Discrimination of vowel sounds of very short duration*

CHING YEE SUEN and MICHAEL P. BEDDOES
Department of Electrical Engineering
The University of British Columbia, Vancouver 8, British Columbia, Canada

An experiment was conducted to investigate the discrimination of six vowel sounds of 10 msec duration. Vowels of equal pitch and intensity level were generated by computer. Both phonetically trained (PT) and untrained (UPT) Ss were used. Rapid learning took place, and the PT Ss showed much better discrimination than the UPT Ss. Confusion matrices of the last four learning blocks indicated that /i/ and /u/ sounded very much alike when they were short. The pattern of the test scores was discussed with reference to pure tone perception.

In this paper, we aim to shed some light on the fascinating problem of perception of very short speech sounds; the sounds themselves are very short segments (little more than a period) taken from the quasiperiodic vowels. Specifically, we ask, "Can we perceive a vowel if only 10 msec of it is heard?"

Bürck, Kotowski, and Lichte (described in Stevens & Davis, 1937) found that the duration of 10 msec of a low-frequency tone was not long enough for identification of the pitch of the tone. However, Peterson (1939), Gray (1942), and Joos (1948) reported that vowel fragments less than 10 msec could be recognized.

Although there has been intensive study of the durations of vowels (e.g., Sharf, 1964, and references therein), relatively less work has been done relating the durations of vowels to their recognition (Siegenthaler, 1950; Tiffany, 1953; Schwartz, 1963; Fujisaki & Kawashima, 1968). Powell and Tosi (1970) found that the median recognition threshold varies from vowel to vowel from about 10 to 30 msec. The results of their experiments, however, were not in good agreement with those previously obtained by Peterson (1939), Gray (1942), and Joos (1948) on vowel fragments. In other studies (Schwartz, 1963; Fujisaki & Kawashima, 1968), it was reported that vowels could be correctly identified at durations of the order of 30 msec. This experiment shows that both PT and UPT Ss can learn to discriminate vowels of 10 msec duration.

PROCEDURE

Preparation of Vowels

Since the fundamental frequency limen is about 0.3% to 0.5% of the fundamental frequency (Flanagan, 1965) and intensity discrimination may be acute with short sounds (Schwartz, 1963), precise control of intensity and fundamental frequency of the vowels is necessary.

Six vowels, /a/, /ɛ/, /e/, /o/, /u/, and /i/, were chosen. They were sustained by a male American phonetician. Since irregularities of amplitude and pitch may occur when a vowel is sustained by a human speaker (e.g., see Thomas et al, 1970), a computer-controlled method was employed to extract a basic segment (the pitch period of the voice) of the vowel waveform. This segment was then repeated a number of times to simulate the vowel waveform. The scheme of basic segment extraction was done by a segmentation program (Suen & Beddoes, 1971) with the aid of a PDP-9 computer and a precision display unit. This scheme allowed the operator to detect and extract accurately a desired segment of the vowel waveform. The sustained vowels were first low-pass filtered at 8 kHz and sampled at a rate of 20 kHz. The waveforms were then displayed on the screen of the display unit. A pitch period of 7.65 msec duration (i.e., fundamental frequency = 131 Hz) occurred in all the sustained vowels. A basic segment of each vowel with this pitch period was then extracted. The starting point of this basic segment was taken to be the zero-crossing before the major peak in the period of the vowel waveform. After these basic vowel segments had been extracted, synthesized vowels were generated and presented to both PT and UPT listeners for identification. When all these vowels were correctly identified, they were recorded on a tape. Subsequently, these synthesized vowels were played and their intensities were equalized by measurement with a rms voltmeter (Hewlett-Packard, Model 3400A). From these synthesized vowels, a second basic segment of each vowel was extracted and formed the basic segment of the synthesized vowels used in this study. All the synthesized vowels were low-pass filtered at 8 kHz before presentation to the Ss. The first three formants of these resulting vowels were measured with a variable band-pass filter (Krohn-Hite Model 3342R) and the rms voltmeter. The formant frequencies in hertz were: /a/, F1 = 760, F2 = 1,050, F3 = 2,500; /ɛ/, F1 = 580, F2 = 1,900, F3 = 2,450; /e/, F1 = 510, F2 = 2,050, F3 = 2,700; /o/, F1 = 530, F2 = 820, F3 = 2,420; /u/, F1 = 270, F2 = 660, F3 = 2,350; /i/, F1 = 260, F2 = 2,200, F3 = 2,950.

Experimental Design and Testing Procedure

A pilot experiment indicated that it was possible to discriminate among six 10-msec vowels used in this study and that there might be a great difference between PT and UPT Ss. As a result, vowels of 10 msec duration were used and two groups of Ss were employed. Six university students who had no

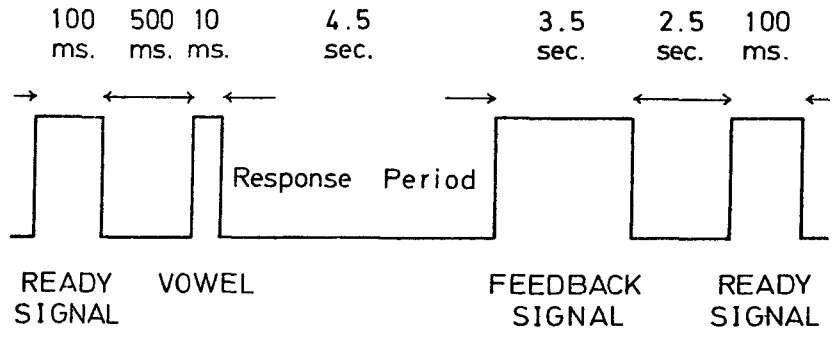

Fig. 1. Presentation of the vowel stimuli.

*This research was supported by grants from the National Research Council and the Medical Research Council of Canada. The authors are indebted to Dr. John H. V. Gilbert of the Division of Audiology and Speech Sciences of the University of British Columbia for frequent consultations and constructive suggestions. They also benefited from discussion with Dr. A.-P. Benguerel. The authors also wish to thank the reviewers of this article for suggestions and comments.

training in articulatory phonetics before formed the group of UPT Ss. Four graduate students and two faculty members, all of speech sciences, formed the group of PT Ss. The graduate students had had about 1 year of training in articulatory phonetics, and the faculty members had had about 5 years of teaching experience in phonetics and speech sciences.

The test materials consisted of six different blocks of 36 vowel stimuli each. These 36 stimuli were composed of the six tested vowels occurring six times in a block in a constrained manner so that each vowel followed itself and every other vowel in the whole block. To minimize order effect, each of the six Ss of the two groups was assigned to a given row in a 6 by 6 Latin square. At the end of the sixth block, the first two blocks were presented to the S again.

This experiment was conducted in a quiet room. Prior to the presentation of a vowel sound, a 100-msec "ready" signal of 1 kHz was generated (see Fig. 1). After hearing the vowel sound, the S was required to associate it with one of the six needle positions (indicated by Nos. 1-6) corresponding to different deflections on the meter in front of him. He was required to write down the number in the response period of 4.5 sec on a response sheet provided. The feedback signal would then deflect the needle to the position with which the vowel was to be correctly associated. After this, the ready signal would again be heard before the next vowel was presented, etc. To ensure uniformity of stimuli presentation, both indicating signals and all vowel stimuli were generated by the computer. Both the 1-kHz signal and the vowel stimuli were recorded on one channel of a stereo tape recorder (Tandberg Model 64). The signal which monitored the meter was recorded on another channel. Both the 1-kHz signal and the vowel stimuli were presented to the Ss through a loudspeaker (Ampex F2044 speaker amplifier). Prior to the test, 6 to 10 stimuli of a block were presented to the S to familiarize him with the testing procedure and to adjust the sound to his comfortable level. He was also told the six different vowels used in this experiment. To the UPT Ss, key words (father, set, chaotic, notation, pool, and beet) were used to illustrate the vowels; explanation was also provided to them when there was doubt. To avoid obvious relations between the stimulus and the number put on the meter, the numbers were changed from block to block following another 6 by 6 Latin square. Thus, deflection of the needle of the meter was the same for the same vowel throughout the whole test, but the numbers put on the meter were changed from one block to another. There was a rest period of 3-4 min between blocks, and each S spent about 1 h 20 min for this experiment. All the Ss were paid and were encouraged to try their best by

Table 1
Means and Standard Deviations of the Percent Correct Discrimination Scores of PT and UPT Ss as a Function of Test Blocks

Block	PT Ss		UPT Ss	
	Mean	SD	Mean	SD
1	51.4	17.48	25.0	10.27
2	69.0	17.23	38.4	13.56
3	76.9	23.61	61.1	12.73
4	85.2	13.68	66.7	16.11
5	89.8	7.46	68.5	21.60
6	91.7	6.00	69.0	21.84
7	94.9	5.88	76.9	19.08
8	95.8	5.96	74.1	15.83

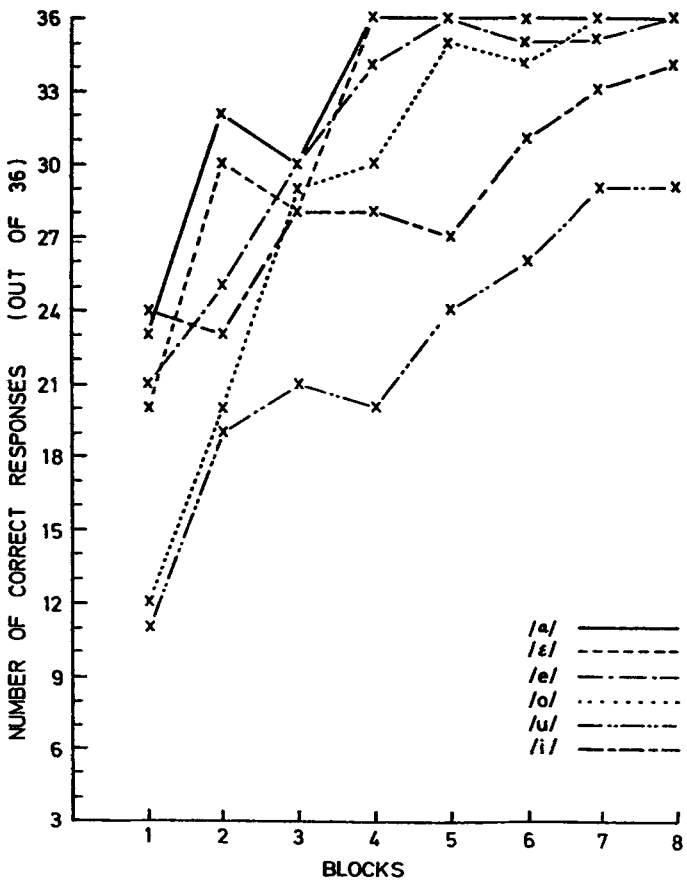

Fig. 2. Discrimination scores of PT Ss for the six tested vowels.

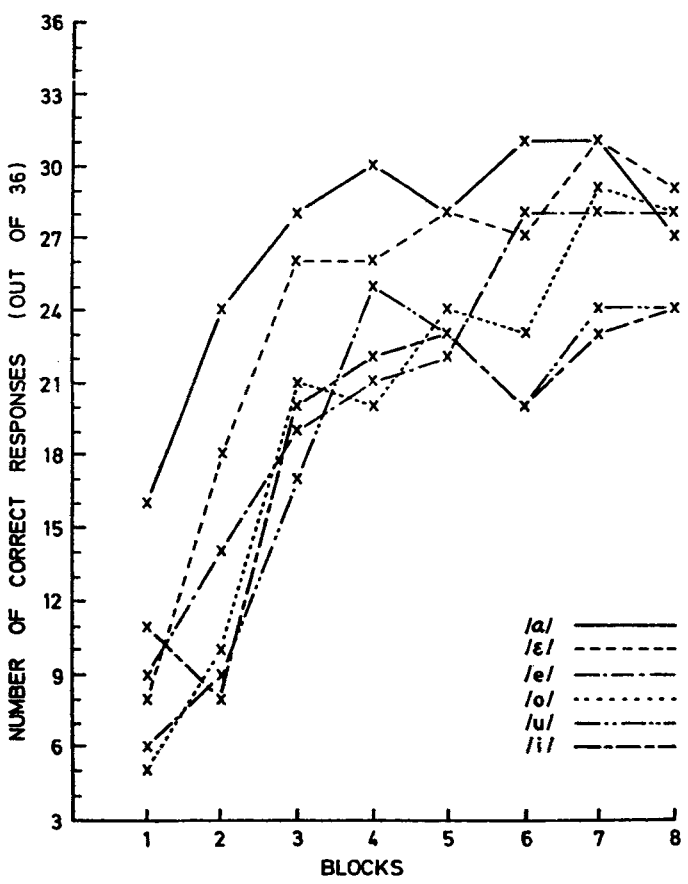

Fig. 3. Discrimination scores of UPT Ss for the six tested vowels.

Table 2
Confusion Matrix of PT Ss for the Last Four Blocks (144 Vowel Stimuli)

Stimulus	Response α	ϵ	e	o	u	i
α	144					
ϵ		144				
e			142	1	1	
o				141	3	
u				3	108	33
i				2	17	125
Total	144	144	142	147	129	158

Table 3
Confusion Matrix of UPT Ss for the Last Four Blocks (144 Vowel Stimuli)

Stimulus	Response α	ϵ	e	o	u	i
α	117	22	5			
ϵ	16	115	10	2	1	
e		9	106	15	7	7
o	1	2	13	103	13	12
u			3	7	91	43
i	3		12	12	26	91
Total	137	148	149	139	138	153

giving them a bonus if they got a good average percent correct discrimination.

RESULTS AND DISCUSSION

The means and standard deviation of the percent correct discrimination scores for both groups of Ss are shown in Table 1. The large standard deviations indicate that initially there was quite a big spread among the test scores of the different Ss. Deviations among the scores of the PT group, however, were not great in later blocks of discrimination learning.

It must be emphasized that, even though the vowels were only 10 msec in length, they did sound like vowels to the PT Ss after several exposures. In fact, some of the vowels, particularly /a/ and /ϵ/, were recognized by most of the PT Ss on first hearing them. During the test, they also mimicked the vowels and tried to map them into their own vowel system. To the UPT Ss, the vowels sounded like clicks.

Large differences were obtained in the performance of the two groups of Ss. The scores of the UPT group ranged from 15% to 25% below those of the PT group. Rapid learning took place in the first four learning blocks, after which the scores rose steadily. The PT Ss approached perfect discrimination of the six vowels, and there were four Ss (including one UPT S) who reached the 100% correct discrimination scores towards the end of the experiment.

An analysis of variance was performed on the discrimination scores shown in Table 1. Significant differences were obtained in learning blocks ($p < .001$) and groups ($p < .05$), but not their interaction ($p > .10$).

Discrimination learning curves for the different vowels are shown in Figs. 2 and 3. The PT Ss could learn the vowels /a/, /ϵ/, /e/, and /o/ to 100% correct discrimination. The UPT Ss also had high scores for these four vowels. However, both groups of Ss had lower scores in /i/ and /u/. Confusion matrices for the last four blocks are shown in Tables 2 and 3, respectively, for the PT and the UPT Ss. These matrices reveal that /i/ and /u/ sound very much alike when they are short. This kind of /i/ and /u/ confusion has also been observed before by Powell and Tosi (1970).

Clarification concerning the scores of the different vowels can be achieved by reference to the perception of short tones. Bürck, Kotowski, and Lichte (described in Stevens & Davis, 1937) found that the duration of a tone required to produce the experience of a definite pitch decreased from about 50 to 11 msec as the frequency of the short tone increased from 50 Hz up to 2 kHz. Beyond 2 kHz, duration increased with frequency. For a tone of 250 Hz, about 20 msec was required to produce a definite pitch and only about 12 msec was required for a tone in the range of 1-4 kHz. Since /i/ and /u/ have the lowest first formant frequencies, 260 and 270 Hz, respectively, their discrimination might be expected to be poorest at 10 msec duration. This is in agreement with the results. Likewise, the first formants of /a/ and /ϵ/ lie in the highest frequency range among the six tested vowels, and their discrimination is best. Both /e/ and /o/ have a first formant frequency lower than that of /a/ and /ϵ/, and their scores were correspondingly lower. This suggests that discrimination of vowels of a very short duration is like the perception of short tones, and, for a short duration, vowels with a higher first formant frequency are better discriminated.

REFERENCES

FLANAGAN, J. L. *Speech analysis synthesis and perception.* New York: Academic Press, 1965. P. 214.

FUJISAKI, H., & KAWASHIMA, T. The influence of various factors on the identification and discrimination of synthetic speech sounds. Paper presented at the 6th International Congress on Acoustics, Tokyo, Japan, 1968.

GRAY, G. W. Phonemic microtomy: The minimum duration of perceptible speech sounds. Speech Monographs, 1942, 9, 75-90.

JOOS, M. Acoustic phonetics. Language Monograph, 1948, 24(No. 2, Suppl. 77-78).

PETERSON, G. E. The significance of various portions of the wave length in the minimum duration necessary for the recognition of vowel sounds. Unpublished doctoral dissertation, Department of Speech, Louisiana State University, 1939.

POWELL, R. L., & TOSI, O. Vowel recognition threshold as a function of temporal segments. Journal of Speech & Hearing Research, 1970, 13, 715-724.

SCHWARTZ, M. F. A study of thresholds of identification for vowels as a function of their duration. Journal of Auditory Research, 1963, 3, 47-52.

SHARF, D. J. Vowel duration in whispered and in normal speech. Language & Speech, 1964, 7, 89-97.

SIEGENTHALER, B. M. A study of the intelligibility of sustained vowels. Quarterly Journal of Speech, 1950, 36, 202-208.

STEVENS, S. S., & DAVIS, H. *Hearing—its psychology and physiology.* New York: Wiley, 1937. P. 102.

SUEN, C. Y., & BEDDOES, M. P. Some applications of a small digital computer in speech processing. Paper presented at the 81st meeting of the Acoustical Society of America, April 20-23, 1971, Washington, D.C.

THOMAS, I. B., HILL, P. B., CARROLL, F. S., & GARCIA, B. Temporal order in the perception of vowels. Journal of the Acoustical Society of America, 1970, 48, 1010-1013.

TIFFANY, W. R. Vowel recognition as a function of duration, frequency modulation and phonetic context. Journal of Speech & Hearing Disorders, 1953, 18, 289-301.

(Accepted for publication January 17, 1972.)