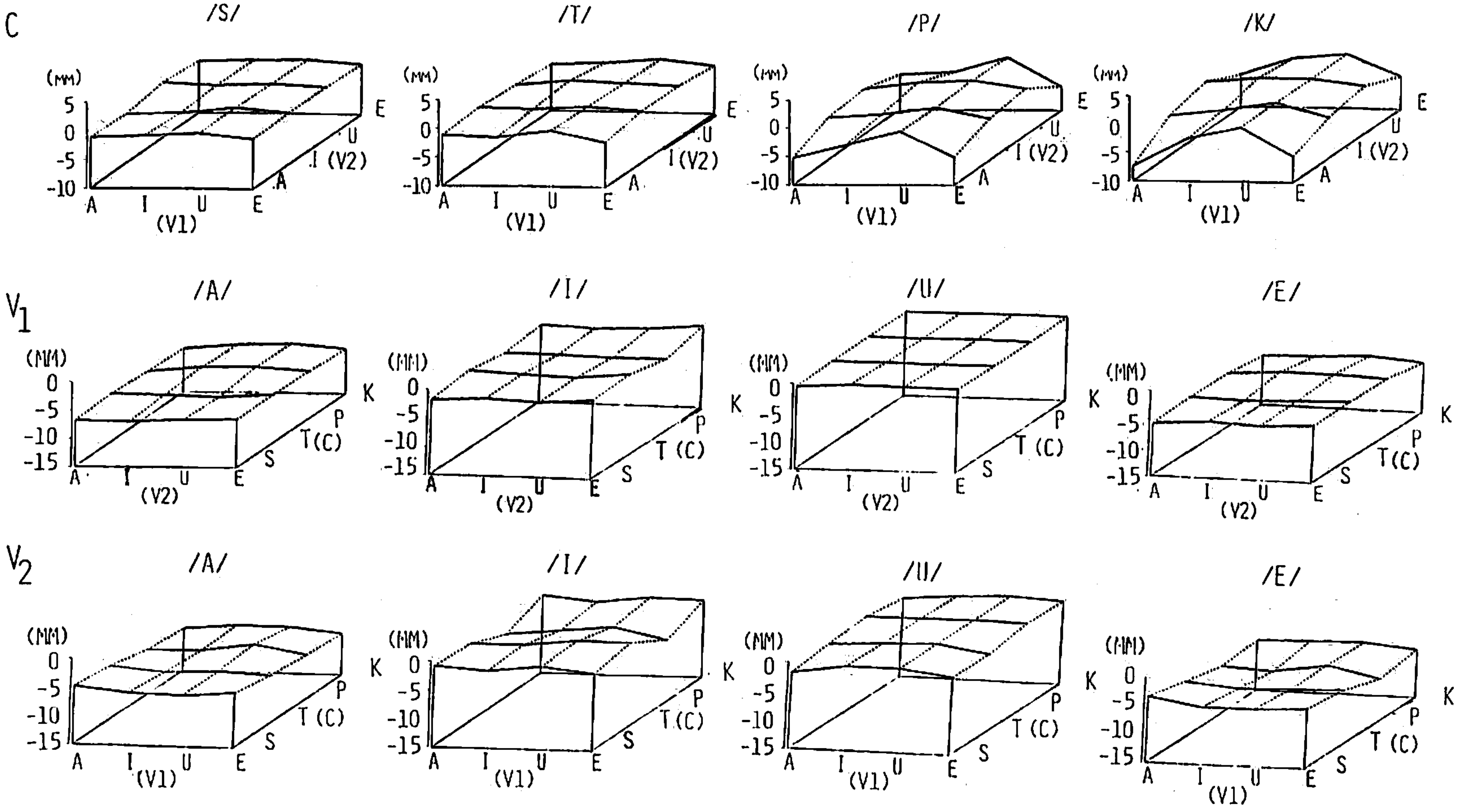

Fig. 8 *Target levels estimated using a critically damped second order system with compensatory feedback.*

2. *Results*

Figure 8 shows the target levels obtained for the time constant of 60 msec and the feedback gain of 3.0. It was observed that, compared to the results shown in Fig. 5, the target levels of the consonants /s/ and /t/ were nearly constant over different vowel contexts. This can be interpreted as indicating that the 'contrast effect' observed in the previous section was effectively realized by the compensation feedback, and, at the input stage to the system, the variations in the command level due to the vowel context was reduced to a large extent.

This reduction of the 'contrast effect' was also observed for the consonants /p/ and /k/. As a result, the remaining component of the contextual variation in the target levels of /p/ and /k/ can be considered to be mostly due to the 'assimilatory effect.' That is, in the context of open vowels, the target levels of these consonants are lower.

The reduction of the 'contrast effect' was also observed for the second vowels /a/ and /e/. For these vowels, the variation in the target levels due to the difference in the first vowel was smaller than in Fig. 5.

For a feedback gain value greater than 3.0, there appeared again a contextual variation in the target levels of /s/ and /t/. In this case, the target levels for these consonants were higher when the first vowel was an open vowel. Thus, the pattern of the contextual variation in the target levels was not as simple as that shown in Fig. 8.

Summary

In the present study, a jaw movement recording device using a PSD (optical spot position sensitive detector) was devised, and the movement of the jaw for VCV sequences was observed. The jaw movements were approximated as the response of a linear second order system to an input step function which specified the target positions of the jaw for the successive phonemes. The pattern of the contextual variations in the target position for each phoneme was analyzed.

Analysis using a critically damped linear second order system showed that the target levels for the consonants /s/ and /t/ varied according to the nature of the preceding vowel, i.e., when the preceding vowel was open, the target level was higher.

When a critically damped second order system with a compensatory feedback was adopted, the target levels for /s/ and /t/ became nearly constant, regardless of the vowel context. The remaining contextual variations in the target levels of /p/ and /k/ can be considered as mostly due to the influence of the open vowels which lowered the target levels of the adjacent consonants.