Ann. Bull. RILP (1982) No. 16, 103-124

A STUDY OF RHYTHM PERCEPTION IN REPETITIVE SOUND SEQUENCE

Seishi Hibi

1. Introduction

There is little doubt that rhythmic variables play a significant role in the intelligibility of speech. Not only in speech, a simple rhythmic structuring is inherent in most other human behaviors also. There have been many studies which have elucidated both qualitative and quantitative characteristics of rhythmic activities.

In perception of a regular sequence of stimuli that is neither too rapid nor too slow, we tend to perceive it as rhythmic (Woodrow, 1951; Fraisse, 1963). Allen (1975) stated that the time interval between pulses had to be greater than about 0.1 second in order to be heard as a succession of pulses, and the interval had to be less than about 3.0 seconds in order to be heard as a group of pulses. Woodrow (1951) reported that the perceptually "preferred" rate of succession, which was indicated by the time interval between successive clicks, was between 0.2 and 1.3 seconds. He also introduced the investigation conducted by Blakely (1933), who measured the accuracy of discrimination of intervals bounded by clicks, and found discrimination to be most accurate at intervals of 0.6 and 0.8 seconds. At these lengths the just noticeable relative difference was slightly less than 8 percent of the standard, and increased both above and below this middle region.

When a regular sequence of sounds, like short tones, clicks, or monosyllables, is presented to listeners, it is perceived as rhythmic. However, if the temporal regularity of the series is distorted or perturbed, we hear some irregularity therein and it is perceived as arhythmic. The perception of a rhythm line of music involves a discrimination of different time intervals. For example, if the tape speed happens to change because of a short break in the electric current, one might well point it out. In human speech, we find several instances of apparent temporal irregularity not only in dysarthric patients (Hirose et al., 1978, 1979, 1980) but also even in normal speech, on the evidence of acoustic measurements (e.g. Klatt, 1976; Lehiste, 1973). It must be noted here, however, that the temporal irregularity in normal speech which was found in their experiments was interpreted as a perceptual cue which conveyed information about the linguistic content of an utterance.

On production of rhythm, many studies have investigated the rate of succession of rhythmic beats. The rhythmic beats in speech are, however, not necessarily placed on the similar position of the sound sequence. The interval between the stressed syllables alternating with all intervening unstressed syllables (Chomsky and Halle, 1968) is held nearly equally in the so-called "stress-timed" languages and, as opposed to this, the length of each syllable is maintained to be almost equal in the

---

A dissertation submitted to the Faculty of Medicine in partial fulfillment of the requirements for the degree of Doctor of Medical Science, University of Tokyo.

"syllable-timed" languages (Pike, 1945). Several studies (Abe, 1967; Allen, 1972) showed interstress intervals in the stress-timed language fell between 0.2 and 0.8 seconds. Sato (1977) measured the rate of succession of morae and the dependency of duration of each syllable on the number of morae combined in a rhythmic group in Japanese speech. Most of his measurements fell between 0.1 and 0.2 seconds per mora and the duration of each syllable decreased with the increase in number of morae per group. When the number of morae was equal to or greater than 6, duration was independent of this parameter. In French, the rate of succession of syllables was about 0.15 to 0.2 seconds per syllable, and the number of syllables in an utterance group was 2 to 11 (Malécot et al., 1972). They found a strong positive correlation between the rate of succession of syllables and the number of syllables in an utterance group. The rate of each utterance group was, therefore, between 0.4 and 1.7 seconds. In production of non-speech rhythm, Fraisse (1963), for example, showed that the time interval between the key notes in a musical composition was statistically found to be between 0.15 and 0.9 seconds.

In terms of other characteristics of rhythmic action, various researchers (Woodrow, 1951; Fraisse, 1963; Michon, 1966, 1967; Hibi, 1980) have explored the variability thereof. They have shown that the overall range of standard errors was about 3 to 11 percent of the length of the interval, when the subject produced an even tempo. Allen (1975) reported that the variability of production of speech sequence matched the variability of other rhythmic activities such as finger-tapping. Namely, short speech segments had variability of about 10 percent, longer stretches of speech about 4 percent, while the overall range for standard errors for motor rhythms was, as mentioned above, 3 to 11 percent of the length of the time intervals being produced.

The argument so far is that we perceive some sequential sounds as rhythmic when they have regular sequential time intervals and we produce regular sequential time intervals when we act rhythmically. The rate of succession within a range of so-called perceptually "preferred" rates matches the rate of succession of rhythmic activities. Moreover, the variability of production of rhythmic activities is not far different from the most accurate level of discrimination of time intervals.

Since it is the general notion that production of rhythm and perception thereof are dynamically coupled, there should be a mechanism which governs the timing controls of temporal sequences in both perception and production of rhythm. The objective of the present study is, therefore, to explore an identical timing-related threshold which would be common both in perception and production of rhythm.

2. Perception of Rhythm

The problem for experimental investigation here concerns the ability of listeners to tell whether or not there is a distortion in a sequence. It turns out that there are many examples, as mentioned in the preceding section, in daily listening which seem to require such a judgment for ordinary perceptual recognition to take place. We therefore wish to know how greatly a temporal distortion must intervene in uniformly spaced members of sounds in order for a listener to be able to report an irregularity.

2.1 Experiment A

2.1.1 Method

2.1.1.1 Stimulus Materials

Uniformly spaced temporal sequences consisting of 15 tone bursts, each of which was 1 kHz in frequency and 5 milliseconds in duration and with different rates of succession ranging from 0.7 to 7.0 times per second, were used as basic (i.e. undistorted) sequences. The rates of succession were 0.7, 1.0, 1.3, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0 and 7.0 times per second. Translated into a time interval measure, they were 1500, 1000, 750, 500, 400, 333, 250, 200, 167 and 143 milliseconds, respectively. Distorted versions of each basic sequence were prepared as follows. One interval in the sequence was either lengthened or shortened by 2, 4, 6, 8, 10, 12, 14 and 16 percent of the basic interval, yielding 16 distorted versions of the sequence. Fig. 1-A shows the examples of lengthening and shortening of the interval in the sequence schematically.

The distorted time interval appeared in serial position 7, 8, or 9. Namely, the length of a time interval bounded by the 7th and 8th tone bursts, by the 8th and the 9th, or by the 9th and the 10th, was either lengthened or shortened. There were therefore a total of 480 distorted versions, depending upon the serial position of the distorted interval. With three copies of each basic sequence version included, there were therefore a total of 510 sequences in the experiment. The 510 sequences were assembled into three blocks of 160 distorted versions and 10 basic sequence versions. Experimental conditions were assigned quasirandomly in such a fashion that experimental conditions were distributed evenly within and between blocks.

All of the sequences were prepared by means of PDP 11/34 computer routines as follows. The following three parameters were accepted: a basic time interval which corresponded to a rate of succession of stimulus sequence, a degree* of distortion, and a serial position where the distortion was placed. Then a digitized tone was readout repeatedly with the aid of a built-in clock. The synthesized sequence of the digitized tones was then digital-to-analog converted, and recorded on videotape by CPM recording system through a low-pass filter set at 4 kHz.

2.1.1.2 Subjects and Procedure

Three subjects were employed in the present experiment: two male adults who were skilled in listening tests of synthesized speech, and a female otolaryngologist. All had normal hearing. The subjects participated in a total of 12 sessions, each of which lasted approximately 40 minutes. There was a rest interval after every one-third of the session. During each session, the subjects sat in a sound-proof room and listened to the test stimuli through a loudspeaker. Tone bursts were set at 75 dB SPL.**

---

*NOTE: The degree of distortion is defined as $\Delta T/T$, where ΔT is the length of the time interval either added or subtracted, and T is the length of the basic time interval.

**NOTE: The sound pressure level of test stimulus is expressed in the case of continuously presented 1000 Hz sound.

The subjects were required to judge whether there was a distortion in the sequence by writing down either "detected" or "not detected." The choice between these two alternatives was forced. A pause of 7 seconds in duration was placed between each presentation of the sequence, during which the subjects wrote their judgments on a form.

2.1.2 Results

2.1.2.1 Data Analysis

For each of the three subjects, the number of occasions out of a total of 12 trials that he or she reported "detected" was recorded. Then the average percentage of judgments of the "detected" distortions for all subjects was plotted on normal probability paper, on the assumption that the function relating the probability that a subject would respond "detected" to the degree of distortion would be a normal ogive. One example of the values is presented in Fig. 2, where each asterisk (*) represents the mean relative frequency of judgments "detected" among 12 trials on the 3 subjects.

The abscissa of Fig. 2 shows the degree of distortion in percentage. The assumption that the relation between response probability and the degree of distortion would be a normal ogive appears tenable because the distributions for all of the different conditions, i.e. different rates of succession and "lengthening" or "shortening," appear to fall reasonably well on a straight line. (Graphs are not shown for all conditions.) The line in the figure represents straight-line fits to the data by means of the least-square solution using Müller-Urban weights.

The degree of distortion which afforded fifty percent judgments of detection (hereafter, DL) for either the lengthened or shortened sequence of each rate of succession was calculated. These values are presented in Fig. 3-A, where each filled circle (●) and open circle (○) represents the degree of distortion (DL) under the condition of "lengthened" and "shortened," respectively. The abscissa of the figure shows the rate of succession ranging from 0.7 to 7.0 times per second in logarithmic scale.

The "detected" or "not detected" data were submitted to analyses of variance using the above-mentioned DLs as cell entries, and using either subjects or serial positions of distortion as sampling units. Since almost all analyses yielded similar results, only mean values of the degree of distortion (DL) across subjects (serial positions of distortion as sampling units) are reported below.

The experimental factors in the overall analysis (see Table 1-A) were rate of succession (0.7 to 7.0 times per second), position of distortion (serial position 7, 8, or 9), type of distortion ("lengthened" or "shortened"), and interactions among these factors.

2.1.2.2 Results

As Fig. 3-A shows, the DL value of the degree of distortion varied depending upon the rates of succession of the sequences. The results of statistical analyses of data support this; namely, the "rate of succession" factor was significant. The interaction between "rate" and "type of distortion" reached significance, although the "type of distortion" factor was not significant.

Since we could readily see from the figure that there was no great difference among the DL values which were obtained at the lower rates of succession, i.e. 1.0 to 2.5 times per second, and that there seemed to be some difference among higher rate data, the DL data were divided into two groups in terms of rate of succession and analyses of variance were carried out for the data in each group separately. The following results were obtained by the local analyses of variance. First, the variations of DL among the lower rates of succession, i.e. 1.0 to 2.5 times per second, were relatively small. (F ratio was just smaller than the significant F(0.05).) There was no significant difference between "lengthened" and "shortened" among these rates of succession. The mean value of DLs among these rates was 6.1 percent. Second, among the higher rates of succession, i.e. 4.0 to 7.0 times per second, the variations of DL were smaller too (F ratio was much smaller than F(0.05)), but this time the difference in DL values between the two types of distortion was significant. The mean values of DLs among these higher rates were 8.9 percent and 8.2 percent for "lengthened" and "shortened," respectively.

Of particular interest here is a difference between "lengthened" and "shortened" at a rate of 3 times per second of succession (see Fig. 3-A). The values of DL for the different types of distortion differed to a great extent. A more precise experiment will be introduced later in Experiment C.

2.2 Experiment B

2.2.1 Method (Stimulus Materials, Subjects and Procedure)

Whereas a temporal interval in the sequence was either lengthened or shortened in Experiment A, a temporal allocation of the tone burst was slightly changed for Experiment B. In other words, when the time interval just preceding a tone burst was lengthened or shortened, the time interval which succeeded this particular tone burst was shortened or lengthened by the same amount in order to maintain the whole length of the sequence to be the same. Fig. 1-B shows this type of temporal distortion, i.e. lengthening/shortening or shortening/lengthening, of the sequence schematically. The distorted temporal allocation of a tone burst appeared in serial position 7, 8, or 9 in this experiment too. Namely, the temporal allocation of the 7th, 8th, or 9th tone burst was moved either forward or backward by ΔT. For example, the distortion of the temporal allocation of the 8th tone burst was realized by "lengthening" or "shortening" the time interval bounded by the 7th and 8th tone bursts, and by "shortening" or "lengthening" the interval bounded by the 8th and 9th tone bursts.

The method for preparing the tone burst sequence, subjects and procedure were the same as those described for Experiment A.

2.2.2 Results

The raw data were processed in the same way as described for Experiment A, and the degrees of distortion which afforded fifty percent judgment of detection (DL) were calculated. These values are presented in Fig. 3-B, where each filled square (■) and open square (□) represents the DL value of degree of distortion under the condition of "lengthened/shortened" and "shortened/lengthened," respectively. The abscissa in the figure shows the rate of succession in logarithmic scale.

It was found that the DL value of degree of distortion varied depending upon the rate of succession in this experiment too. In the analysis of variance (see Table 1-B), the "rate of succession" factor was also significant here. The difference between "types of distortion" ("lengthened/shortened" vs. "shortened/lengthened") was not significant. Variance for this factor was negligibly small so that there were no different DLs between "types of distortion" at all. This result was sound, holding across all rates of succession except for 3.0 times per second. Although variance for this factor was very small, the interaction between rate and type was significant.

Since we could readily see from the figure that there seemed to be no particular variations among DL values obtained in the lower regions of rate of succession (i.e. 1.0 to 2.5 times per second), or among these obtained in the higher regions of rate (i.e. 4.0 to 7.0 times per second), the DL data were divided into two groups in terms of rate of succession and analyses of variance were carried out for the data in each group separately. The local analyses of variance showed the following. The variations of DL among the lower rates of succession, i.e. 1.0 to 2.5 times per second, and those among the higher rates of succession, i.e. 4.0 to 7.0 times per second, were small. There was no significant difference between types of distortion both among the above-mentioned lower and higher regions of rate of succession. The mean values of DL were 6.0 percent and 7.6 percent for the lower rates and the higher rates of succession, respectively. Almost the same values of DL for the lower rates of succession were found in both Experiment A and Experiment B.

2.3 Experiment C

Of particular interest in the results of Experiment A is, as mentioned earlier, the difference between the DL values which were obtained by lengthening and shortening at a rate of 3 times per second.

In addition to Experiment A, the present Experiment C was carried out in order to explore a threshold by which "detected" or "not detected" judgment would be divided into two categories. The stimulus materials were prepared in the same way as described in Experiment A. The rates of succession of the stimulus sequence were, however, set at 370, 350, 330, 310 and 290 milliseconds this time. The other conditions were exactly the same as in Experiment A.

The results are shown in Fig. 3-C. The degree of distortion which afforded fifty percent judgments of detection for either lengthened or shortened sequences of each rate of succession is represented by a filled circle (●) and open circle (○), respectively. The abscissa of the figure shows the basic rate of succession in logarithmic scale. The figure shows the same data as those shown in Fig. 3-A, except that it carries additional results also. In the figure, we can see the difference in DL values between "lengthened" and "shortened" at around the rate of 3 times per second. The lengthened interval might act as the slower rate of succession and the shortened interval vice versa, even if the rest of the intervals were left unchanged. Fig. 3-C′ is redrawn so as to make it clearer in order to show what happens in this region. The abscissa of the figure shows the particular time interval which is either lengthened or shortened by the above-mentioned degree in logarithmic scale. The ordinate contains DL values. The figure shows that the time interval of about 330 milliseconds forms the boundary between the two categories.

3. Production of Rhythm

The problem for experimental investigation here concerns the ability of speakers to produce a rhythmic sequence with as few fluctuations as possible. It turns out that there are numerous examples in music performance which seem to require such a procedure. For example, the conductor of an orchestra brings his baton up and down in order to transmit his own "tempo" to his members, and they reproduce or follow their conductor's rhythm. In production of speech, we seem to be able to count numbers in time with the ticking of a clock. We therefore this time wish to know how accurately we can produce a temporal sequence.

3.1 Experiment D

3.1.1 Method

The stimulus signals were uniformly spaced temporal sounds which have a rate of repetition ranging from 1 to 6 times per second. A sequence consisting of 100 tone bursts, each of which was 1 kHz in frequency and 5 milliseconds in duration, was recorded in advance on audio tape to serve as signal stimuli.

Nine subjects participated in the present experiment: seven male and two female adults, between 25 and 45 years old. The subjects were requested to repeat the monosyllable /pa/ at least 50 times in time with the pre-recorded signal stimuli, and their utterances and stimulus signals were recorded on a 2-channel tape recorder simultaneously. The stimulus signals were presented to the subjects in a sound-proof room through headphones at 60 dB SPL.

The recorded utterances and stimulus signals were played back and stored digitally on data files in a laboratory computer. The digital sound waves of the utterances were displayed on a graphic terminal, from which the vowel onset of each utterance /pa/ was determined visually. The time interval between each utterance was thus obtained from the inter-vowel-onset interval, based on which statistical analyses were carried out.

3.1.2 Results

3.1.2.1 Mean and Standard Deviation

The mean and standard deviation of the time intervals were calculated from about 40 reproduced utterances for each different rate of succession of stimulus signal.

The mean time interval of the repetition of the temporal sequence, as responses in time with stimulus signals, was synchronous with the interval of the stimulus signals for every rate of succession (Fig. 4). The differences between the mean time intervals and the intervals of stimulus signals were within 2 percent. In the figure, the ordinate shows the mean time interval of the reproduction, and the abscissa shows the interval of the stimulus signals. Both are indicated in logarithmic scale. Standard deviations were small; namely, coefficients of variation (CV) fell between 3 and 9 percent almost independent of the rate of succession of stimulus sequence (Fig. 5). In the figure, the ordinate and abscissa show the coefficient of variation in percentage and the rates of succession of stimulus sequence, respectively.

3.1.2.2 Autocorrelation in Reproduction Time Intervals

The mean time interval of reproduction of the temporal sequence was clearly maintained to be synchronous with that of the stimulus signal, although each time interval of reproduction deviated to some extent. It is, therefore, plausible that there should be some adjusting mechanism which does work in order to maintain this synchronization. This possibility is suggestive of a negative correlation between the time intervals of reproduction.

To investigate the possibility, an autocorrelation function was introduced. On the assumption that the response time sequence would be a regular stochastic process, an autocorrelation function of the sequence $R(j)$ is defined as follows.
First, the covariance of the sequence is,

$$C(j) = \frac{1}{n-j} \sum_{i=1}^{n-j} (T_i - \bar{T})(T_{i+j} - \bar{T})$$

where, T_i is *i-th* time interval, T_{i+j} is *(i+j)-th* time interval, $\bar{T}$ is the mean time interval, and n is the total number of time intervals.

Of course when $j=0$,

$$C(0) = \frac{1}{n} \sum_{i=1}^{n} (T_i - \bar{T})^2$$

is the variance of the sequence. Then, $R(j)$ is obtained from $C(j)$ which is normalized by the variance $C(0)$:

$$R(j) = C(j)/C(0)$$

This parameter represents the correlation between time intervals which were distanced by j. Fig. 6-i and 6-ii show examples of $R(j)$ for the reproductions of the sequence in subject A, at rates of 2 times per second and 5 times per second, respectively. In the figure, the ordinate contains $R(j)$ values while the abscissa shows j, which is the distance between the time intervals whose correlation is shown by

$R(j)$. $R(j)$ values show a damped oscillation pattern in Fig. 6-i. In other words, the absolute values of $R(j)$ reduce as j increases. Of particular interest here, $R(1)$ and $R(3)$ in Fig. 6-i represent negative values, so that the neighboring time intervals have negative correlation. But in Fig. 6-ii, we can no longer see either the damped oscillation pattern or the negative correlation in $R(j)$ values. $R(j)$ values appeared randomly around zero. The time intervals come to be independent of each other, and the distribution of time intervals is non-correlative.

To clarify the trends, joint histograms are shown in Fig. 7-i and 7-ii, where the abscissa shows a time interval which is normalized by the mean interval of the sequence ($T_i/\overline{T}$) and the ordinate shows the time interval distanced by j also normalized by the mean ($T_{i+j}/\overline{T}$). Namely, each dot in the figure has horizontal and vertical coordinates of ($T_i/\overline{T}$, $T_{i+j}/\overline{T}$). In the figure, the value of j is limited within the range from 1 to 4, in order to examine the correlations between the neighboring time intervals. When there exist some negative correlations between the neighboring time intervals, most dots fall in the second and fourth quadrants and the joint histogram shows an oval shape which inclines toward left. The non-correlative distribution of the time intervals, on the contrary, does not result in such an oval shape. The results for subject A when he produced the time sequence at a rate of 2 times per second is, for example, shown in Fig. 7-i, in which we can see the oval shape. The results for the same subject when he produced the time sequence at a rate of 5 times per second is also, for example, shown in Fig. 7-ii, in which we find only non-correlative distributions.

Table 2 shows the results of inspection by autocorrelation and the joint histogram studies, where each circle represents that the negative correlation between the neighboring time intervals seemed to exist. In almost all of the subjects, the negative correlations were found only in the region of the lower rates of reproduction (i.e. 1.0 to 3.0 times per second).

4. Discussion

In the present experiments the author has attempted to examine the ability of human beings both in perception and in production of the simplest rhythm which has an even tempo. Fig. 3-A, 3-B and 3-C show the results in terms of the Weber ratio $\Delta T/T$ where ΔT is the increment or decrement necessary to give an average performance of fifty percent detection when added to a basic duration T. In Experiments A, B and C, Weber's law has, in a sense, been found to hold approximately for discrimination of the irregularity in the rhythmic sequence, but only in the restricted regions of the rate of succession. Namely, the ability of listeners to report whether there was a distortion in a sequence as a function of the rate of succession did not vary in the regions of from 1 to 2.5 times per second and from 4 to 7 times per second. However, the difference in the degrees of distortion which afforded fifty percent judgments of the "detection" in terms of the ratio $\Delta T/T$ between these two regions was significant.

The picture emerging from these results, then, is of the subjects' accommodating to the temporal sequences which have different rates of succession ranging from 0.7 to 7.0 times per second. After several sessions, the author asked the subjects how they felt the distorted part, or in which way they paid attention to the sequence. All answered that they listened to one tone burst at a time and assigned the time orientation (cf. Fraisse, 1957) of the next in perception of time sequence of the lower rates (i.e. from 1.0 to 2.5 times per second). Then, they attempted to look for a discrepancy between the actual timing and the predicted one. In perception of sequences of the higher rates of succession (i.e. from 4.0 to 7.0 times per second), on the other hand, they heard a group of tone bursts in which the distorted part might cause an impression of irregularity. In fact, some subjects did write down "detected" when they were only under the impression that there was a "stumbling" or a "slip" in the sequence.

There might be different sorts of processing in perception of temporal sequences, namely, one might be an ongoing processing in the region of lower rates of succession, and another a cumulative processing in the region of higher rates. This possibility led to the formulation of a modeling hypothesis, based on the results of both experiments in perception and in production of the temporal sequences.

In the rhythmic concept, it is inherent that perception of preceding events in a temporal sequence generates expectancies concerning forthcoming events (Martin, 1972). This mechanism appears to work both in an ongoing processing and in a cumulative processing of the temporal sequence, although it plays a different role in each processing.

Ongoing Processing

In an ongoing processing, one might first predict the timing of the next coming event, and then, check the actual event with the predicted one. The subject then reports "detected" if he finds a discrepancy in this checking process. The whole routine, therefore, appears to be an ongoing processing in the sense that the event is processed one-by-one.

The most direct evidence for the ongoing processing mechanism is that all types of distortion in both Experiment A and Experiment B yielded similar values of DL when the rate was set at 1.0 to 2.5 times per second. The ongoing processing mechanism should yield similar values of DL both in Experiment A and Experiment B because the listeners were required to detect only the first distorted interval which they encountered in perception of the sequence. In Experiment B, once a lengthened (or shortened) interval was detected, the following shortened (or lengthened) interval could play only a role in confirming the detection for the listeners. In Experiment A, the "lengthening" and "shortening" of the same degree could give rise to the same amount of discrepancy between the actually perceived timing and the predicted one. It was found, therefore, that the "ongoing processing" mechanism can work only in perception of the lower rates of succession (i.e. from 1.0 to 2.5 times per second).

As mentioned earlier in Results of Experiment D, the negative correlation among neighboring time intervals tended to restore the timing of the following utterances to their original regular relationship with the preceding utterances only in reproduction at the lower rates (i.e. 1.0 to 3.0 times per second). The ongoing processing mechanism accounts for the experimental results. In order to compensate a discrepancy between the timing of utterance and that of stimulus signal in an adjacent or neighboring utterance, one has to detect the discrepancy in an "ongoing" manner.

Wholistic Processing

In perception of sequences of the higher rates of succession (i.e. from 4.0 to 7.0 times per second), on the other hand, the detection of the distorted interval might require the following two steps. First, it is necessary to postulate a regular temporal pattern, that is, the pattern of time intervals that would yield an even tempo, and then to detect a departure from these regular values in the observed intervals by pattern-matching or similar routine. The mechanism of detection of the distorted interval which requires a two-stage analysis with the pattern-matching stage depending upon the preceding postulating stage bears some resemblances to the analysis-by-synthesis model of speech perception (Halle and Stevens, 1964). In this model, the determination of the time interval pattern is realized by means of a "wholistic" processing.

An intuitive interpretation for wholistic processing is that one can hardly process the ongoing event in a period of time as short as less than 250 milliseconds, although he can predict the timing pattern of the forthcoming events. In wholistic processing, one might first predict the timing pattern of the forthcoming events, but he can hardly check the ongoing event. He perceives and cumulates the successive events, and then attempts to compare the actually perceived pattern with the predicted one. Therefore, the routine appears to be a wholistic processing in the sense that the sequence is processed as a whole.

When we hear the stimulus sequence as a whole, or at least we hear the neighboring vicinity of the distorted interval in the sequence as a whole, the DL values obtained under one condition should differ from those obtained under another condition, because the virtual degree of distortion differs depending upon the way in which the time sequence is distorted. In fact, lengthening affected the discrimination less than shortening in Experiment A, since for a given degree of change ΔT added to or subtracted from the basic interval T, the relative change was less in the case of lengthening than in the case of shortening. In other words, the greater degree of distortion was needed in order to afford fifty percent judgments in the case of lengthening (8.9 percent) than in the case of shortening (8.2 percent). Another evidence for wholistic listening is that different values of DL were obtained in Experiment A and Experiment B. While the shortened (or lengthened) time interval, which succeeded the lengthened (or shortened) one, had little effect on the discrimination among the lower rates of succession in Experiment B, it enhanced the detectability of distortion among the higher rates. The probabilities that the distortion was detected at the lengthened interval and/or at the shortened interval were calculated from the results obtained in Experiment A, when the degree of distortion which afforded the fifty percent judgment in Experiment B was placed in the sequence. The probabilities were calculated on the assumption that the detection of the lengthened (or shortened) interval and that of the shortened (or lengthened) interval were independent phenomena, which can occur simultaneously (see Table 3). The calculated probabilities were nearly equal to 0.5 and appeared to match the observed ones. The distortion in Experiment B was, therefore, detected at the lengthened (or shortened) and/or the shortened (or lengthened) intervals. The present results which show the difference between Experiment A and Experiment B thus suggest a "wholistic" processing in the higher rates of succession.

As mentioned earlier in Results of Experiment D, the negative correlation among neighboring time intervals disappeared in reproduction at the higher rates of succession (i.e. 4.0 to 6.0 times per second). Although the rate of reproduction was fixed at almost the same as that of the signal stimuli, the detection of a time-lead or a time-lag by means of ongoing processing was no longer found. However, there seemed to be another kind of adjusting mechanism which contributed toward the synchronization between the reproduction and the signal stimuli, since the wholistic processing mechanism could analyze a dismatching of the pattern in due course.

The discussion so far makes it clear that different processing mechanisms govern the timing-related controls at the lower and higher rates of succession both in perception and in production of a rhythmic sequence. However, the measurements of DL value at a rate of 3 times per second fell into two categories, namely, the DL for the "lengthened" sequence showed a value which was similar to those obtained in the lower rate group, and the DL for the "shortened" sequence showed the value of the higher rate group. The more precise experiment in this region, i.e. Experiment C, showed that the time interval of 330 milliseconds formed the boundary between these two categories.

The results of other experiments in the literature with different methods and purposes are consistent with those of the present experiments. In perception of nearly identical temporal spacing, the prediction-comparison routine can be interpreted as a "rehearsal" of the sequence. In reproduction of the temporal sequence, on the other hand, the reproduction in time with the stimulus sequence can be interpreted as a "shadowing" of the sequence. Various authors (Norman, 1972, 1977; Waugh et al., 1965) described that we are able to rehearse or shadow the time sequence up to 3 to 6 times per second. From the time interval measurement view, every "rehearsal" or "shadowing" requires a time interval of about 170 to 330 msec, which would be interpreted as the time interval needed in order to readout an output from the short-term memory and then to rewrite it into the same memory for the rehearsal routine, and as that needed in order to readout an output from the short-term memory and then to effect an operation for the shadowing routine.

5. Concluding Remarks

There were three specific aims in the present study. They were:

1) to know how greatly a temporal distortion must intervene in uniformly spaced members of sounds in order for a listener to be able to report an irregularity.
2) to know how accurately one can produce a regular temporal sequence.
3) to explore an identical timing-related threshold which would be common both in perception and production of rhythm.

First, as concerns 1), it was found that the degree of distortion which afforded fifty percent judgments of detection varied dependently upon the rates of succession of the sequence. Weber's law was, however, found to hold approximately for discrimination of irregularity in sequences in the regions of the lower and higher rates of succession. In Experiment A, the mean value of DLs among the lower rates of succession, i.e. 1.0 to 2.5 times per second, was 6.1 percent. The mean values were 8.9 percent and 8.2 percent for the "lengthened" and the "shortened," respectively, in the regions of the higher rates of succession (i.e. 4.0 to 7.0 times per second). In Experiment B, the mean values were 6.0 percent and 7.6 percent in the regions of the lower and higher rates of succession, respectively. We may, at this stage, assume that there are two types of processing mechanisms which govern the perception of the temporal sequence.

Second, as concerns 2), it was found in Experiment D that it is possible to reproduce the regular temporal sequence synchronously with the stimulus signals. The variabilities of the reproduced time intervals were found to be 3 to 9 percent of their mean intervals, almost independent of the rate of succession of the stimulus sequence. The autocorrelation study showed that an adjusting mechanism worked only in the region of the lower rates of succession (i.e. 1.0 to 3.0 times per second). Here again, we may assume that there are two types of processing mechanism which govern the perception of the temporal sequence; one allows the adjustment mechanism to work in reproduction of the sequence and the other does not.

Third, as concerns 3), it was found in Experiment C that a time interval of about 330 milliseconds was a threshold which distinguished one region from another.

From the results of the present study, we may tentatively conclude that the onging processing mechanism works in the region of rates slower than 3 times per second and the wholistic processing mechanism works, on the contrary, in the region of rates more rapid than 3 times per second.

The present study has revealed some characteristics of man's ability in perception and in reproduction of the temporal sequence which have not been explicitly shown so far.

Acknowledgments

The author wishes to express his sincerest gratefulness to Professor Masayuki Sawashima, without whose assistance, advice and patient supervision this work would never have been accomplished.

I am also deeply indebted to Professor Shuzo Saito and Dr. Hajime Hirose for their helpful comments and constructive criticism. Thanks are also due to all other members of the Research Institute of Logopedics and Phoniatrics for their help and advice throughout the course of this study, and to my subjects for their willingness to participate in the experiments.

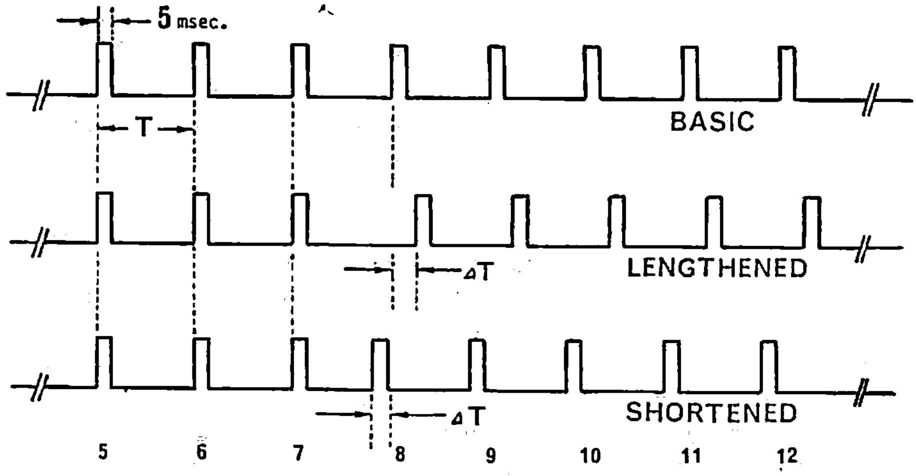

Fig. 1-A *Examples of distorted interval in Experiment A. The length of the time interval bounded by the 7th and 8th tone bursts is either lengthened (middle) or shortened (bottom) by* Δ*T.*

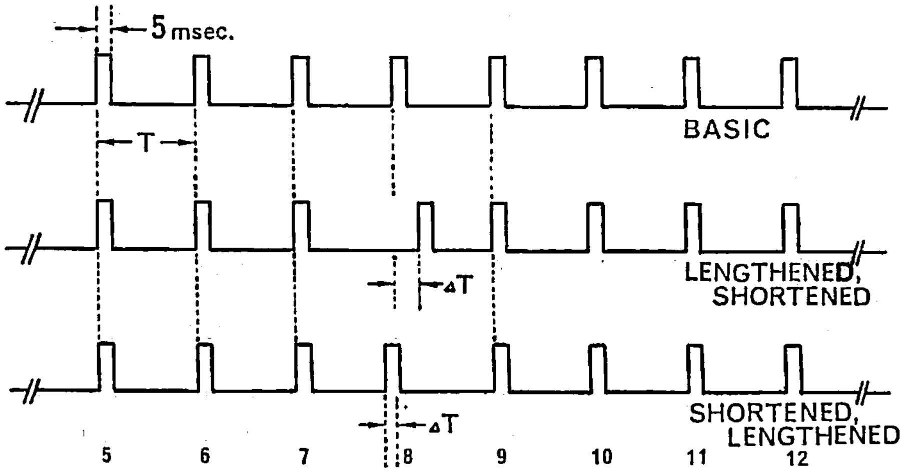

Fig. 1-B *Examples of distorted temporal allocation in Experiment B. The temporal allocation of the 8th tone burst is moved either backward (middle) or forward (bottom) by* Δ*T.*

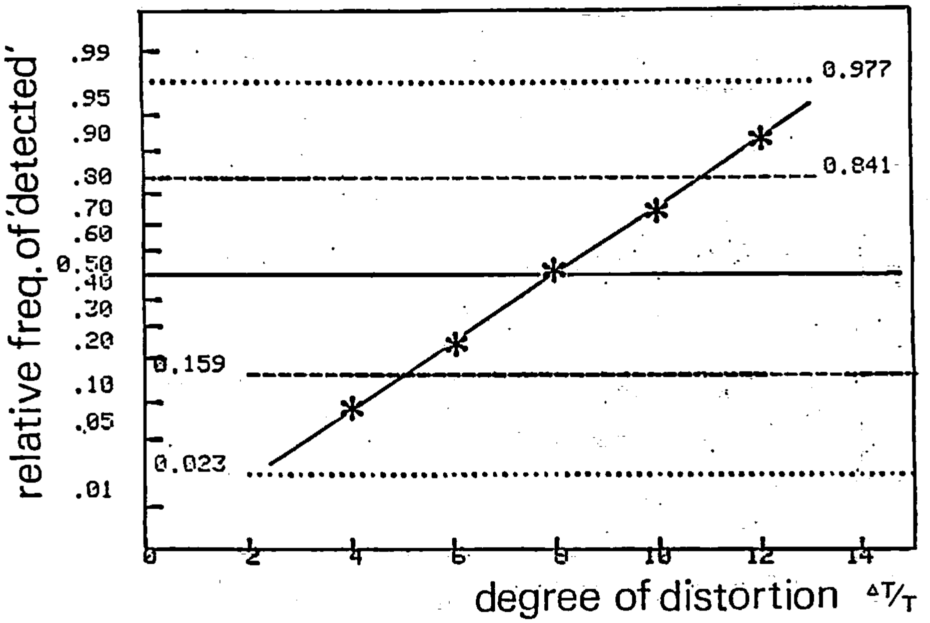

Fig. 2 *Probability that distortion will be detected as a function of the degree of distortion. The ordinate represents relative frequency of judgments "detected," while the abscissa shows the degree of distortion. One example of the results is presented here for "shortened" sequences of "6 times per second."*

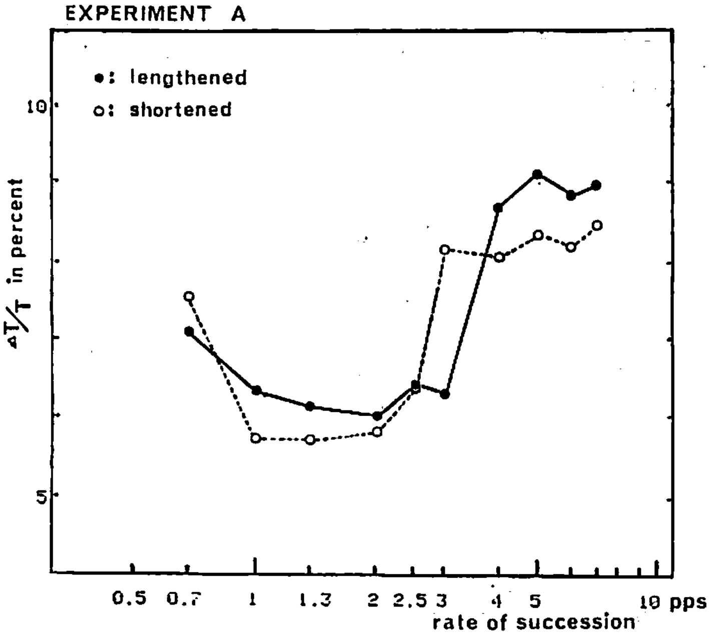

Fig. 3-A *Degree of distortion which afforded fifty percent judgment (DL) as a function of the rate of succession. The ordinate contains DL values, while the abscissa shows the rate of succession.*

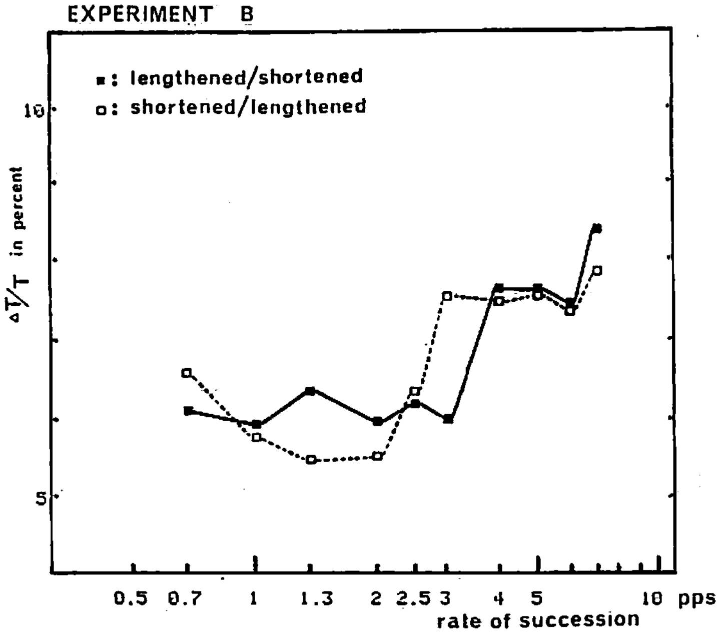

Fig. 3-B *Degree of distortion which afforded fifty percent judgment (DL) as a function of the rate of succession. The ordinate contains DL values, while the abscissa shows the rate of succession.*

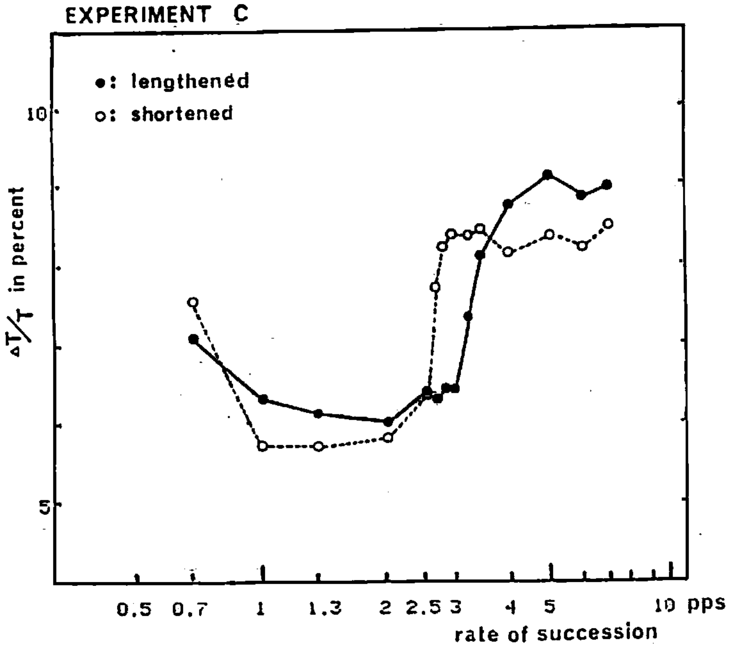

Fig. 3-C *Degree of distortion which afforded fifty percent judgment (DL) as a function of the rate of succession. The ordinate contains DL values, while the abscissa shows the rate of succession.*

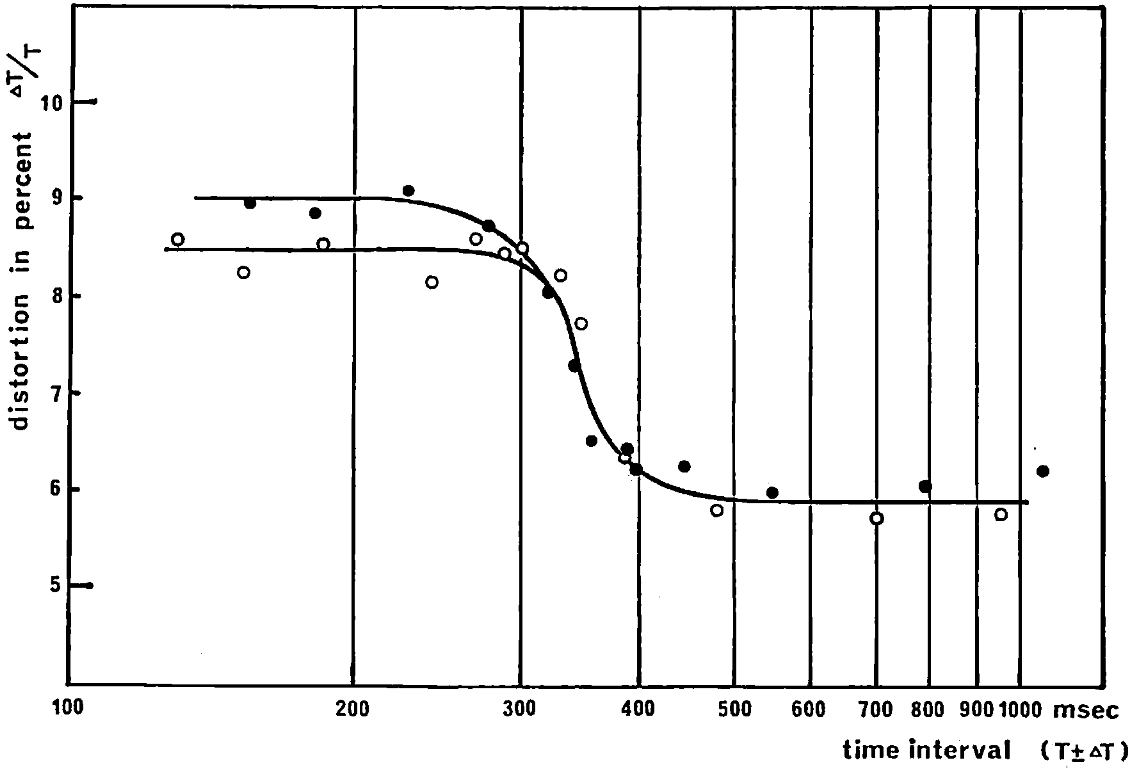

Fig. 3-C′ *Degree of distortion which afforded fifty percent judgment (DL) as a function of the length of the distorted interval. The ordinate contains DL values, while the abscissa shows the length of the distorted interval.*

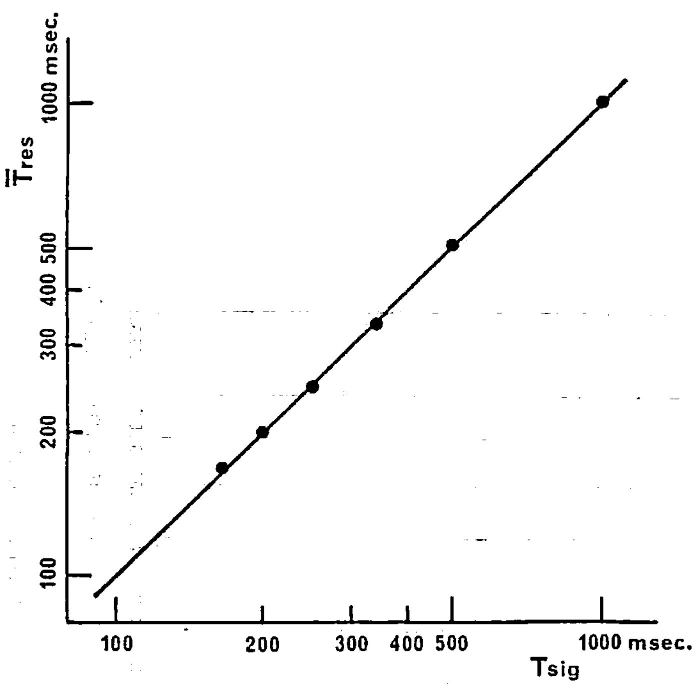

Fig. 4 *The relation between the mean time interval of the reproduction of the sequence as response in time with the stimulus sequence ($\overline{T}_{res}$) and the time interval of the stimulus signals (T_{sig}). The mean values were calculated from among all of 9 subjects.*

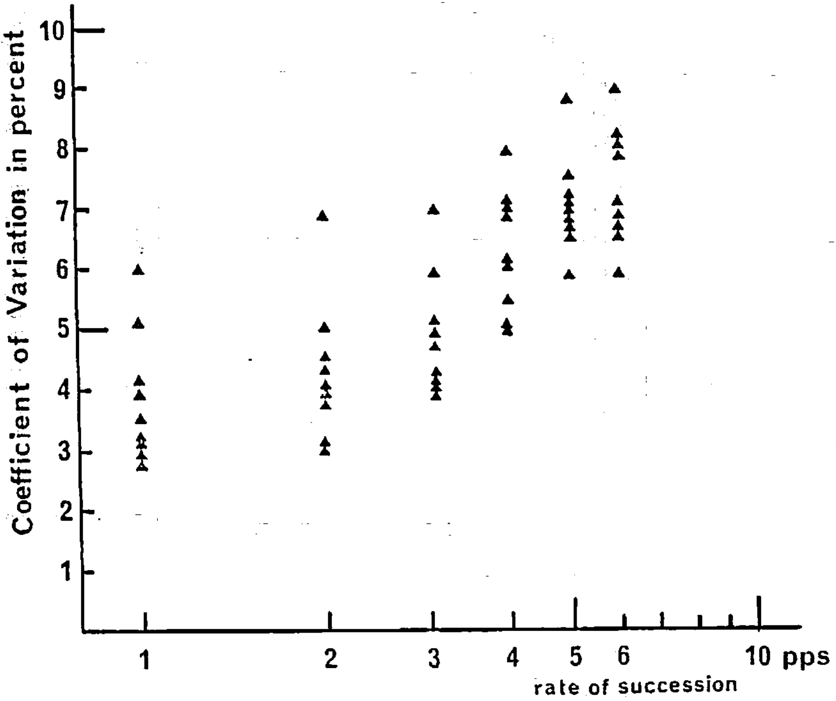

Fig. 5 *The relation between the coefficient of variance and the rate of succession of the stimulus signals. The results of 9 subjects were shown by triangles.*

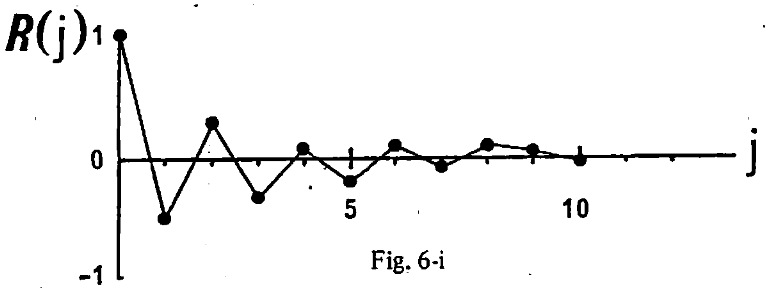

Fig. 6-i

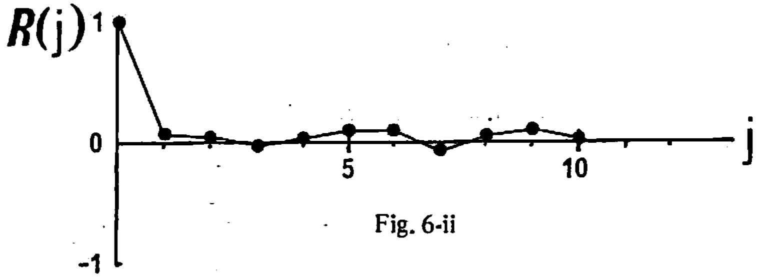

Fig. 6-ii

Fig. 6-i *and* Fig. 6-ii *show the autocorrelation function R(j) as a function of j, where j is the distance between two time intervals whose correlation is shown by R(j). R(j) values are for the reproduction of the time sequence at rates of 2 times per second (Fig. 6-i) and 5 times per second (Fig. 6-ii). Both R(j) values were obtained from the data of subject A.*

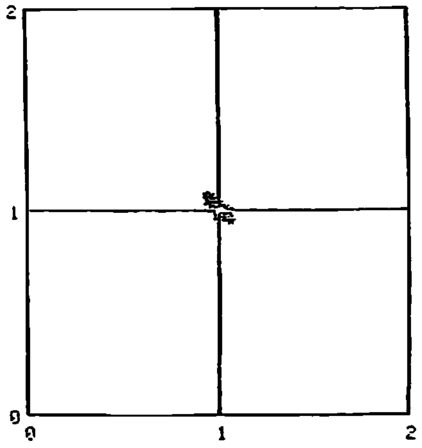

Fig. 7-i *Joint histogram for the reproduction of the time sequence at a rate of 2 times per second. (Subject A)*

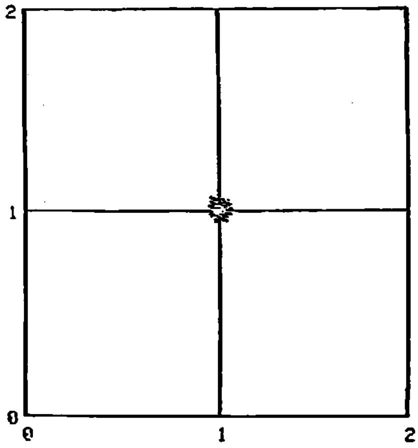

Fig. 7-ii *Joint histogram for the reproduction of the time sequence at a rate of 5 times per second. (Subject A)*

Table 1-A *Analysis of variance*

Experiment A

Factor	*Sum of Square*	*D.F.*	*Variance*
Rate of succession	80.54	9	8.95**
Serial position	0.35	2	0.18
Condition (L vs. S)	0.13	1	0.13
R x S	3.40	18	0.19
S x C	0.22	2	0.11
C x R	8.42	9	0.94**
R x S x C	2.04	18	0.11

**: $p < 0.01$ L: "lengthened", S: "shortened"

Table 1-B *Analysis of variance*

Experiment B

Factor	*Sum of Square*	*D.F.*	*Variance*
Rate of succession	38.43	9	4.27**
Serial position	0.65	2	0.33
Condition (L/S vs. S/L)	–	1	–
R x S	1.70	18	0.09
S x C	0.52	2	0.26
C x R	6.73	9	0.75**
R x S x C	2.10	18	0.12

**: $p < 0.01$ L/S: "lengthened and shortened", S/L: "shortened and lengthened"

Rate Subject	1	2	3	4	5	6
A	●	●	●	●		
B	●	●	●			
C		●	●			
D	●	●	●	●		
E	●	●	●			
F	●	●	●			
G	●	●	●			
H		●	●			
I	●	●	●			

Rate: Rate of succession

Table 2 *Results of inspection by means of joint histogram and autocorrelation studies, where each circle represents that the adjustment mechanism seemed to work.*

	Rate of succession	4	5	6	7	PPS
Results of EXPERIMENT B	L/S	7.66	7.63	7.45	8.39	%
	S/L	7.47	7.57	7.38	7.85	%
	Mean	7.57	7.60	7.42	8.12	%
Probability calculated from the results of EXPERIMENT A when distorted by the degree of the above mean value	P_1	0.27	0.28	0.13	0.28	
	P_2	0.40	0.29	0.33	0.35	
	$(1-P_1)\times(1-P_2)$	0.44	0.51	0.58	0.47	
	$1-(1-P_1)\times(1-P_2)$	0.56	0.49	0.42	0.53	

L/S: lengthened/shortened,
S/L: shortened/lengthened,
P_1: Probability of "detected" in EXPERIMENT A for "lengthened" sequence,
P_2: Probability of "detected" in EXPERIMENT A for "shortened" sequence.

Table 3 *The calculated probabilities that the distortion was detected at the lengthened interval and/or at the shortened interval from the obtained in Experiment A, when the degree of distortion which afforded fifty percent judgment in Experiment B was placed in the sequence.*

Bibliography

Abe, I. (1967): English sentence rhythm and synchronism. Bulletin of the Phonetic Society of Japan, 125, 9-11.

Allen, G.D. (1972): The location of rhythmic stress beats in English speech. Parts I and II. Language and Speech 15, 72-100, 179-195.

Allen, G.D. (1975): Speech Rhythm: its relation to performance universals and articulatory timing. Journal of Phonetics, 3, 75-86.

Chomsky, N. and Halle, M. (1968): The Sound Pattern of English. New York: Harper and Row.

Fraisse, P. (1963): The Psychology of Time, New York: Harper and Row.

Halle, M. and Stevens, K.S. (1964): Speech recognition: A model and a program for research. In: J.A. Fordor J.J. Kats (Eds), The Structure of Language: Reading in the Philosophy of Language. Englewood Cliffs, N.J.: Prentice Hall.

Hibi, S. (1980): On the rhythm pattern of repetitive utterances. Ann. Bull. RILP. 14:85-90.

Hirose, H., Kiritani, S., Ushijima, T., Yoshioka, H. and Sawashima, M. (1978): Patterns of dysarthric movements in patients with Parkinsonism. Ann. Bull. RILP. 12:73-81.

Hirose, H. and Sawashima, M. (1979): Disturbance of the rhythmic structure of speech in the dysarthrias. Ann. Bull. RILP. 13:113-121.

Hirose, H., Kiritani, S. and Sawashima, M. (1980): Patterns of dysarthric movements in patients with ALS and PBP. Ann. Bull. RILP. 14:263-272.

Klatt, D.H. (1976): Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. J.A.S.A. 59, 1208-1221.

Lehiste, I. (1973): Rhythmic units and syntactic units in production and perception. J.A.S.A. 54, 1228-1234.

Malécot, A., Johnston, R. and Kizziar, P.A. (1972): Syllabic rate and utterance length in French. Phonetica 26, 235-251.

Martin, J.G. (1972): Rhythmic (hierarchical) versus serial structure in speech and other behavior. Psychological Review Vol. 79. No. 6, 487-507.

Michon, J. (1966): Tapping regularity as a measure of perceptual motor load. Ergonomics Vol. 9.

Michon, J. (1967): Timing in temporal tracking. Inst. for Perception, Soesterberg, the Netherlands.

Norman, D.A. (1977): Human Information Processing. N.Y. Academic Press.

Norman, D.A. (1972): Memory and Attention. N.Y.: John Wiley.

Pike, K.L. (1945): The Intonation of American English. Ann Arbor: University of Michigan Press.

Sato, H. (1977): Segmental Duration and Timing Location in Speech. Reports of the 1977 Autumn Meeting of the Acoust. Soc. Japan. Transactions of Committee of Speech Research. Acoust. Soc. Japan No. S77-31.

Waugh, N.C. and Norman, D.A. (1965): Primary Memory. Psychological Review 72, 89-104.

Woodrow, H. (1951): Time Perception. Handbook of Experimental Psychology (S.S. Stevens Ed.) 1224-1236. New York: Wiley.