# AUDITORY PERCEPTION OF DURATION OF SPEECH AND NON-SPEECH STIMULI*

H. Fujisaki, K. Nakamura** and T. Imoto**

## I. Introduction

Temporal features of speech, such as duration and rate of change, play no less important roles in the transmission of information than spectral features. In fact, the rate of change of spectral features is crucial in all languages for the transmission of information concerning certain classes of phonemes, while the segmental duration in some languages primarily carries prosodic and emotional information,[1] and its inter-segmental variations, if not insignificant, are almost always accompanied by more reliable and distinctive spectral cues. There exist, however, languages in which the segmental duration serves as the primary cue for the distinction between certain classes of phonemes. In this respect, the sound pattern of Japanese presents examples of considerable interest since all the vowels and some consonants possess 'longer' counterparts that can be discriminated primarily by their durations. Though the results of an extensive study[2] have been published on the measurement of durations of these segments, the perceptual role of the durational cue still remains to be investigated.

The present paper deals with an experimental investigation into the roles played by segmental durations in the perception of Japanese vowels and consonants in various contexts, and their relationships to the perception of duration of various non-speech sounds which possess acoustic features similar to the speech sounds under study. The investigation has been performed both on subjects with normal hearing and on hard-of-hearing children.

---

* Paper to be presented at Symposium on Auditory Analysis and Perception of Speech, Leningrad, August 1973.

** Department of Electrical Engineering, Faculty of Engineering, University of Tokyo.

## II. Segmental Durations in Spoken Japanese

The sound pattern of Japanese is characterized by the existence of distinctions in segmental durations of certain classes of speech sounds, namely vowels, nasal consonants and some voiceless consonants. Each of the five vowels has a longer counterpart which is phonetically identical except for the duration, so that the contrast in their duration is phonemic, as illustrated in the following examples:

| | | | |
|---|---|---|---|
| [haku] | 'foil' | [ha:ku] | 'grasp' |
| [in] | 'rhyme' | [i:n] | 'clinic' |
| [kucu] | 'shoe' | [kucu:] | 'pain' |
| [keši] | 'poppy' | [ke:ši] | 'contempt' |
| [oi] | 'nephew' | [o:i] | 'cover' |

Similarly, with the exception of [h], each of the voiceless consonants in intervocalic position contrasts in duration with its longer counterpart, commonly called a 'geminate' consonant, and is often denoted by the juxtaposition of two identical phonetic symbols. In voiceless fricatives, it is the duration of the quasi-stationary frication that constitutes the phonemic contrast, as in the following examples:

| | | | |
|---|---|---|---|
| [ise] | 'a place-name' | [isse] | 'a unit of area' |
| [išoku] | 'transplantation' | [iššoku] | 'one color' |

In voiceless plosives and affricates, on the other hand, the phonemic contrast is realized by a difference in duration of the stop gap preceding the plosion, as in the following examples:

| | | | |
|---|---|---|---|
| [supai] | 'spy' | [suppai] | 'sour' |
| [ita] | 'existed' | [itta] | 'went' |
| [iki] | 'breath' | [ikki] | 'riot' |
| [icu:] | 'stomach-ache' | [iccu:] | 'one (letter)' |
| [iči] | 'one' | [ičči] | 'agreement' |

Finally, nasal consonants in intervocalic position may be converted into their longer counterparts by an increase in duration of the nasal murmur, as in the following examples:

| | | | |
|---|---|---|---|
| [ama] | 'nun' | [amma] | 'massage' |
| [ana] | 'hole' | [anna] | 'such' |
| [iŋa] | 'burr' | [iŋŋa] | 'negative picture' |

From one phonemic point of view,[3] a long vowel may be regarded as a vowel phoneme followed by a syllable-final phoneme, /H/, which does not possess an inherent phonetic value and is assimilated to its immediate predecessor, while a geminate consonant may be regarded as a consonant phoneme preceded by another phoneme, /Q/, which also does not possess an inherent phonetic value but is assimilated to its immediate successor. A long nasal consonant may also be regarded as a nasal consonant preceded by another syllable-final nasal phoneme /N/, whose phonetic value is determined by its immediate successor, but unlike /Q/, may occur anywhere except in the word-initial position. Thus the phonemes /H/ and /N/ are actualized by an increase in the duration of a quasi-periodic vowel and a nasal murmur, respectively, while the phoneme /Q/ is actualized either as an increase in the duration of a quasi-stationary aperiodic frication or of a silent interval preceding plosion.

In spite of these differences in their roles in the phonemic system as well as in their spectral features, however, these phonemes share a peculiar feature in that, unlike other phonemes, they occupy approximately one basic unit of duration of Japanese, a 'mora.' A mora usually consists of either a consonant-vowel pair or a vowel. Thus the duration of a long vowel or a geminate consonant has been observed to be more than twice as large as its shorter counterpart.[2] Our own measurement, which is to be reported elsewhere, also confirms these observations.

Perceptually, these phonemes present instances of particular interest, since they involve a distinction of both vowels and consonants on a single continuum of temporal duration. They are also of interest because of varieties in the spectral features of the interval in question, namely periodic in vowels and nasal consonants, aperiodic in voiceless fricatives, and nil in voiceless plosives. It is the purpose of the present study to investigate the perception of segmental durations of these peculiar phonemes, and to compare their perception to that of acoustically similar non-speech sounds.

III. Discrimination of Durations of Non-speech Stimuli

Though the perception of segmental durations in speech is the main object of the present investigation, it is also important to know how this perception is related to that of non-speech sounds. Although considerable amounts of data have recently been published on temporal discrimination,[4-7] they cannot be adopted for the purpose of comparison because of discrepancies in the method as well as in the analysis of results. Consequently, experiments were designed to measure temporal discrimination of i) pure tone at 500 Hz, ii) white noise (within the frequency range of 50 - 3400 Hz), iii) filtered noise (with center frequency of 3000 Hz and bandwidth of 500 Hz), and iv) temporal gap between two tone bursts of 100 msec each. Stimuli i), ii) and iv) were adopted to serve for comparison with vowels, fricatives and plosives, respectively, while stimuli iii) were acoustically identical to the [s]-sound adopted in the perceptual experiments using synthetic speech to be described in the next section.

Both the tone and noise bursts had a linear rise and decay of 10 msec each to avoid clicks, and the total duration including rise and decay times was varied in stimuli i), ii) and iii) while the duration of the gap was varied in stimuli iv). The experiments were performed to measure discriminability at 100msec of each of the durations in question. In the case of 500 Hz tone, discriminability was also measured at durations of 50, 150, 200 and 300 msec. The measurement of discriminability was based on the two-alternative forced-choice paradigm, in which two stimuli were presented to a subject in temporal succession and the subject was required to select the one he believed to be longer in duration. In the present experiments, the interval between a pair of stimuli was kept at one second, while successive pairs were separated by four seconds for written response, and a brief tone of 1000 Hz was inserted at every 10 pairs.

For a measurement of discriminability at a duration T, the durations of the first and the second stimuli, denoted respectively by $Ts_1$ and $Ts_2$, were selected as follows:

$Ts_1 = T + n\cdot T,$ $\qquad$ $Ts_2 = T - n\cdot T,$

where $n = -3, -2, -1, 0, 1, 2, 3,$

thus resulting in seven different diads whose mean values were always kept equal to T. The value of T was determined for a particular value of T on the basis of a preliminary test such that the two extreme diads, corresponding to $n = \pm 3$, were almost always discriminated correctly. One set of test materials consisted of a randomized sequence of 140 diads containing 20 each of the seven diads, preceded and followed by five dummy diads. The synthesis of the stimulus waveforms as well as their compilation were performed on a digital computer, and the output was fed to a digital-to-analog converter at a rate of 8 kHz with an accuracy of 8 bit/sample, to be recorded on an analog tape recorder for off-line experiments.

The subjects were four adults with normal hearing, and sat for three test sessions for each value of T, so that the number of judgments for each diad was 60. Because of differences in individual performances, the test results of each subject require separate analysis.

The results of one discrimination test can be illustrated by Fig. 1, where the probability that the second stimulus $S_2$ is judged longer than the first stimulus $S_1$ is plotted on a normal scale against the difference in their durations, $Ts_2 - Ts_1$. If the comparative judgment is assumed to be disturbed by some internal random processes, the experimental data can best be approximated by a normal distribution. Within the range of values of T under study, the approximation holds quite well for all the individual data. The solid line in Fig. 1 indicates an approximation to the measured data on the basis of the least-mean-squared error weighted by Müller-Urban coefficients, characterized by its mean $\mu_D$ and the standard deviation $\sigma_D$ as listed in the figure.

The mean of the distribution indicates the time-order error, while the standard deviation $\sigma_D$ can be regarded as an index of the accuracy of comparative judgment necessary for discrimination. The conventional difference limen (DL), defined by the abscissa corresponding to 75% correct judgment, is equal to $0.675\ \sigma_D$ in the absence of the time-order error, but tends to vary with the absolute value of the time-order error. In view of the fact that the time-order error is not negligible in most of the test results, conventional DL as well as the index of signal detectability $d'$, both based on the analysis of correct/incorrect responses, are considered inappropriate for describing the accuracy of a subject's performance in the task of discrimination.

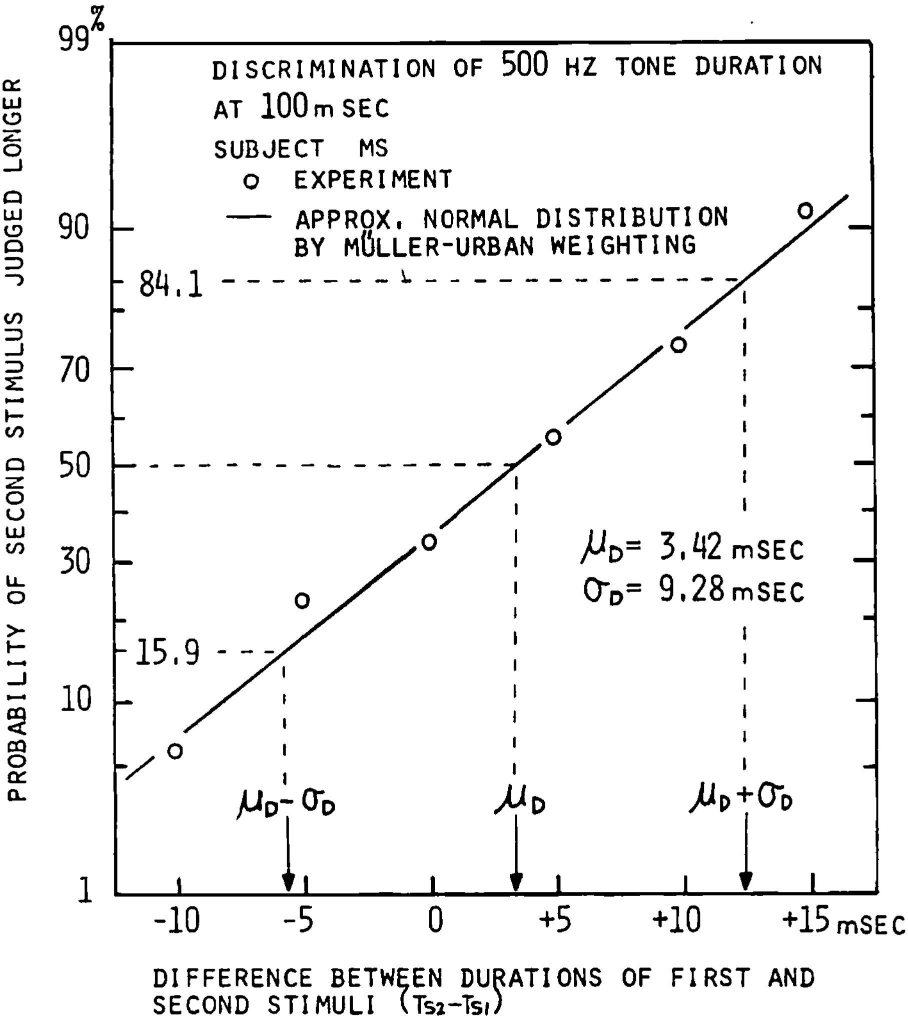


Fig. 1. An example of results of discrimination test for 500 Hz tone duration at 100 msec, and the approximate normal distribution by Müller-Urban weighting. The mean ($\mu_D$) indicates time-order error, while the standard deviation ($\sigma_D$) serves as an index of accuracy of discrimination.

Table 1 lists the mean value and the standard deviation of $\sigma_D$ 's for the four subjects. Analysis of variance indicates that the difference between the value of $\sigma_D$ for pauses (temporal gaps) and those for tone and noise bursts

Table 1. Accuracy of discrimination ($\sigma_D$) for duration of various non-speech stimuli. Mean and standard deviation of four subjects.

| Stimuli | 500 Hz tones | | | | | white noise | filtered noise | pause between tones |
|---|---|---|---|---|---|---|---|---|
| | 50 msec | 100 msec | 150 msec | 200 msec | 300 msec | 100 msec | 100 msec | 100 msec |
| Accuracy of discrimination | 7.6 msec | 9.6 | 11.6 | 14.5 | 23.1 | 9.1 | 6.7 | 21.5 |
| Standard deviation | 1.7 msec | 1.5 | 1.9 | 1.3 | 1.1 | 1.9 | 0.86 | 4.9 |

of the same duration is highly significant at the 1% level (t = 4.1). This is in agreement with the published results,[6, 7] and suggests that the neural processes for the representation and comparison of filled and empty intervals may not be identical.

The mean value of $\sigma_D$ 's of the four subjects for tone bursts is plotted against tone duration T in Fig. 2, and is seen to increase rather slowly with T for smaller values of T, but tends to increase almost linearly with T for larger values. The solid curve is a linear approximation to $\sigma_D^2$ versus $T^2$ with the least-mean-squared error criterion, and can be expressed by

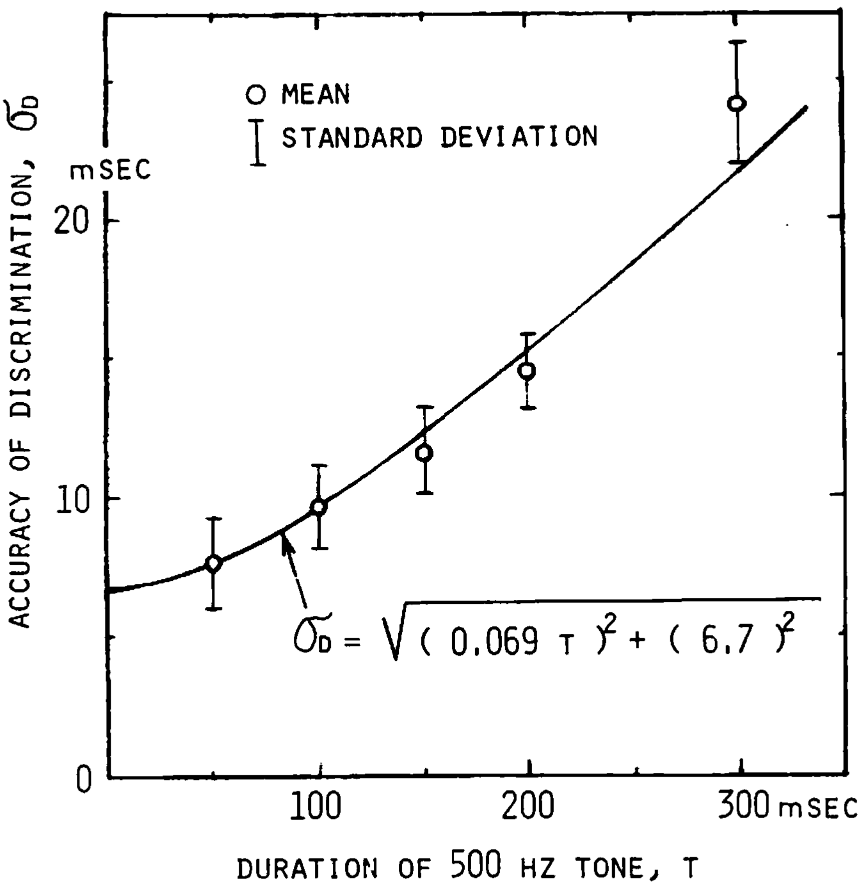


Fig. 2. Accuracy of discrimination in $\sigma_D$ for the duration of tone bursts of 500 Hz. Mean and standard deviation of four subjects. The solid curve is an approximation to the measured data.

$$\sigma_D^2 = (0.069T)^2 + (6.7)^2, \qquad (\text{msec})^2.$$

The approximation is seen to hold fairly well for the range of 50 - 300 msec of T, and suggests that the variability of discrimination within the observed range may be caused by the participation of two rather independent random processes, both being characterized by variances expressed by the first and second terms in the right hand side of the above equation. We may speculate that they correspond to noises in the memory process and the decision process, respectively. Physiological interpretation of these results, however, is beyond the scope of the present paper and will be discussed elsewhere.

## IV. Identification and Discrimination of Segmental Durations of Synthetic Speech

In accordance with the considerations given in section II, the following four stimulus groups were synthesized to investigate the perception of duration of various speech segments. Namely,

1) Group 1 (the vowel group), in which the duration of the vowel [o] (defined arbitrarily as the interval from the onset of the vowel [o] to the onset of the formant transition toward the following vowel) is varied to cover the range of duration from [oi] ('nephew') to [o:i] ('cover') found in natural utterances.
2) Group 2 (the fricative group), in which the total duration of the fricative noise interval is varied from [ise] ('a place-name') to [issɛ] ('a unit of area').
3) Group 3 (the plosive group), in which the duration of the alveolar stop (defined as the sum of a variable stop gap and a constant interval of 30 msec between plosion and the onset of the next vowel) is varied from [ita] ('existed') to [itta] ('went').
4) Group 4 (the nasal group), in which the duration of the nasal murmur is varied from [ama] ('nun') to [amma] ('massage').

These pairs of words were selected such that they should have the same type of pitch accent,[8] and should contain the segment under study in the minimal context, so that they could be discriminated solely by the durational cue.

Based on our spectrographic measurements of segmental durations in natural utterances, the durations of all the vowels adjacent to the variable segment were kept equal to 100 msec, including rise and decay times of 20 msec each, corresponding to a talking rate of five morae per second. The duration of the variable segment ranged from 80 to 240 msec in jumps of 10 msec, resulting in 17 different synthetic words. In the identification test, those synthetic words were presented either in isolation (identification in word context) or embedded in a short carrier sentence of synthetic speech with the same talking rate as the test word. One set of test materials consisted of a randomized sequence of 85 items containing 5 each of the 17 words or sentences, preceded and followed by five dummies. Successive items were separated by four seconds for response, and a brief tone of 1000 Hz was inserted at every 10 test items. The procedures in the discrimination test were the same as for the non-speech stimuli.

The stimuli were synthesized by digital computer simulation of a terminal-analog speech synthesizer with control of seven parameters: voice fundamental frequency ($F_0$), three formant frequencies ($F_1$, $F_2$, $F_3$, ), intensity of voice source (Av), center frequency of fricative filter (Fc), and intensity of noise source (Ac). The bandwidths of the formants and the fricative filter were varied with their frequencies, and the fourth formant frequency was fixed at 3600 Hz, while the sampling frequency was 8 kHz. Except for the duration of the segment under study, the parameters were controlled by segmental and prosodic rules which specified their values by piecewise-linear models. Compilation of test stimuli was also performed by a digital computer. A group of five subjects, of which four were the same as in the discrimination tests of non-speech stimuli, had six test sessions per each/set of test materials. Test results of these subjects were analyzed individually.

The result of an identification test is illustrated by Fig. 3, which shows the performance of a subject in the identification of a geminate fricative consonant [ss] in the word context [ise]-[isse]. The probability of identification of the geminate fricative is plotted on a normal scale against the duration of the fricative noise interval. As in the analysis of results of discrimination tests, the results conform quite well to the approximation by a normal distribution calculated with Müller-Urban weighting. In this case, however, the

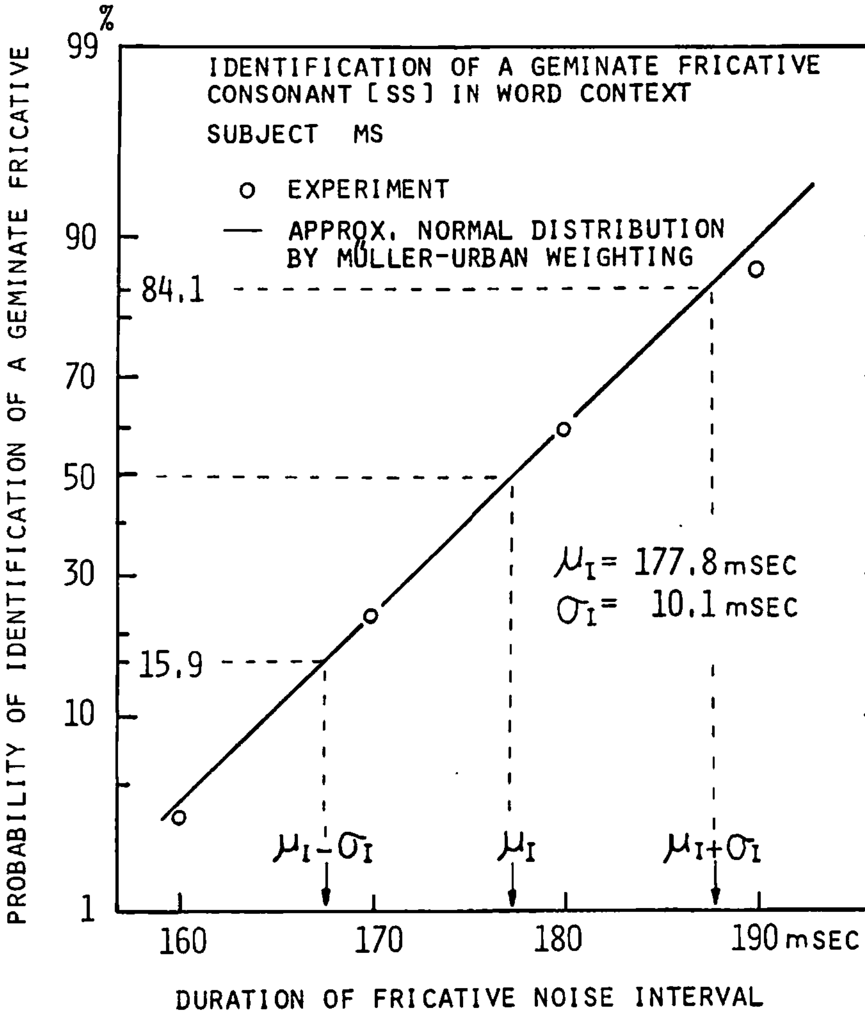


Fig. 3. An example of results of identification test for a geminate fricative consonant in word context [ise]- [isse], and the approximate normal distribution by Müller-Urban weighting. The mean ( $\mu_I$) indicates the phoneme boundary, while the standard deviation ( $\sigma_I$) serves as an index of accuracy of identification.

mean $\mu_I$ indicates the phoneme boundary, and the standard deviation $\sigma_I$ can be regarded as an index of the accuracy of categorical judgment necessary for identification. Table 2 lists the phoneme boundaries and their standard deviations obtained from the analyses of individual data and then averaged over all the subjects. The results indicate that both the phoneme boundary $\mu_I$ and the accuracy of identification $\sigma_I$ remain quite stable regardless of the spectral features of the segments as well as their contexts. It may also be noted that the accuracy is somewhat improved as the context is extended from word to sentence. The difference, however, is only significant for the fricative group at the 5% level.

In comparison with the results of discrimination tests of acoustically similar non-speech stimuli of Table 1, it should be noted that the values of $\sigma_I$ for the vowel and the nasal groups are very nearly equal to $\sigma_D$ for tone bursts of 100 or 150 msec, while $\sigma_I$ for the plosive group is significantly smaller than $\sigma_D$ for pauses between two tone bursts and is essentially the same as $\sigma_I$ 's for the vowel and the nasal groups. This leads one to suspect that the same mechanism is utilized for the identification of these speech segments in word context, while the mechanisms for the discrimination of durations of filled and empty time intervals, which certainly requires short-term memory for these acoustically different stimuli, may not be identical.

The fricative group, on the other hand, presents an interesting exception in this regard, since the accuracy of identification $\sigma_I$ in word context is found to be inferior to the accuracy of discrimination $\sigma_D$ when the same fricative noise is presented in isolation as non-speech stimulus. For the sake of further comparison, a discrimination test was also performed on the fricative interval of 100 msec in word context, and the mean value of $\sigma_D$'s for the five subjects was 17 msec, being almost equal to $\sigma_I$ in word context, but significantly larger than $\sigma_D$ for the same noise in isolation. In view of the differences in the mechanisms and processes involved in identification and discrimination of speech sounds,[9] however, there seems to be no specific reason that accuracies in these two tasks should be equal, and the results of our previous investigation on the perception of vowels[10] also indicate that $\sigma_D$ is generally greater than $\sigma_I$ both in ABX and AX tests. It is not possible to decide whether the parallelism between $\sigma_D$ and $\sigma_I$ of speech stimuli observed

Table 2. Phoneme boundaries and accuracy of identification for various synthetic speech stimuli in word and sentence context at a mean mora duration of 200 msec. Mean of five subjects.

| Stimuli | vowel /oi/-/ooi/ | | fricative /ise/-/isse/ | | plosive /ita/-/itta/ | | nasal /ama/-/amma/ | |
|---|---|---|---|---|---|---|---|---|
| Context | word | sentence | word | sentence | word | sentence | word | sentence |
| Phoneme boundary | 156 msec | 168 | 166 | 165 | 169 | 164 | 141 | 152 |
| Accuracy of identification | 9.5 msec | 7.1 | 16 | 10 | 11 | 8.9 | 10 | 8.5 |

in these experiments is a result of fortuitous selection of experimental conditions, or is an intrinsic characteristic of the temporal perception, and further investigation is in progress for the elucidation of this problem.

Though our recent investigation on identification and discrimination of speech sounds provided both experimental evidence and a theoretical model for the so-called categorical phenomenon in discrimination of speech,[9, 10] the phenomenon has not been demonstrated in classes of sounds where the phonemic distinction depends solely on the durational cue. In order to test the validity of our theory for these classes of sounds, discrimination tests were performed both for the duration of the vowel [o] in Group 1 ( /oi/ - /ooi/ ) and the duration of the fricative consonant [s] in Group 2 ( /ise/ - /isse/ ). The durations of the respective segments in the test stimuli were selected following a monotone relationship to yield approximately equal discriminability at both extremes of the stimulus continuum, as shown in Fig. 4. The measurement of discriminability followed the same experimental procedure as adopted in discrimination tests of non-speech stimuli, and typical results are also shown in Fig. 4. The peaks of the two discrimination curves roughly correspond to locations of phoneme boundaries of the subject for the respective stimulus groups, and clearly indicate that the discrimination performance is influenced by categorical judgments to almost the same extent both in vowels and in fricative consonants, when the stimuli are varied on the continuum of duration.

Although the results thus far indicated the stability of phoneme boundaries $\mu_1$, i. e. the perceptual criteria for categorization of segmental durations, regardless of phoneme class and context, it is expected that they are essentially relative and are influenced by the segmental durations or talking rates of their context. In order to obtain quantitative estimates for the magnitude as well as the extent of such influences, further identification tests were performed using synthetic speech stimuli of the fricative group both in word and in sentence context, but at talking rates increased by 25% and reduced by 20% from the original rate of 5 mora/sec, corresponding to mora durations of 160 msec and 250 msec, respectively. The stimuli were exactly the same as for the original rate except that the time scale for the control parameters was varied by the fractions indicated above. In sentence-context experiments, the carrier sentence: [Sorewa _____ desu] ('It is ____. ') was used, so that the test word was preceded by 3 morae and followed by 2 morae, while in

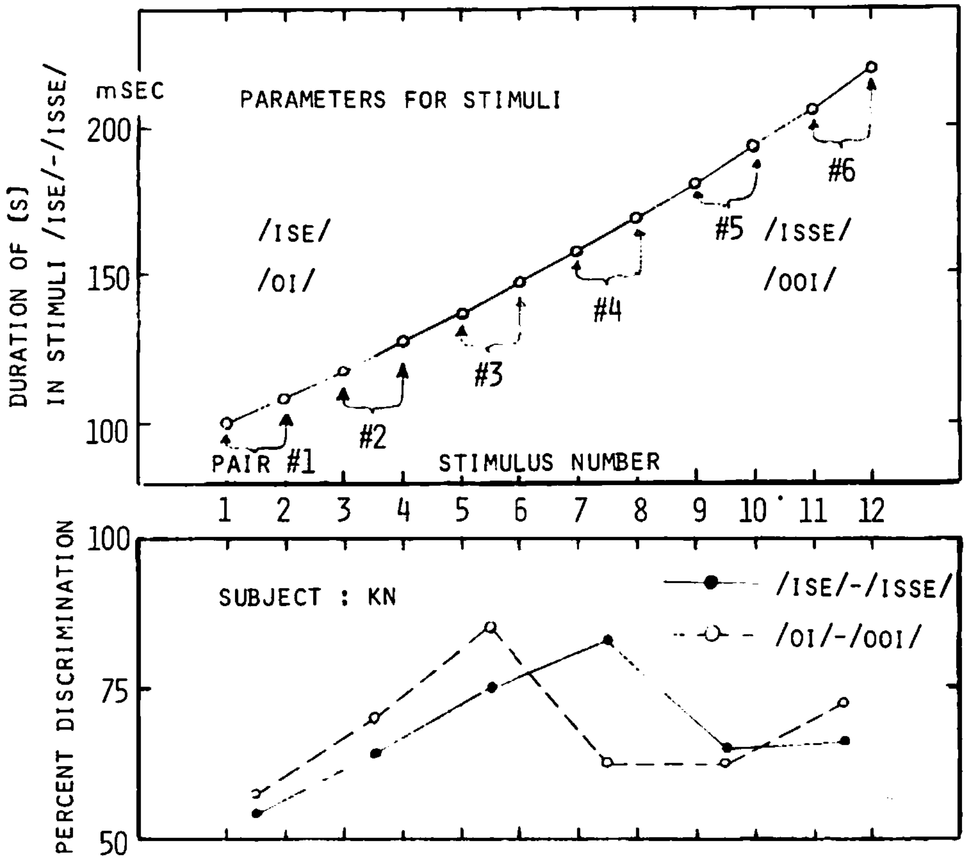


Fig. 4. Discrimination of speech stimuli within and across phoneme boundaries of the vowel [o] in /oi/ - /ooi/ and of the fricative consonant [s] in /ise/ - /isse/.

word-context experiments, the context for the fricative segment was minimal, being composed of only one vowel on each side. The mean phoneme boundaries of the five subjects, obtained both in word- and sentence-context experiments, are plotted against the duration Tm of one mora in Fig. 5, and the lines indicate the following linear approximations:

$\mu_{IW}$ = 0.70 Tm + 25 (msec), in word context,

$\mu_{IS}$ = 0.85 Tm (msec), in sentence context.

These results indicate the short-term adaptability of the temporal criterion in quantitative terms, and show that a context of about one second is almost sufficient for complete adaptation.

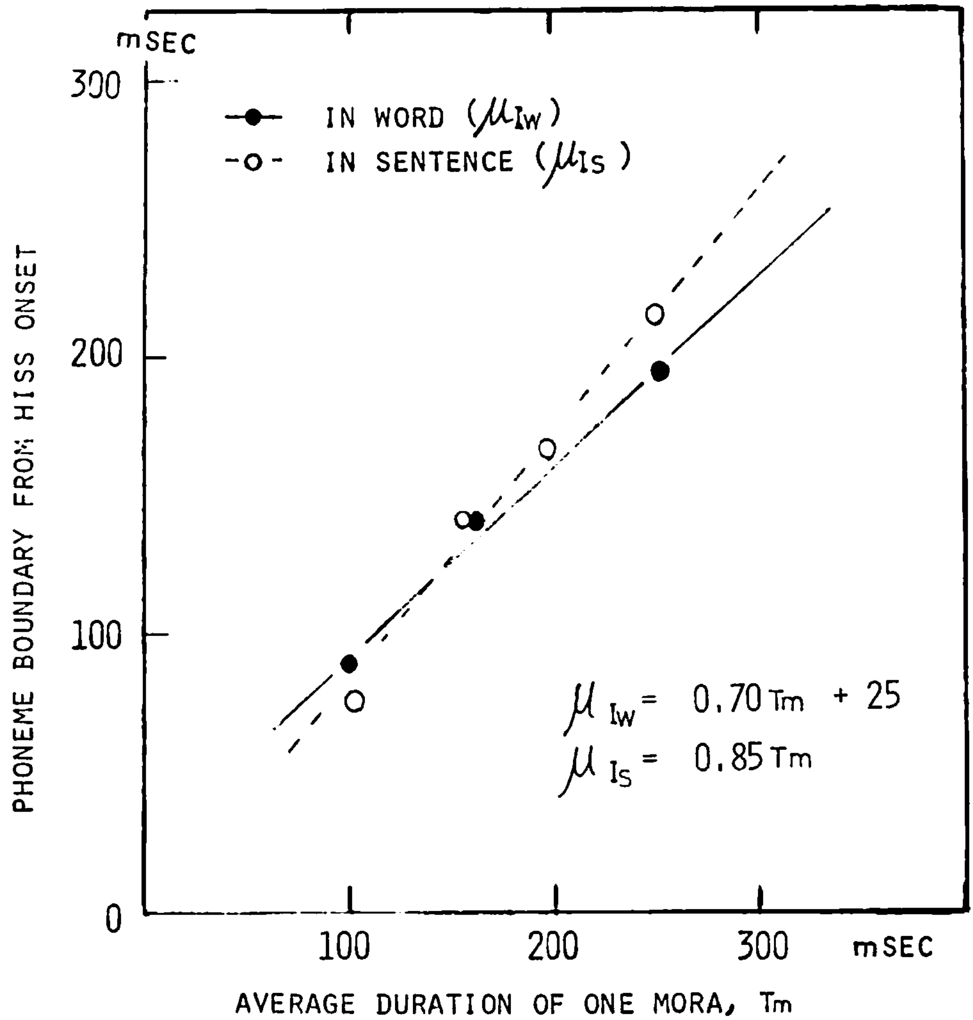


Fig. 5. The influence of talking rate on the identification of a geminate fricative consonant in word and sentence context.

## V. Experiments on Hard-of-Hearing Children

Quantitative estimation for the impairments of comprehension of speech by hard-of-hearing children and detection of the causes of these impairments are indispensable steps toward the alleviation and remedy of these impairments through proper training. For this purpose, a preliminary speech audiometry was performed on hard-of-hearing children using natural utterances of 2- and 3-mora words, and revealed specific impairments in the perception of speech sounds that were discriminated by the durational cue,[11] which could not be detected by conventional speech audiometry using only monosyllables.

In order to obtain more precise estimates of the degree of impairments in the perception of these speech sounds, and to locate their causes, discrimination and identification tests of duration were performed using synthetic speech and non-speech materials similar to those used for normal hearing subjects, but somewhat reduced in scale. The subjects were four children of 11 and 12 years of age with severe sensory-neural hearing impairments, all of them being educated in a special class for the hard-of-hearing. Three normal hearing children of the same age were also tested as the control. The stimuli were presented monaurally through a headphone to the better ear at the most comfortable level for each subject.

While normal hearing children can be considered as samples from a single population as far as the perception of the present test stimuli is concerned, such is not the case for hard-of-hearing children because of diversities both in types and degrees of their hearing impairments and in their previous educational history. Consequently, performances of hard-of-hearing children were individually analyzed and tested for significance of difference from the corresponding performances of the normal group. The results of these tests for significance are summarized in Table 3 along with descriptions of individual hearing loss and previous educational history. These results indicate that, in spite of their severe hearing impairments as measured by pure tone audiometry, none of the hard-of-hearing children differ significantly from the normal group in their perceptual ability of duration of non-speech stimuli when their hearing loss is compensated by proper amplification, but their performances in perception of speech stimuli show extensive individual variations,

Table 3. Hearing loss of individual cases of hard-of-hearing children and comparison of their accuracy of discrimination and identification with normal hearing children.

D : Discrimination.
Iw : Identification in word context.
Is : Identification in sentence context.

| Case | Age | Hearing loss at | | | Type of loss | Non-speech | | /ita/ - /itta/ | | | oi - ooi | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 500Hz | 1000HZ | 2000Hz | | tone | noise | D | Iw | Is | Iw | Is |
| MJ | 12 | 35dB | 60dB | 80dB | high tone | - | - | - | - | - | - | - |
| IK | 12 | 75 | 75 | 70 | flat | - | - | + | - | + | + | + |
| HT | 11 | 70 | 85 | 70 | dip | - | - | + | + | + | + | + |
| TC | 12 | 40 | 45 | 60 | flat | - | - | / | + | + | / | / |

Previous educational history
MJ : ordinary school since 6 years old.
IK : school for the deaf from 4 till 8.
HT : school for the deaf from 2 till 6.
TC : school for the deaf from 6 till 8 (also in dormitory).

Difference from normal group
+ : significant at 1 % level.
- : not significant at 5 % level.

ranging from no significant difference from the normal group at one extreme to highly significant differences in all the tests at the other. These individual differences are found to be more strongly correlated with past history of education and training than with types and degrees of hearing loss. Case TC provides a typical example of severely impaired speech perception due to retarded training. These results testify the importance of appropriate education and training at earlier ages.

## VI. Summary and Conclusions

Synthetic speech and non-speech stimuli were used to investigate perception of temporal duration which plays a distinctive role in the sound pattern of Japanese. For the quantification of results of discrimination and identification tests, methods were described to obtain indices of accuracy $\sigma$, which proved to be more appropriate than the conventional difference limen based on analysis of correct/incorrect responses. Accuracies of discrimination of filled and empty non-speech intervals in normal hearing adult subjects were found to be significantly different, suggesting a difference in the underlying mechanisms. The dependency of $\sigma$ on the duration of tone burst stimuli showed discrepancies from published data, leading to a new interpretation of the mechanisms for temporal discrimination.

Identification tests of elongated speech segments both in word and in sentence context, on the other hand, showed marked uniformity of phoneme boundaries as well as accuracies of identification for vowels, nasals, and voiceless plosives, suggesting that the durational cues of these stimuli are processed by the same mechanism in spite of the differences in their acoustical characteristics. The influence of the timing rate of the context upon the phoneme boundaries was measured both in word and in sentence contexts, indicating the short-term adaptability of the decision criteria in quantitative terms.

Similar tests were also performed on hard-of-hearing children with normal hearing children as the control, and the results from the hard-of-hearing children were analyzed and tested individually for significance of difference from the normal group. While none of them were found to be significantly different in discriminating durations of non-speech stimuli when their hearing loss was compensated by proper amplification, their performances for speech stimuli showed extensive individual differences, depending more heavily on previous history of education than on types and degrees of hearing loss, and testifying to the importance of proper training at earlier ages.

## References


1. Fry, D. B.: "Duration and Intensity as Physical Correlates of Linguistic Stress," J. Acoust. Soc. Am., 27, 765-768 (1955).
2. Han, M. S.: "The Feature of Duration in Japanese," Study of Sound (J. Phonetic Soc. Japan) 10, 65-80 (1965).
3. Fujimura, O.: "Fundamentals of Speech Science," in J. Oizumi and O. Fujimura, eds.: Speech Science, 3-91, Tokyo University Press (1972) (in Japanese).
4. Chistovich, L. A.: "Discrimination of the Time Interval between Two Short Acoustic Pulses," Soviet Physics - Acoustics, 5, 480-484 (1959).
5. Creelman, C. D.: "Human Discrimination of Auditory Duration," J. Acoust. Soc. Am., 34, 582-593 (1962).
6. Abel, S. M.: "Discrimination of Temporal Gaps," J. Acoust. Soc. Am., 52, 519-524 (1972).
7. Abel, S. M.: "Duration Discrimination of Noise and Tone Bursts," J. Acoust. Soc. Am., 51, 1219-1223 (1972).
8. Fujisaki, H. and Sudo, H.: "Synthesis by Rule of Prosodic Features of Connected Japanese," Proceedings of the 7th I. C. A., 3, 133-136 (1971).
9. Fujisaki, H. and Kawashima, T.: "A Model of the Mechanisms for Speech Perception - Quantitative Analysis of Categorical Effects in Discrimination," Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 30, 59-68 (1971).
10. Fujisaki, H. and Kawashima, T.: "A Quantitative Model for the Mechanisms of Discriminating Speech Sounds," Proceedings of the 7th I. C. A., 3, 433-436 (1971).
11. Nakamura, K.: "Perception of Geminate Consonants in Hard-of-Hearing Children (in Japanese), unpublished M. A. Thesis, Faculty of Education, Tokyo University of Education (1973).