Ann. Bull. RILP, No. 11, 93-103 (1977)

DISCRIMINATION AND SHORT-TERM RETENTION OF SPEECH AND NON-SPEECH STIMULI BY NORMAL AND HEARING-IMPAIRED SUBJECTS*

Hiroya Fujisaki and Akira Yokkaichi**

1. Introduction

Extraction of information by sensory processing of signals and its retention in short-term memory are the essential requisites for the hierarchical structure of information acquisition, and are especially important in processing linguistic information conveyed by speech, because of the sequential characteristics of acoustic signals. It is apparent that multiple stages of processing and short-term retention are involved in speech perception where the ultimate decoding of information contained in a sentence of several seconds must be based on recognition and integration of units with much smaller time span such as words, syllables and phonemes.

Difficulties of speech perception in cases of serious hearing impairment may thus arise not only from reduced ability of peripheral auditory processing, but also from impairments at various intermediate stages of processing and retention of information. It is therefore desirable that the design of hearing and speech processing aids as well as the training procedures should be based on the analysis of the nature of impairments in individual cases.

Although a number of studies have been published on auditory processing and short-term retention of simple stationary acoustic stimuli such as pure tones and wide band noises,[1-5] little has been studied on more complex, transient acoustic stimuli such as tones with time-varying frequencies and complex sounds with time-varying resonance frequencies,[6] which respectively constitute basic acoustic correlates of suprasegmental and segmental information of speech. The objective of the present investigation is to clarify the effects of basic acoustic characteristics upon detection and retention of simple non-speech stimuli, and to relate the results with those found in the case of speech stimuli both in normal and hearing-impaired subjects, thereby obtaining an insight into various causes of difficulties of speech perception and effective means for their remedy. The present paper describes our preliminary analysis of the temporal effects on discrimination and short-term retention of frequency of pure tones and complex sounds with a formant structure, as well as of speech sounds consisting of nonsense syllables.

* Paper presented at Research Conference on Speech-Processing Aids for the Deaf at Gallaudet College, May 1977.

** Department of Special Education, Faculty of Education, Tokyo University of Education

2. Experiments on Non-Speech Stimuli

Temporal variation in the fundamental frequency of speech is the basic acoustic correlate of accent and intonation, while rapid temporal variations in formant frequencies characterize a major class of consonants as contrasted to vowels with more or less stationary formants. As a basis of quantitative understanding of such temporal effects in processing and retention of acoustic stimuli, a set of frequency discrimination tests was conducted using synthetic non-speech stimuli with both stationary and transient characteristics.

Method

The measurement of difference limen was performed by the A-X procedure for paired comparison, with parameters of stimulus presentation shown in Fig. 1. Each stimulus had linear rise and decay to avoid clicks, and the inter-stimulus interval (ISI) was set at 0.5 sec or 2.5 sec. Each stimulus pair was followed by a pause of 5 sec for written response, and a brief tone of 1000 Hz was inserted at every 10 pairs. For stationary stimuli the subjects were instructed to indicate which of the two stimuli in a pair sounded higher in pitch, while for transient stimuli, they were instructed to indicate which had more rapid transients than the other. The stimuli were generated by a digital computer and were read out at a sampling rate of 10 kHz and an accuracy of 10 bits, and were recorded on an analog tape recorder for off-line experiments.

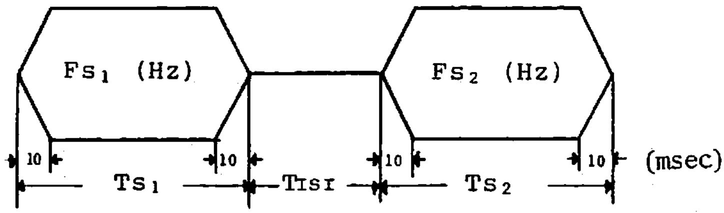

Fig.1. Parameters of stimulus presentation used in discrimination tests of stationary and transient tones.

Stimulus Parameters

1) Frequency discrimination tests of stationary tones and formants

Discrimination tests were designed to measure frequency difference limen (DL) of stationary tones and formants both at 500 Hz. For pure tones, DL was measured at durations of 0.1 sec and 0.5 sec, and at ISI's of 0.5 sec and 2.5 sec. Stimuli for discrimination tests of formant frequency had a formant bandwidth of 60 Hz, with duration and ISI selected at 0.1 sec and 0.5 sec, respectively. Furthermore, frequencies of the first stimulus (S_1) and the second stimulus (S_2) in a pair, were selected such that

$$F_{S1} - F_{S2} = \pm n \cdot \Delta F, \quad n = 0, 1, 2, 3,$$

and

$$(F_{S1} + F_{S2})/2 = 500 \text{ Hz},$$

thus resulting in a set of seven stimulus pairs constructed from seven stimuli, as shown in Fig. 2. The parameters of stimuli used in these experiments are listed in Table 1.

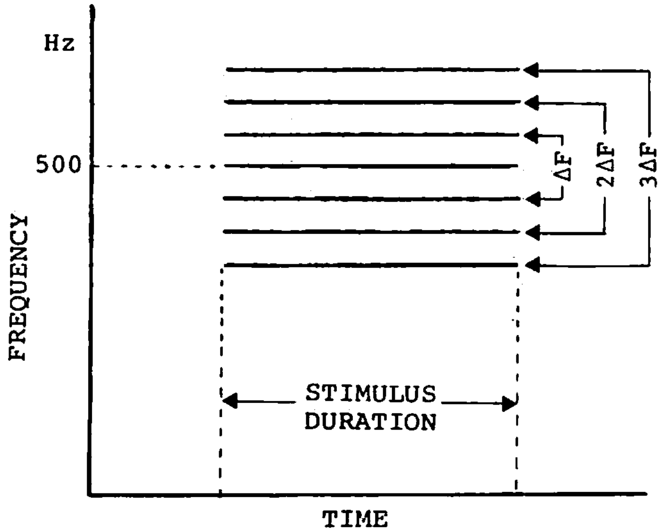

Fig. 2. Seven stimuli used in a paired comparison test of frequency discrimination for stationary tones.

Table 1. Parameters of stationary stimuli used in frequency discrimination tests.

STIMULUS	DURATION (SEC)	ISI (SEC)
500Hz TONE	0.1	0.5
"	0.5	0.5, 2.5
500Hz SINGLE FORMANT	0.1	0.5

2) Frequency discrimination tests of transient tones and formants

Discrimination tests were also designed to measure frequency difference limen of transient tones and formants. The frequency of a transient tone stimulus increased linearly from the initial frequency chosen at around 250 Hz, to the final frequency fixed at 500 Hz. The initial frequency was varied in seven equal steps to produce seven transient stimuli as shown in Fig. 3, which were then utilized for discrimination tests in the same way as in the case of stationary tone stimuli. Frequencies of transient formant stimuli were selected in a similar manner. Difference limen for transient tones were measured at durations of 0.1 sec and 0. 5 sec, while DL for transient formants were measured at a duration of 0.1 sec. Initial frequencies of these transient stimuli were selected in the same way as in the case of stationary stimuli. The parameters of stimuli used in these experiments are listed in Table 2.

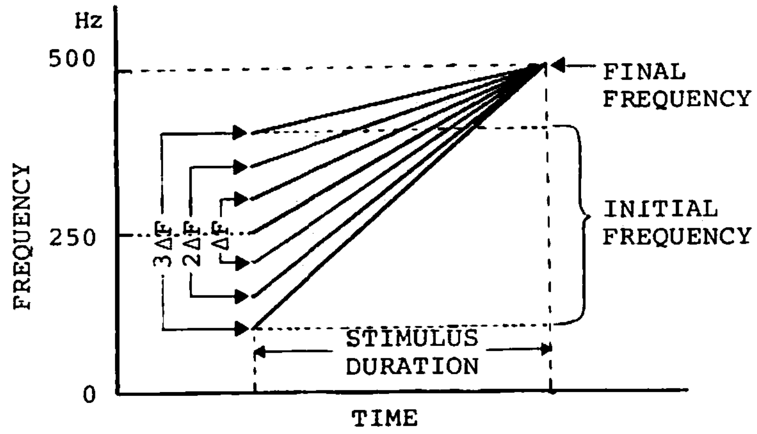

Fig. 3. Seven stimuli used in a paired comparison test of frequency discrimination for transient tones.

Table 2: Parameters for transient stimuli.

STIMULUS	DURATION (SEC)	ISI (SEC)
TONE	0.1	0.5
"	0.5	0.5, 2.5
SINGLE FORMANT	0.1	0.5

Subjects and Experimental Procedure

Eight subjects with normal hearing and seven with sensorineural deafness took part in the experiments. The ages of hard-of-hearing subjects ranged from 15 to 30, and their hearing losses at 500 Hz ranged from 55 to 75 dB. The stimuli were presented monoaurally at a sensation level of 50dB in the case of normal subjects, while there were presented at the most comfortable level for each individual in the case of hard-of-hearing subjects. The number of subjects for each test condition was 3 or 4 both in normal and in hard-of-hearing groups, and each subject made 20 to 80 judgments on each stimulus pair. Because of individual differences, the results were analyzed separately for each individual subject.

The result of a discrimination test can be illustrated by Fig. 4, where the probability that the second stimulus S_2 is judged higher in pitch than the first stimulus S_1 is plotted on a normal scale against the difference in their frequencies, $F_{S2} - F_{S1}$ in the case of stationary tones. The result can generally be approximated fairly closely by a normal distribution. The solid line in Fig. 4 indicates an approximation to the measured data on the basis of the least-mean-squared error weighted by Müller-Urban coefficients. The mean value μ and the standard deviation σ respectively indicate the time-order error and the accuracy of comparative judgment. The value of σ thus obtained shall be defined as the difference limen (DL) throughout the present paper. The dependency of σ upon ISI can be regarded as indicating the decay of short-term memory of S_1. The conventional

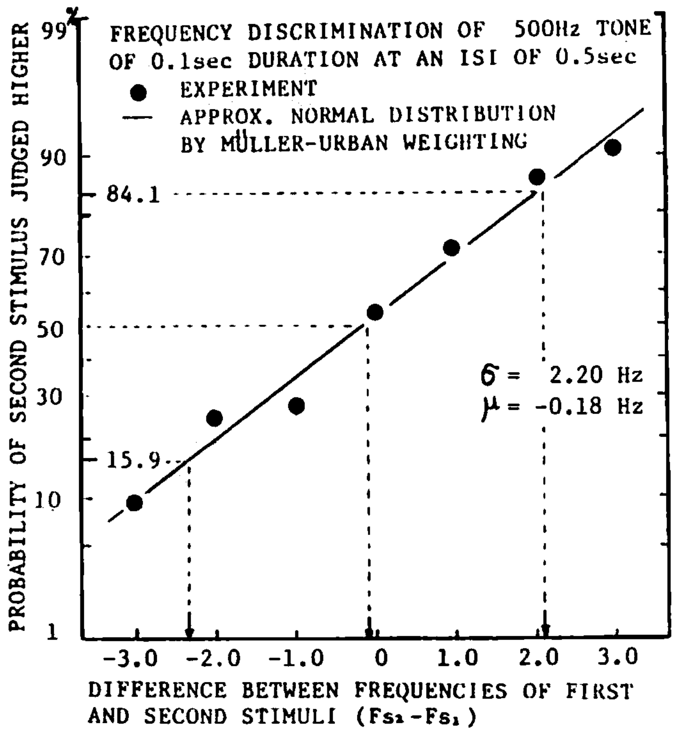

Fig. 4. An example of results of frequency discrimination test for 500 Hz tone of 0.1 sec duration at an inter-stimulus interval of 0.5 sec.

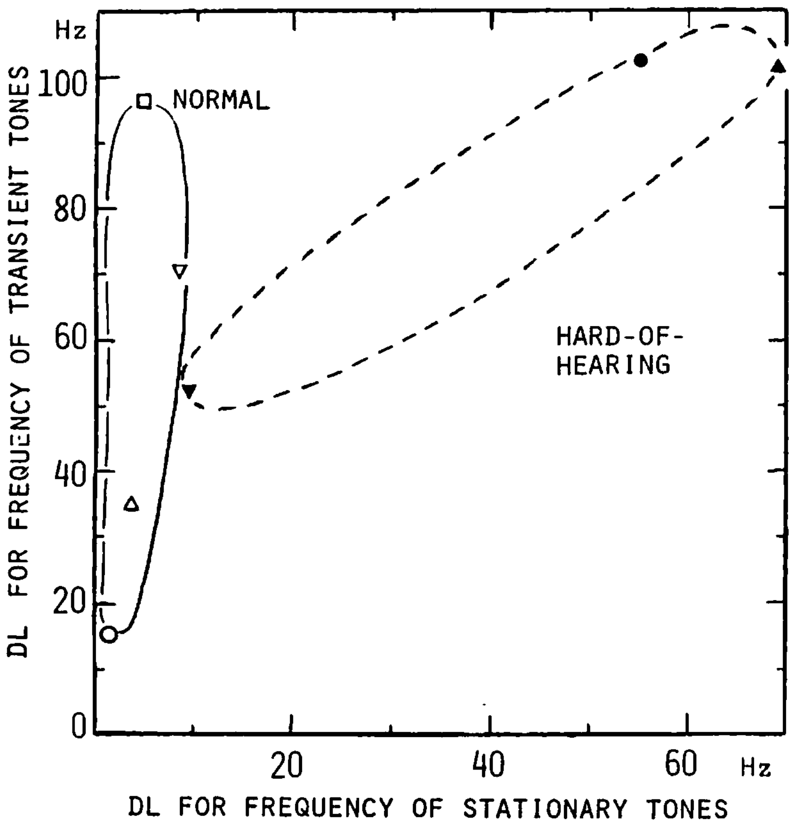

Fig. 5. Relationships between difference limen for frequency of stationary and transient tones of 0.1 sec duration.

difference limen, which is defined by $|F_{S2} - F_{S1}|$ corresponding to 75 % correct judgment, is equal to 0.675 σ in the absence of the time-order error μ, but tends to vary with μ and is always greater than 0.675 σ for non-zero μ.

Difference limen for stationary formant frequencies can be defined and measured in exactly the same way as for stationary tones. On the other hand, difference limen for transient signals can be defined on the scale of the initial frequency of linear transition.

Results

1) Comparison of frequency difference limen for stationary and transient tones

Figure 5 shows a comparison of DL for stationary tones and for transient tones at a duration of 0.1 sec and an ISI of 0.5 sec. Results of four normal subjects indicate comparatively large individual differences, but the ratio of DL for transient tones to that for stationary tones is found to fall within the range of 8 to 18. Results of three hard-of-hearing subjects show still greater individual differences in DL for stationary tones, ranging from one case which falls within the nromal group to two other cases whose DL's are at least several times larger than those of the normal group. It is interesting to note, however, that their DL for transient tones show much smaller individual differences and are very close to the normal range. A similar relationship can also be found in Fig. 6, which shows the results for stimulus duration of 0.5 sec.

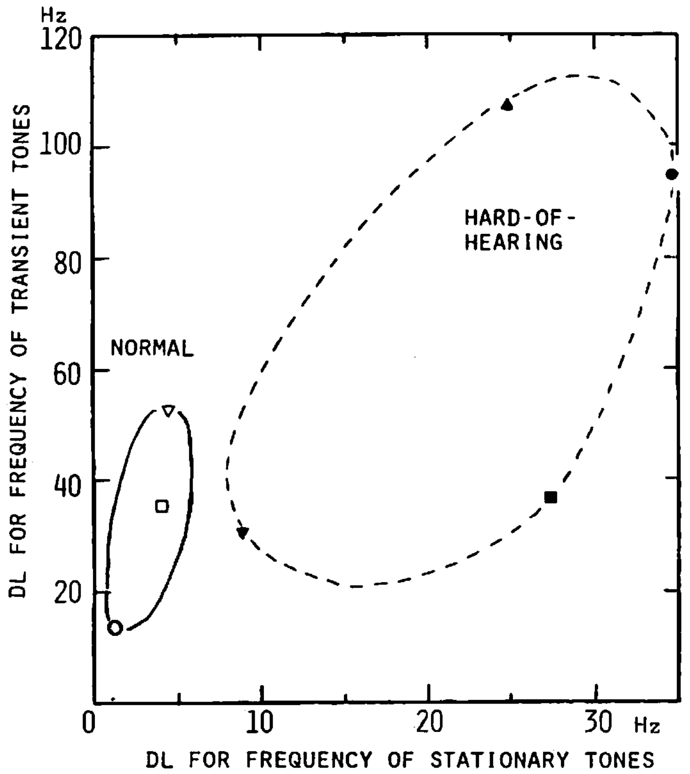

Fig. 6. Relationships between difference limen for frequency of stationary and transient tones of 0.5 sec duration.

2) Comparison of frequency difference limen for stationary and transient formants

Figure 7 shows a comparison of DL for stationary formants and for transient formants at a duration of 0.1 sec and an ISI of 0.5 sec. Results of four normal subjects again indicate fairly large individual differences, and the ratio of two DL's for each subject ranges from 2 to 8. One of the hard-of-hearing subjects falls within the region of normal subjects, while the other two fall far outside. In contrast to the results for tones, however, the ratio of two DL's for the hard-of-hearing group fall within the range of normal subjects, indicating that the degradations in both difference limens are approximately the same.

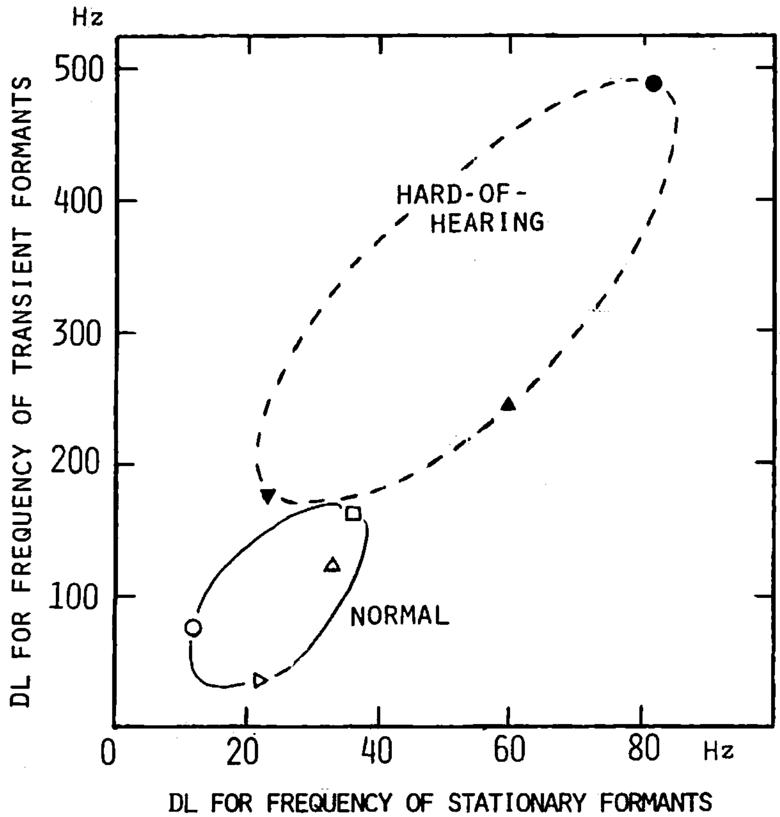

Fig. 7. Relationships between difference limen for frequency of stationary and transient formants.

3) Influence of ISI on frequency difference limen for stationary and transient tones

Figure 8 shows the effect of ISI upon DL for stationary tones. Individual differences within each group are found to be quite large, but the rate of degradation of DL with ISI is comparatively small for both normal and hard-of-hearing groups, being slightly larger for the latter. The effect of ISI is found to be still smaller and the difference between normal and hard-of-hearing groups are also smaller in Fig. 9, which shows the results for transient tones. The relative stability of DL for transient tones as

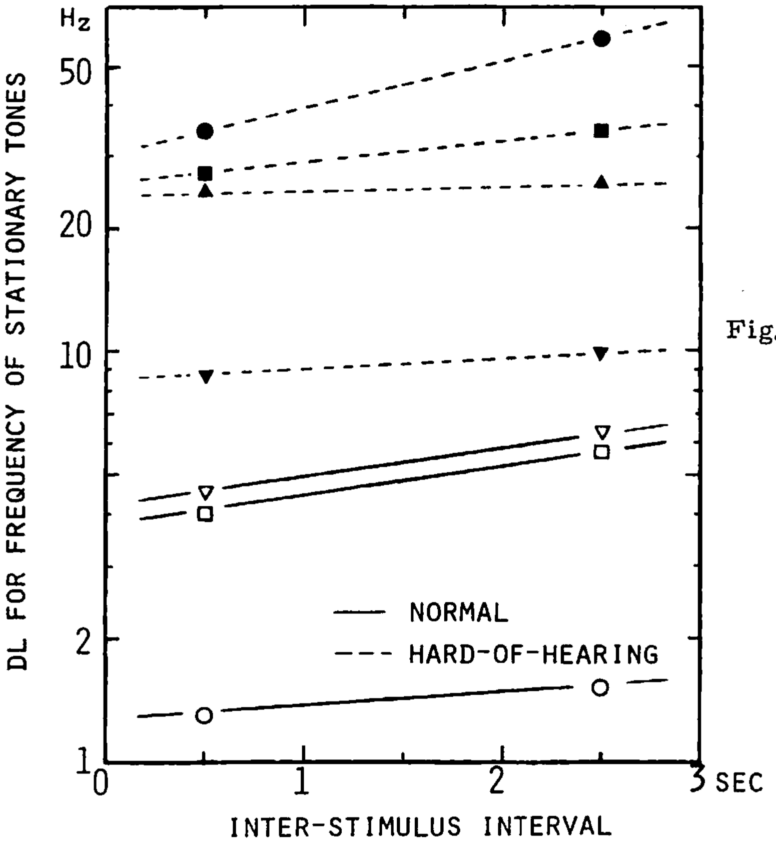

Fig. 8. Influence of inter-stimulus interval on difference limen for frequency of stationary tones.

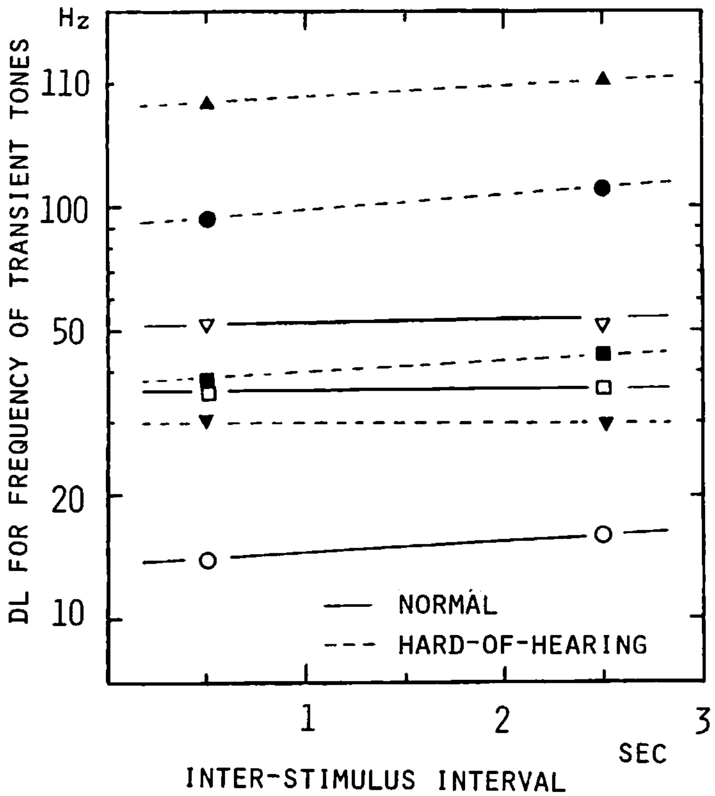

Fig. 9. Influence of inter-stimulus interval on difference limen for frequency of transient tones.

compared to that for stationary tones may be ascribed to the fact that transient stimuli are perceived more categorically than stationary stimuli.[7]

3. Experiment on Speech Stimuli

While it is still too early to relate the findings of non-speech experiments of the preceding section with various impairments of speech perception observed in hard-of-hearing subjects, a preliminary study was also conducted to investigate short-term retention of segmental information using natural speech sounds.

Stimuli

The speech materials consisted of 15 nonsense disyllables as shown in Table 3. These disyllables were uttered and recorded by a female speaker of the Tokyo dialect. The durations of these disyllables ranged from 0.43 to 0.58 sec. Sixteen sequences of four disyllables, randomly selected out of the fifteen, were then constructed. The intervals between successive disyllables were set at 0.5 sec in eight sequences, while they were set at 2.5 sec in the remaining eight sequences. Frequency-modulated tones were further inserted in these intervals to interfere with subject's rehearsal, as shown in Fig. 10.

Table 3. Fifteen nonsense disyllables used in short-term retention test of speech sounds.

[ri-hi]	[ta-mu]	[dʒi-te]	[sa-de]	[nu-to]
[me-zu]	[ze-no]	[tsu-wa]	[ki-da]	[so-bu]
[go-he]	[pu-bo]	[za-ʃi]	[ma-jo]	[ga-pe]

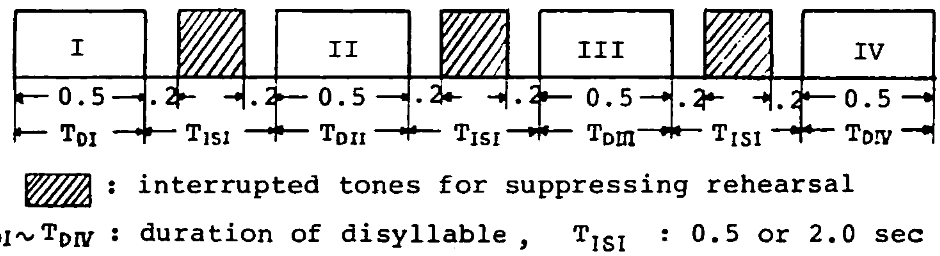

Fig. 10. Parameters of stimulus presentation in short-term retention tests of speech sounds using quadruplets of nonsense disyllables.

Experimental Procedure and Subjects

The experiment on short-term retention of speech sounds consisted of two tests: 1) the identification test of all disyllables, and 2) the sequence reproduction test conducted a week after the identification test. In the identification test, each disyllable was presented five times in random order to the subject, who was instructed to write down what he heard immediately after each disyllable. In the reproduction test, on the other hand, the sixteen sequences, each consisting of four disyllables, were presented in random order to the subject, who was instructed to write down his judgment after the end of each sequence. The results of identification test was analyzed and the subject's most consistent response was adopted as the correct response for the subject in scoring his performance in the reproduction test.

The subjects were nine children of 10 to 12 years of age with medium sensorineural hearing loss, being educated in a special class for the hard-of-hearing. Nine children with normal hearing of the same age were also tested for comparison.

Results

Results of normal and hard-of-hearing groups are compared in Fig. 11. In the normal group, a majority of subjects show little effect of ISI on the percentage of correct recall, while a majority of subjects in the hard-of-hearing group shows degradation of retention for increased ISI. The difference in the mean performances of the two groups are highly significant for both ISI conditions. Further analysis of results indicated that errors due to misplacement of correct responses occur more frequenctly in the hard-of-hearing group than in the normal group, the difference being again highly significant.

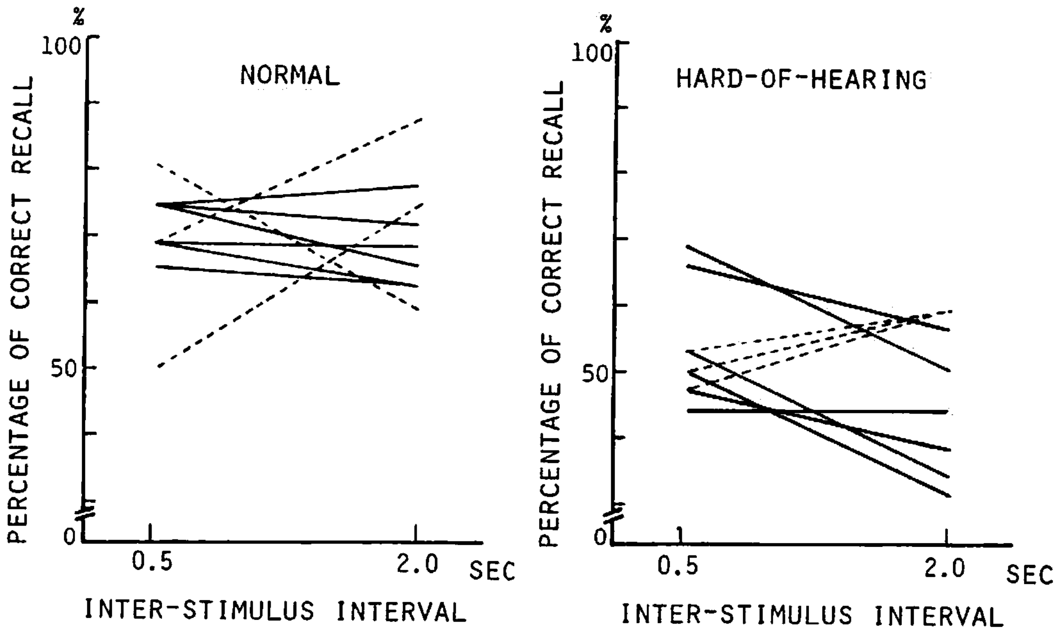

Fig. 11. Percentage of correct recall of speech sounds in quadruplets of nonsense disyllables by normal and hard-of-hearing subjects.

4. Summary and Conclusions

A series of frequency discrimination tests, using synthetic non-speech stimuli with various degrees of similarity to the speech signal, has benn conducted on both normal and hard-of-hearing subjects, to investigate and analyze the nature of impairments in processing and retention of auditory information in hard-of-hearing subjects. Experiments on stationary and transient tones have indicated that the hard-of-hearing subjects generally show much greater deterioration for stationary tones than for transient tones. The influence of inter-stimulus interval on frequency difference limen has also been investigated to examine the temporal decay of short-term memory, indicating a slightly larger decay rate in the hard-of-hearing group than in the normal group. Experiments on stationary and transient formants have indicated a similar tendency. Preliminary experiments have also been conducted on short-term retention of segmental information in the recall of sequences of nonsense disyllables, also indicating a greater rate of decay in the hard-of-hearing subjects than in the normal subjects.

Acknowledgments

The research reported here was supported by a Grant in Aid for Scientific Research (No. 120107) from the Ministry of Education. The authors wish to express their thanks to Prof. Kenkichi Suhara of Tokyo University of Education for his kind encouragement, and to Drs. Sadao Shibata, Osamu Tamaki and Nobuyuki Fujino of the National Center of Speech and Hearing Disorders for their cooperation.

References

1. Postman, L. (1946), The Time-error in Auditory Perception, American Journal of Psychology, 59, 193-219.
2. Harris, J. D. (1952), The Decline of Pitch Discrimination with time, Journal of Experimental Psychology, 43, 96-99.
3. Bachem, A. (1954), Time Factors in Relative and Absolute Pitch Determination, The Journal of the Acoustical Society of America, 26, 751-753.
4. Harris, J. D. (1966), Masked DL for Pitch-Memory, The Journal of the Acoustical Society of America, 40, 43-46.
5. Pickett, J. M. and E. S. Martin (1967), Discrimination of Vowel Spectrum in Sensorineural Hearing Impairment, The Journal of the Acoustical Society of America, 42, 1191.
6. Danaher, E. M., M. J. Osberger and J. M. Pickett (1973), Discrimination of Formant Frequency Transitions in Synthetic Vowels, Journal of Speech and Hearing Research, 16, 439-451.
7. Fujisaki, H. and T. Kawashima (1970), Some Experiments on Speech Perception and a Model for the Perceptual Mechanism, Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 29, 207-214.