Ann. Bull. RILP, No. 12, 149-156 (1978)

IDENTIFICATION OF SYNTHETIC SPEECH STIMULI BY HEARING-IMPAIRED SUBJECTS

Akira Yokkaichi* and Hiroya Fujisaki

1. Introduction

The auditory capability of a hearing-impaired person is most commonly evaluated by means of the pure-tone audiometry. Since the acoustic features of speech are quite complex, however, the evaluation based on the pure-tone audiometry is not necessarily sufficient for the purpose of finding the causes of various impairments in speech perception as well as finding effective means for their remedy.

One of the directions to obtain deeper understanding of the impairments in speech perception is to investigate the difference limen for various attributes of simple as well as complex acoustic stimuli, and a number of studies has already been published on discrimination of pure-tones, wide-band noises, synthetic vowels and vowel-like sounds [1~8]. Difference limens obtained in these studies, however, do not serve for the direct evaluation of speech perception, and the relationship between these limens and the speech perception is yet to be clarified.

On the other hand, the speech audiometry using natural utterances of monosyllables has been introduced for directly measuring the capability of speech perception in hearing-impaired subjects. The use of natural utterances in the conventional speech audiometry, however, introduces various causes of variability and does not allow free and accurate control of stimulus parameters. These difficulties can be circumvented by using synthetic speech sounds, and identification tests using synthetic stimuli can be a more powerful means for quantitative investigation of the ability of speech perception than the conventional speech audiometry.

From this point of view, synthetic vowels were adopted by Fujisaki, Tomisawa and Sato [9, 10] in a series of identification tests for hearing-impaired subjects, and the results were correlated with the results of the pure-tone audiometry. Few studies, however, have been made on the identification of synthetic speech stimuli with time-varing characteristics, which are known to be quite important in transmitting a major part of the linguistic information.

The present study aims at investigating the effects of acoustic features of speech on difficulties of speech perception by means of identification tests using synthetic speech. The accuracies of identification for vowels with stationary formants, a semivowel or a liquid with relatively slow formant transitions, and voiced stop consonants with relatively rapid formant transitions are measured and compared with each other, and the performance of hearing-impaired subjects are compared with those of normal subjects [11].

* Faculty of Engineering, University of Tokyo

2. Method

2.1 Stimulus parameters

The speech stimuli used in this study were steady-state vowels, or vowel-like sounds generated by appropriate control of formant frequencies of vowels [12,13]. The third, fourth, and fifth formant frequencies of all the stimuli used in the following experiments were fixed at 2700, 3500, and 4500 Hz respectively, while the bandwidths of the five formants were fixed at 60, 130, 200, 270, and 340 Hz, respectively. In the case of CV syllables, the vowels were always /a/, with the first and second formant frequencies at 780 Hz and 1200 Hz, respectively. Furthermore, the fundamental frequency of the stimuli was always 116 Hz, and the duration of the stimuli was 180 msec including the rise and decay times.

Experiment 1: Identification of /u/ and /i/.

Figure 1 (a) shows the amplitude pattern and the range of formant frequencies of the vowel stimuli. By varing the second formant frequency from 900 Hz to 2340 Hz at nine equal steps, ten stimuli were synthesized to cover the range from the vowel /u/ to the vowel /i/. In these stimuli, the first formant frequency was 300 Hz, and the rise and decay times of the amplitude were 25 msec.

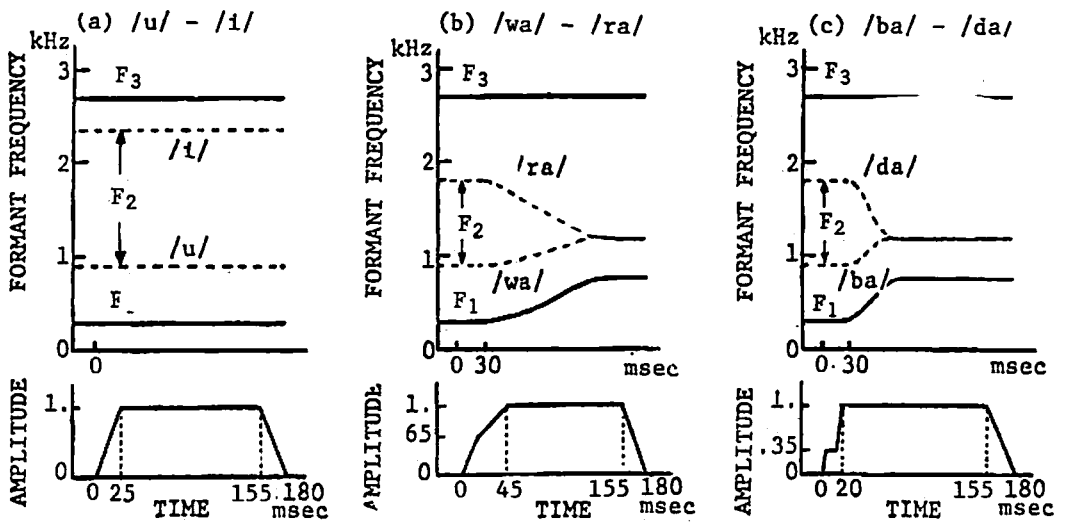

Fig. 1. Schematic drawings of formant and amplitude patterns used in identification tests of /u/-/i/, /wa/-/ra/, and /ba/-/da/.

Experiment 2: Identification of /wa/ and /ra/.

Figure 1 (b) shows the amplitude pattern and the range of formant trajectories of the semivowel/liquid stimuli. By varing the initial frequency of the second formant from 900 Hz to 1800 Hz at nine equal steps, ten stimuli were synthesized to cover the range from /wa/ to /ra/. The initial frequency of the first formant was always 300 Hz, and the trajectory of the

formant transition from the consonant to the following vowel was approximated by the step response of a critically-damped second-order linear system. The time constants of transition for the first and second formants were 15.0 msec and 17.5 msec, respectively. The rise time of the amplitude was 45 msec.

Experiment 3: Identification of /ba/ and /da/.

Figure 1 (c) shows the amplitude pattern and the range of formant trajectories of the voiced stop consonant stimuli. The stimulus parameters were the same as used in Experiment 2, except that the time constants of the first and second formant transitions were 5 msec, and the stimuli had a 20 msec buzz bar preceding the formant transitions. In this way, ten stimuli were synthesized to cover the range from /ba/ to /da/.

Some of the important stimulus parameters used in Experiment 1 through 3 are listed in Table 1.

Table 1. Stimulus parameters used in Experiment 1 - 3.

Experiment & stimuli		F2 (Hz)	TC1 (msec)	TC2 (msec)
Exp. 1	/u/-/i/	900 - 2340	—	—
Exp. 2	/wa/-/ra/	900 - 1800	15.0	17.5
Exp. 3	/ba/-/da/	900 - 1800	5.0	5.0

F2 : range of initial frequencies of second formant
TC1 : time constant of first formant transition
TC2 : time constant of second formant transition

2.2 Experimental procedure

A test material consisted of a randomized sequence of 110 stimuli containing 10 each of the original stimuli, preceded and followed by five dummies. Successive stimuli were separated by five seconds for response, and a brief 1000 Hz tone was inserted at every 10 stimuli. The synthesis and compilation of the stimuli were performed on a digital computer, and the output was fed to a digital-to-analog converter at a rate of 10 kHz with an accuracy of 10 bit/sample, to be recorded on a tape recorder for off-line experiments.

Four subjects with normal hearing and four with medium-to-severe sensorineural hearing loss took part in the experiments. The age of hearing-impaired subjects ranged from 16 to 23, and the averaged hearing losses of individual subjects ranged from 63 to 79 dB. The stimuli were presented monoaurally at a sensation level of 50 dB in the case of normal subjects, while they were presented at the most comfortable level for each individual in the case of hearing-impaired subjects. Each subject was asked to select, by forced judgment, /u/ or /i/ in Experiment 1, /wa/ or /ra/ in Experiment 2, and /ba/ or /da/ in Experiment 3. The total number of judgments on one stimulus ranged from 20 to 70 per subject. The result of an identification test is illustrated by Fig. 2, where the probability that the stimulus is judged as /i/ is plotted on a normal scale against the initial frequency of the second formant transition. The straight line in Fig. 2 indicates an approximation of the measured data by a normal distribution

obtained by the maximum-likelihood method. The mean value μ and the standard deviation σ respectively indicate the phoneme boundary and an index of the accuracy of categorical judgment necessary for identification. The value of σ thus obtained shall be called as the identification accuracy throughout the rest of the paper.

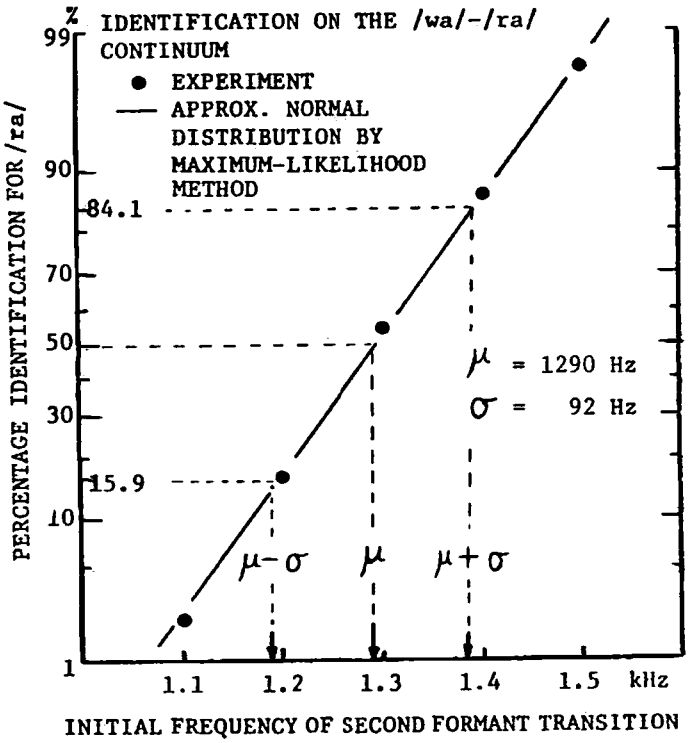

Fig. 2. An example of results of identification test on the /wa/-/ra/ continuum.

3. Results and Discussion

3.1 Identification accuracy for stationary vowels, a semivowel and a liquid, and voiced stop consonants.

The results of the three identification tests are summarized in Fig. 3. The ordinate indicates the initial frequency of the second formant transition, and the symbols "●▲■▼" and "○△□▽" respectively indicate identification accuracies of the four hearing-impaired subjects and the four normal subjects. An arrow in the figure indicates that it was impossible to measure the identification accuracy for stimuli with faster formant transitions. The accuracy of identification for all three groups of stimuli fell within the range from 70 Hz to 100 Hz in normal subjects. In the case of hearing-impaired subjects, all the four subjects could identify the stationary vowel stimuli, but only two of them could consistently identify the semivowel/liquid stimuli, of whom only one could identify the voiced stop stimuli. Comparison of identification accuracies for normal and for hearing-impaired subjects indicates a marked deterioration in hearing-impaired subjects for less stationary stimuli.

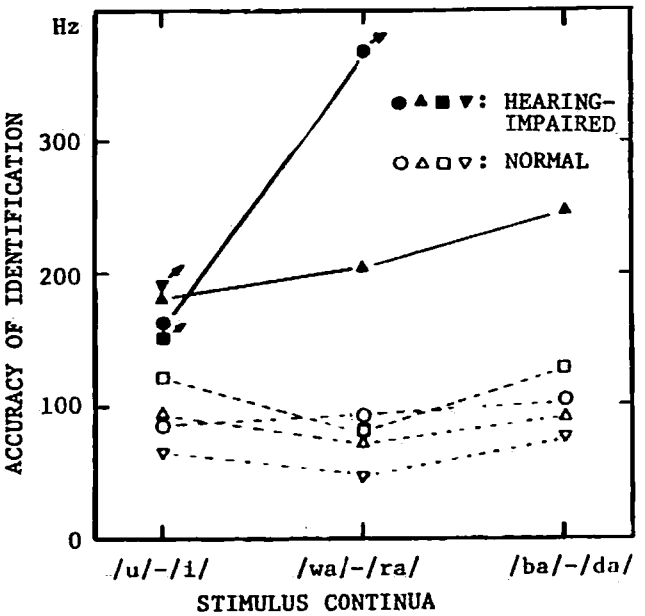

Fig. 3. Accuracy of identification obtained in various identification tests for normal and hard-of-hearing subjects.

3.2 Phoneme boundaries for stationary vowels, a semivowel and a liquid, and voiced stop consonants.

Figure 4 illustrates the phoneme boundaries of /u/-/i/, /wa/-/ra/, and /ba/-/da/ represented by the initial frequency of the second formant.

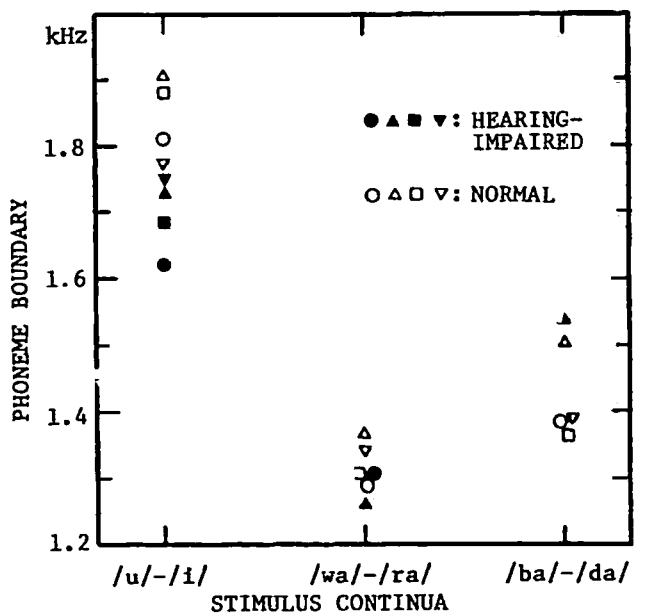

Fig. 4. Phoneme boundaries obtained in various identification tests for normal and hard-of-hearing subjects.

The boundary between the stationary vowels /u/ and /i/ was about 1700 Hz for the hearing-impaired subjects and was about 1800 Hz for the normal subjects. On the other hand, the boundary between /wa/ and /ra/ as well as the boundary between /ba/ and /da/ was not appreciably different in normal and hearing-impaired subjects.

3.3 Relationship between the identification accuracy and the articulation score.

The relationship between the identification accuracies and articuration scores for /u/-/i/, /wa/-/ra/, and /ba/-/da/ is illustrated in Fig. 5.

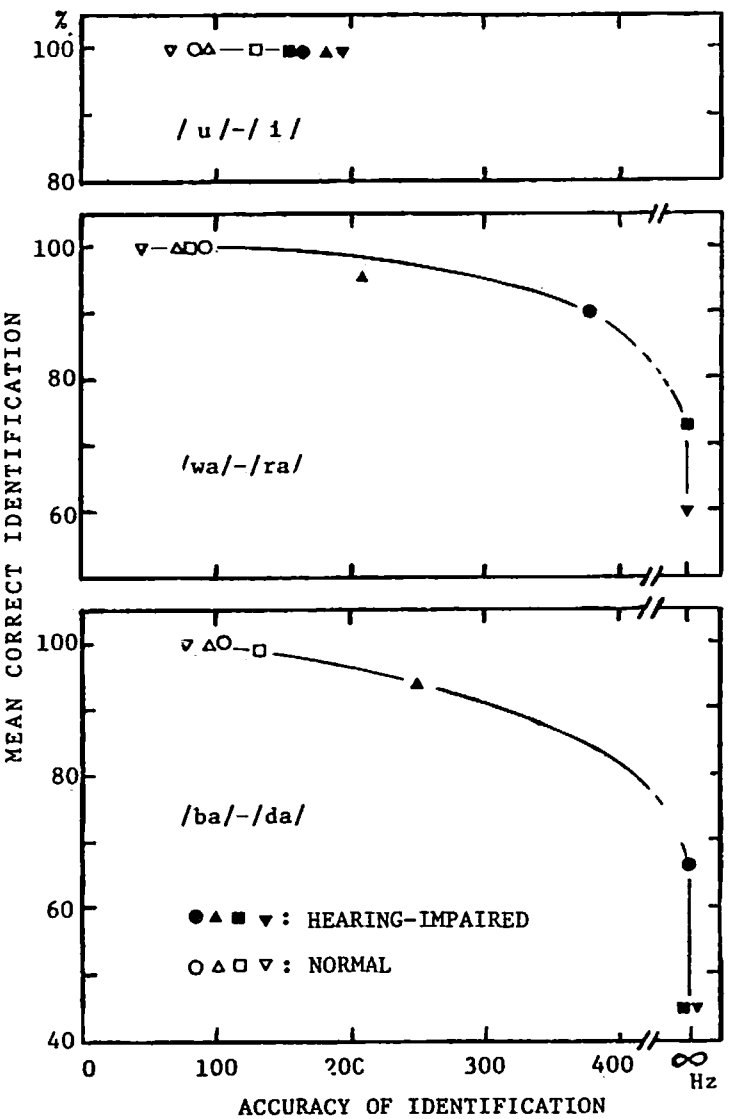

Fig. 5. Relationships between accuracy of identification and mean correct identification.

The ordinate in each panel indicates the mean articulation score obtained by averaging the scores for the two extreme stimuli on the stimulus continuum used for an identification test, and the abscissa indicates the identification accuracy. Each symbol indicates a subject's performance. As seen from the figure, there are many cases where two subjects show a large difference in the identification accuracy, even if no appreciable differences are found in the articulation score. These results indicate that the identification accuracy is a more sensitive measure for the ability of speech perception than the articulation score.

The results of these experiments confirmed that the difficulty of speech perception in hearing-impaired subjects generally increases as the stimulus becomes less stationary. Namely, semivowels and liquids are found to be more difficult, and voiced stops are found to be still more difficult to identify than vowels, possibly because of the decrease in the overall energy as well as in the temporal redundancy.

It is to be noted that the performance data of the hearing-impaired subjects were obtained when the stimuli were presented at the most comfortable level for each individual subject and for each stimulus category. Since this condition is scarcely satisfied in ordinary situations, the difficulties of these hearing-impaired subjects in their daily communication are considered to be greater than in these experimental conditions. Effective utilization of their residual hearing will therefore require not a mere amplification nor a simple shift of the frequency components, but a processing of the speech signal that will transform the acoustic features of individual speech sounds to match their perceptual capability.

4. Summary and Conclusions

For the purpose of obtaining quantitative understanding of the difficulties of speech perception in hearing-impaired subjects, their ability of identifying various speech sounds were measured and compared with normal hearing subjects using synthetic speech stimuli. In order to examine the effect of change of acoustic features on perceptual difficulties, vowels with stationary formant frequencies as well as vowel-like sounds with different rates of change of formant frequencies were used as stimuli. It was found that the difficulty of speech perception in hearing-impaired subjects generally increases as the stimulus becomes less stationary. It was also found that the identification accuracy is a more sensitive index for the difficulty of speech perception than the conventional articulation score.

Acknowledgments

The research reported here was supported by a Grant-in-Aid for Scientific Research (No. 239005) from the Ministry of Education. The authors wish to express their thanks to Drs. Sadao Shibata and Osamu Tamaki of the National Center of Speech and Hearing Disorders for their cooperation.

References

1. Harris, J. D. (1952); Pitch Discrimination, The Journal of the Acoustical Society of America, 24, 750-755.

2. Pickett, J. M. and E. S. Martin (1968); Some Comparative Measurements of Impaired Discrimination for Sound Spectral Differences, American Annals of the Deaf, 113, 259-267.

3. Pickett, J. M. and J. Mártony (1970); Low-frequency Vowel Formant Discrimination in Hearing-impaired Listeners, Journal of Speech and Hearing Research, 13, 347-359.

4. Danaher, E. M., M. J. Osberger, and J. M. Pickett (1973); Discrimination of Formant Frequency Transients in Synthetic Vowels, Journal of Speech and Hearing Research, 16, 439-451.

5. Eguchi, S. (1973); Difference Limen for the Formant Frequency in the Population of Adult and Children Aged from 7 to 15 Years Old, Audiology Japan, 16, 131-136.

6. Mártony, J. (1974); Some Psychoacoustic Tests with Hearing Impaired Children, Speech Transmission Laboratory, Quarterly Progress and Status Report, 2-3, 72-89.

7. Deguchi, T. and S. Kuroki (1975); Frequency Discrimination of Hard-of-hearing Children, Transactions of the Committee on Speech Research, Acoustical Society of Japan, S 74-42.

8. Fujisaki, H. and A. Yokkaichi (1977); Discrimination and Short-term Retention of Speech and Non-speech Stimuli by Normal and Hearing-impaired Subjects, Annual Bulletin, Research Institute of Logopedics and Phoniatrics, No. 11, 93-103.

9. Fujisaki, H., M. Tomisawa and T. Sato (1969); Speech Audiometry by Synthetic Japanese Vowels, Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 28, 74-79.

10. Tomisawa, M. (1971); Speech Audiometry by Synthetic Japanese Vowels, Audiology Japan, 14, 202-229.

11. Yokkaichi, A. and H. Fujisaki (1978); Identification of Synthetic Speech Stimuli in Hard-of-hearing Subjects, Transactions of the Committee on Speech Research, Acoustical Society of Japan, S77-60.

12. Fujisaki, H. and T. Karaki (1966); Synthesis of Semi-vowels, Liquids and Voiced Stop Consonants by a Terminal Analog Speech Synthesizer, Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 25, 105-112.

13. 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.