

# An Experimental Study of the Identification of Filtered Japanese Vowels

Akira Hasegawa\*

## PURPOSE

The intimate connection of formant frequencies with the vowel spectrum which can be completely specified by the four lower formant frequencies proved by Fant (11) calls for a very close look at the perception of Japanese vowels from which various formants have been removed by high- and low-pass filtering. The purpose of the present study is to examine 1) the discontinuous functions of per cent correct identification of Japanese vowels versus filter cut-off frequency, 2) difference between the correct identification of vowels of a male speaker and that of a female speaker under controlled conditions, and 3) the effect of filter sharpness upon the identification of filtered vowels.

### *The Problems*

*Functions of per cent correct identification of Japanese vowels versus filter cut-off frequency.* Contrary to the smooth functions, obtained by some earlier workers: e. g., French and Steinberg (14); Ochiai and Fukumura (21); Ochiai (22); Sekiguchi (28); Lehiste and Peterson (18)<sup>1)</sup>, of per cent correct

---

\* Akira Hasegawa (M. A., International Christian University, 1967) is Lecturer of Audio-visual Communication at the Otsuma University for Women. This article is based on his M. A. thesis.

identification of vowels versus filter cut-off frequency, Carterette and Møller (3), who conducted an experiment of the perception of filtered stimuli made from 14 real and 14 synthesized Swedish vowels, found that discontinuities existed in the identification of 10 real vowels distorted by high- and/or low-pass filtering; i. e., in the identification of [tɪ:] (HP), [y:] (HP and LP), [i:] (HP and LP), [e:] (HP and LP), [ɛ:] (HP), [ɛ] (LP), [a] (HP), [ɔ] (LP), [o:] (LP), and [u:] (HP). The existence of discontinuous functions in the identification of Japanese vowels has been merely suggested by Chiba and Kajiyama (6), whose results of phonetic evaluation of change in quality of filtered Japanese vowels show the possible discontinuities to be found in the phonemic evaluation of Japanese [u]'s distorted by high- and low-pass filtering. But such quite few examples of discontinuous functions in identifying filtered Japanese vowels do not seem to be mainly due to the smaller number of different vowels to be discriminated in Japanese language in comparison with those in Swedish, but strongly due to the smaller number of cut-off frequencies in the region 300 to 3000 Hz used by Chiba and Kajiyama (6), because cut-off frequencies are the important determiners of characteristics of the function of per cent correct identification of each vowel versus filter cut-off frequency. It seems necessary, therefore, to examine the discontinuous functions of per cent correct identification of Japanese vowels versus filter cut-off frequency with methodological exactitude.

*Difference between the correct identification of vowels of a male speaker and that of a female speaker under controlled conditions.* Vowels of adult male speakers have been used in the listening tests of filtered vowels by Lehiste and Peterson (18),

- 
- 1) One of the results of their listening tests suggests, however, a discontinuous function in the identification of American English [i] distorted by high-pass filtering.

Palva (24), and others except Ohta, *et al.* (23) who used a professional female announcer of a wireless broadcasting station for their experiment of the identification of Japanese vowels distorted by band-pass filtering. No comparison has been made between the identification of the filtered vowels of a male speaker and those of a female speaker. In spite of some difficulties in controlling variables in real speech in carrying out such an experiment, it seems necessary to test whether the identification of filtered vowels of a male and a female speaker reflects their vowel spectrum envelopes or not.

*Effect of filter sharpness upon the identification of filtered vowels.* If formant frequencies are important determiners of the perception of vowels, it is apparent that the degrees of attenuation of the filters to be used must be a crucial factor. The appropriate degrees of attenuation of the filters in carrying out of an experiment may be determined if the degrees of our frequency selectivity in vowel perception is calculated. Castle (4) assumed by reference to the works of Steinberg (29), Black (2), and Fletcher (13) that frequency information within a vowel structure which is more than 25 dB below the peak amplitude is probably not important to the perception of the vowel, though there has not been any experimental evidence to verify it. According to the works on masking effects by Greenwood (15), Schafer and Thompson (27), and Feldtkeller and Zwicker (12), the analysis of hearing difficulty by Békésy (1), and the studies of the neural responses by Katsuki (17), the overall frequency selectivity of our ears is approximately 80 dB to 120 dB per octave. The value cannot, however, be directly used for the selection of filter sharpness in the perception of filtered vowels, because the value has been obtained from the studies of the perception of some pure tones which are quite different in acoustic structure from those of spoken

vowels. It seems, therefore, at least necessary to carry out the experiment taking the effect of filter sharpness upon the perception of filtered vowels into consideration, and it is even worthwhile to make an attempt to estimate its effect.

### *Hypotheses*

On the basis of the foregoing discussion the author comes to the following working hypotheses for the problems: 1) discontinuous functions of per cent correct identification of Japanese vowels versus filter cut-off frequency exist in more cases than those suggested by Chiba and Kajiyama (6), 2) difference is found to be between the correct identification of vowels of a male speaker and the correct identification of vowels of a female speaker reflecting their vowel spectrum envelopes, and 3) difference is found to be between the identification of vowels distorted by high- and low-pass filtering with three different degrees of attenuation.

## **EXPERIMENT**

### *Recording and Selection of Unfiltered Vowels*

13 native speakers of Japanese, 8 male and 5 female, ranging in age from 19 to 63 years, took part in the recording of the unfiltered vowels. They were two trained phoneticians and their pupils of the English Division of the Department of Education of the Chiba University, who had no history of speech or hearing impairment. Five Japanese vowels /i/, /e/, /a/, /o/, and /u/ were chosen and used because they can be produced in isolated steady state without ambiguity. The recording was carried out in July 1966 at the recording studio of the English Division of the Department of Education of the Chiba University. Each speaker came to the studio individually, and the complete procedure for collecting his vowel samples was accomplished in a

single session.

The speaker was seated in the chamber, and was given the following set of oral instructions in Japanese before the recording was made :

You are asked to record samples of five Japanese vowels /i/, /e/, /a/, /o/, and /u/ in isolation. Certain of these samples will be selected for composing stimuli for a series of listening tests. Your task is to sustain each isolated vowel sample for approximately two seconds in the normal way of phonation with constant vocal effort. The vowels to be spoken are presented to you one at a time through the glass window between the chamber and the control room by the experimenter with flash cards. You are also asked in the production of vowels to maintain the fundamental pitch as close as possible to the pitch of reference sound to which you listen over a monaural earphone. The recording begins after your some practice for adjusting the pitch level to that of the reference signal. You can practice after each change of the pitch level of the signal which will be made three or four times while vowel samples are articulated without difficulty. The pitch level is kept constant during the production of each set of 15 vowels, 3 for each vowel category, which are randomly presented to you. You can freely intermit the reference signal with a switch near at hand. The starts and ends of recording are also led with flash cards.

The reference signal was fed by a Sony Model TC-777A single-channel half-track tape recorder in the control room which played pure tones recorded from an oscillator at a tape speed of 19 cm/sec. While Lehiste and Peterson (18) used recurrent impulse waves at a pitch of 140 Hz as reference signal, continuous single sinusoidal waves were used in this study, which could be intermitted with a switch at the speaker. The pitch levels of the reference signals used were 100 Hz, 150 Hz, 200 Hz and 250 Hz : i. e., 100 Hz, 150 Hz, and 200 Hz for the male speakers, and 150 Hz, 200 Hz, and 250 Hz for the female speakers. There was no time limit for any speaker to do practice for adjusting the pitch level to that of the reference signal. The time spent for it was, in effect, widely different in length from speaker to speaker, which ranged approximately from a quarter of minute to five minutes.

The recording of all vowel samples was accomplished on a system consisting of an Aiwa Model VM-17 vari-directional ribbon microphone and a Sony Model FT-2 single full-track tape recorder operated at 38 cm/sec. Four seven-inch reels of a special brand of low-print magnetic tapes<sup>2)</sup> were used for recording approximately 650 vowel samples of the 13 speakers. An attempt was made to keep the distance from the lips of each speaker to the microphone constant<sup>3)</sup>, approximately 25cm apart. The recording volume control was fixed throughout the recording of vowel samples of each speaker once the recording level was set at the beginning of recording. The vowels in each vowel category of the same speaker were generally recorded within a range of  $\pm 3$  dB in VU reading. But the VU reading in the recording of vowel samples in different vowel categories varied maximally  $\pm 5$  dB to  $\pm 7$  dB. The VU reading of [a] was the highest, that of [i] and [u] was the lowest, and that of [e] and [o] was intermediate between them. The overall frequency response characteristics of the tape recorders are found to be flat from 100 Hz to 7500 Hz ( $\pm 1$  dB), and that of the microphone system is flat from 100 Hz to 8000 Hz ( $\pm 2$  dB).

All the recorded vowel samples were then examined auditorily by the experimenter at the recording studio with a Pioneer Model SE-1 stereo headphone set to eliminate the vowels produced with glottal stops, noticeable nasalization, and/or undesired voice qualities such as creak, whisper, and breathy voice as defined by Catford (5), Crystal and Quirk (7). One male and one female speakers were finally selected, whose five vowel samples [i], [e], [a], [o], and [u] to be used in the experiment showed little auditory difference in fundamental voice frequency. The pitch level of each selected vowel of the male

---

2) Scotch Brand, Type 131-366, Sumitomo 3M Ltd., Tokyo.

3) It was found by Tiffany (30) that this method of controlling intensity is a reliable procedure for recording sustained vowels.

speaker was approximately 150 Hz, and that of each selected vowel of the female speaker was 200 Hz. The onset of each selected vowel sample was marked with ink at the commencement of audible and VU activity from the exposed playback head of the Sony FT-2, and the terminal of the sample was marked with ink exactly 38 cm (1000 msec) beyond the onset ink marker. A cut was then made allowing several centimeters of recorded tape beyond the terminal ink marker. The terminal portions were then removed, and the tape ends were spliced with 190 cm (5 sec) of new timing tapes which were inserted between the segments of unfiltered stimuli to prevent a print effect. A diagonal (45 degree) cut was used throughout the splicing. The spliced master tapes were re-recorded into the six experimental tapes, each of which included 10 unfiltered stimuli, two for each intended vowel category. The transfer was accomplished from the Sony FT-2 to the Sony TC-777A.

### *Formant Frequency Specification*

Formant frequency specification was carried out at the Experimental Phonetics Laboratory of the Tokyo University of Foreign Studies for each unfiltered stimulus which was used in the experiment. A Rion Electric Company sound spectrograph Model SG-04A was used for this purpose. The samples were reproduced on a Sony Model FT-1 tape recorder, using appropriate impedance matching. The first step with each vowel sample was to make a conventional sound spectrogram with the 300 Hz band width, flat frequency response,<sup>4)</sup> and an effective frequency range 85 Hz to 4000 Hz. This was done both to verify the essential constancy of the selected vowel sample and to make it easier to select that portion of the

---

4) It was expressively pointed out by Lindblom (19) that a prefiltering +6 dB/octave gives rise to certain inaccuracies in formant measurements.

signal to be analyzed by the sound spectrograph sectioner. The portion selected for sectioning was always centrally located in the total duration. The sections were made with the 45 Hz band width and flat frequency response,<sup>4)</sup> and were then analyzed by essentially the same methodologies described by Fairbanks and Grubb (10). For specifying formant frequencies for a signal, major regions were determined by delimiting the formant bars in the conventional sound spectrogram. Attention was next directed to those portions of the section which corresponded to those formants. The points of harmonics of which amplitudes were large in these portions of the sections were measured, and the formant frequencies were determined by "the direct evaluation of the frequency positions of the peaks of the vowel envelope curve." (25 : 183) In every case the central frequencies of the first four formants (F 1, F 2, F 3, and F 4) were measured. The results are presented in Tables I and II. The formant frequencies obtained in the present study were compared to the related data in the literature: e.g., Chiba and Kajiyama (6); Doi (8); Nakata (20). The comparisons indicate that the formant frequencies for the vowels of the speakers used in the present study are not atypical for trained Japanese speakers.

Table I  
Formant Frequency Specification of Unfiltered Vowels  
of the Male Speaker Used in the Experiment †

Intended Vowels	F0	Formants			
		F 1	F 2	F 3	F 4
[ i ]	154	300	2145	2930	3710
[ e ]	151	455	1810	2420	3700
[ a ]	153	690	1055	2750	3600
[ u ]	151	300	1210	2135	3785
[ o ]	150	455	875	2530	3450

† Values in Hz.



Table II  
Formant Frequency Specification of Unfiltered Vowels  
of the Female Speaker Used in the Experiment †

Intended Vowels	F0	Formants			
		F 1	F 2	F 3	F 4
[ i ]	206	350	2280	3150	4180
[ e ]	203	575	1350	2800	3980
[ a ]	202	1050	1680	2935	3870
[ u ]	204	425	1625	2740	3985
[ o ]	201	650	955	2780	3845

† Values in Hz.

### *Recording of Filtered Stimuli*

The filtered stimuli were processed individually through the system which includes two tape recorders and a JEIC Model UF-2 Filter set with an approximately 55 dB/octave attenuation. For this experiment 15 high-pass filters and 15 low-pass filters were used. The cut-off frequencies were 180, 224, 280, 355, 450, 560, 710, 900, 1120, 1400, 1800, 2240, 2800, 3550, and 4500 Hz for both high- and low-pass filtering. Difference between the two adjacent filter cut-off frequencies is one third octave of the lower filter cut-off frequency. The effective range from 180 Hz to 4500 Hz is sufficient for considering that frequency information which might be thought important to vowel recognition: i. e., F1 through F4. Different sharpness of filtering was obtained by doing filtering once and filtering repeated once and twice at same cut-off frequencies. As each unfiltered vowel sample was processed separately through each of these 30 filters, tape recording was made of all resultant signals. This was accomplished from the Sony ET-2 to the Sony TC-777A operated at 19 cm/sec through the filter set in the first filtering, and in the second and third filtering from the Sony TC-777A to another Sony TC-777A through the filter

set. In the present study the "gains of the overall system were not readjusted, so that the energy level of any particular formant was altered only according to the filter characteristics employed." (18 : 164) The total number of recorded filtered stimuli was 900 : i. e.,  $5(\text{vowels}) \times 2(\text{speakers}) \times 30(15 \text{ high-pass and } 15 \text{ low-pass filters}) \times 3(\text{sharpness of filtering})$ . 900 filtered stimuli were first divided into six groups according to the speakers and sharpness of filtering. Each group consisted of 150 stimuli made from five vowels of the same speaker distorted by the same sharpness of high- and low-pass filtering. 160 stimuli of each stimulus group including 10 unfiltered stimuli prepared before were finally randomly ordered with pauses between the stimuli. The time interval between the two adjacent stimuli for listeners' response was designed to be 2 seconds on the basis of the results of the preliminary investigation. It was observed in the pilot study that unnecessarily long time interval between stimuli seemed to negatively motivate the subjects to take part in the vowel judgments.

### *Testing of Subjects*

Subjects for the listening tests were 29 male and 19 female adult Japanese ranging in age from 20 to 38 years, who took part in the whole sessions of listening tests carried out in October 1966. The median age was 21 years. 68 prospective subjects were drawn from the two sources : a group of teaching staff, postgraduates, and undergraduates who were or had been associated with the Department of Audio-visual Education of the Graduate School of Education of the International Christian University, and a group of male undergraduates at a student dormitory of the Chiba University. All subjects indicated that they were free from known history of hearing loss or pathology. They were speakers of Japanese since childhood, and most of them were native to the Tokyo area. It was the first time for all

the subjects, who had not been phonetically trained, to take part in such a listening test.

The experiment was conducted in the sound-treated room of the ICU Audio-visual Center and in a quiet room of the student dormitory of the Chiba University. Two subjects were scheduled to attend at a time a session of vowel identification, though dropouts and scheduling conflicts in appointment necessitated testing one subject on some occasions. The subjects were seated, seen individually, at a large table with the experimenter beside them. No difference was attributed to the relative positions at the table in relation to the playback tape recorder which produced some operation noise. The level of recorded stimuli and the effectiveness of seal of the phone sockets were such that this ambient noise level was considered satisfactory for the experiment. Two Pioneer Model SE-1 and Model SE-2 headphone sets were used, of which frequency response characteristics were found to be satisfactorily flat up to 4500 Hz, but showed a drop in sensitivity at 5000 Hz and higher. Both headphone sets were connected to the output of the playback tape recorder put on another table which located three meters away from the subjects. Both phones of each headset were active. Sony Model TC-601 single-channel half-track tape recorder was used at ICU, and Sony Model TC-362B single-channel half-track tape recorder at Chiba.

The oral instructions recorded on the beginning part of the first test tape were first given to the subjects, indicating the stimulus items of the test, the decision-making task involved, and the general testing procedures, which are described as follows:

You are going to hear a series of sounds, each of which you are asked to identify as one of the five Japanese vowels /i/, /e/, /a/, /o/, and /u/. On the answer sheet you are asked to write in the appropriate blank either *hiragana* or *katakana* for the vowel which, in your opinion, each stimulus most closely approximates. A mark of *x* can be used

if you find the decision is impossible to make on the identification of the sound which is not vowel-like, or vowel-like but not identified as one of /i/, /e/, /a/, /o/, and /u/. You are also asked to write one of the figures '1', '2', or '3' printed on the upper margin of the answer sheet by which you estimate the degrees of clearness of the vowel identified. The figure '3' denotes that in your opinion the vowel sound is identified quite clearly, '2' denotes that it is fairly well identified, and the figure '1' indicates that it sounds like the vowel answered. It is of course unnecessary for you to write those figures in case that the identification of the sound is impossible to make, and you use the mark *x* for the stimulus. You can then leave the answer cell blank for the estimation of the degrees of clearness of the vowel to be identified. Your task involves 160 items, which are divided into eight subgroups. In order to be accustomed to the kind of stimuli you are going to respond to, let us have a brief practice session for that you try to answer for five stimuli similar to those in the test.

Five practice sounds were presented to the subjects with two seconds of pauses between the stimuli. After the practice a little more instruction was given on the difference in intensity of stimuli:

Just one thing to remember is mentioned here. That is the degrees of loudness of stimuli will vary considerably from sound to sound as you might have probably noticed in the previous practice. Your careful listening and identification of vowels independent to the loudness of stimuli are, therefore, requested to be made.

The test tape was stopped at the end of the instruction, and some procedural questions were answered. The subjects were then seated comfortably at the table and fitted again with the headsets. The experimenter engaged the playback tape recorder start after an indication that the subjects were ready. The time interval between the end of the instruction and the beginning of the test was generally approximately 5 minutes. The test begun with the carrier sentence, "The test begins now," in Japanese. After the completion of a 160 item stimulus sequence (about 8.5 minutes long), the subjects were given a 8-minute rest. Before another listening test resumed, the answer sheets previously given were collected, another similar sheet was given to each subject, and the experimenter informed

the subject of the stimuli and repeated the condensed instructions described earlier. Two listening tests were conducted in each of three sessions, and total testing time for each session was about 25 minutes. Before leaving, the subjects were asked not to disclose the purpose or procedure of the experiment to other known prospective participants.

## RESULTS

46080 responses were made by 48 subjects for 960 stimuli which consisted of 60 unfiltered and 900 filtered stimuli. Chi square test was adopted for treating the data.

### *Identification of Unfiltered Vowels*

The responses for the unfiltered stimuli intended for the five Japanese vowels are presented in Tables III and IV. There is no statistically significant difference in per cent correct between the identification of the vowels intended for /i/, /e/, /a/, and /u/, which had high identifiability (per cent correct: 100 to 96). The results of the estimation of clearness of correctly identified vowels also show that the unfiltered vowels were identified quite clearly by most of the subjects and that there were very few listeners thought that the identified vowels were poor in clarity with respect to representatives of Japanese vowels. Median rating score of the estimation of clearness of each correctly identified vowel was always '3' except '2' for one of [e]'s presented in the second listening test. Some statistically significant differences were obtained, however, between the estimation of clearness of the correctly identified vowels in the same vowel category in the same listening test: [i] (II), [e] (I and II), and [o] (I and III) of the male speaker; [e] (VI) and [o] (V) of the female speaker, though the per cent correct identification of the vowels in the same vowel category

Table III  
Responses for Unfiltered Vowels of the Male Speaker

Stimulus Groups (Test Tape Nos.)	Order of Presentation	Intended Vowels	Responses						Per Cent Correct
			/i/	/e/	/a/	/u/	/o/	x	
I	2	[ i ]	48 48						100 100
		[ e ]		48 47				1	100 98
		[ a ]			47 48			1	98 100
		[ u ]				48 48			100 100
		[ o ]			2 7		45 38	1 3	94 80
II	3	[ i ]	48 48						100 100
		[ e ]		48 48					100 100
		[ a ]			47 46			1 1	98 96
		[ u ]				48 48			100 100
		[ o ]			1 7		47 39	2	98 81
III	1	[ i ]	48 48						100 100
		[ e ]		48 48					100 100
		[ a ]			48 47			1	100 98
		[ u ]				47 48	1		98 100
		[ o ]			2 1		46 39	1	96 81

did not show any significant difference.

The correct identification of vowels intended for /o/ is somewhat lower than that of the other four vowels. The per cent correct identification of unfiltered [o] of the male speaker ranges from 98 per cent to 80 per cent, and that of the one female speaker ranges 85 per cent to 46 per cent. The results of the chi square tests indicate that there is no statistically significant difference between the correct identification of [o] of the male speaker and the lowest per cent (96 per cent) correct response

Table IV  
Responses for Unfiltered Vowels of the Female Speaker

Stimulus Groups (Test Tape Nos.)	Order of Presentation	Intended Vowels	Responses						Per Cent Correct
			/i/	/e/	/a/	/u/	/o/	x	
IV	4	[ i ]	48 48						100 100
		[ e ]		48 48					100 100
		[ a ]			48 48				100 100
		[ u ]				48 47			100 98
		[ o ]			6 19		41 29	1	85 60
V	6	[ i ]	48 48						100 100
		[ e ]		48 48					100 100
		[ a ]			48 48				100 100
		[ u ]			1	47 48			98 100
		[ o ]			16 22		32 25	1	67 52
VI	5	[ i ]	48 48						100 100
		[ e ]		47 47				1	98 98
		[ a ]			47 48			1	98 100
		[ u ]				48 48			100 100
		[ o ]			11 25		37 22	1	77 46

for the other four vowels of the male and female speakers. But there are differences ( $X^2 \geq 5.7$ ,  $p < .05$ ) between the lowest correct response for the other four vowels and the correct response for each [o] of the female speaker except the one of [o]'s presented in the fourth listening test. Significant differences in per cent correct were also obtained between the identification of the [o]'s presented in the same listening test: (II,  $X^2 = 5.5$ ,  $p < .05$ ; IV,  $X^2 = 7.6$ ,  $p < .01$ ; VI,  $X^2 = 24.4$ ,  $p < .001$ ). The correct identification of each [o] of the male and female speakers was then compared. In 23 out of 36 cases was

the vowel of the male speaker identified significantly better than that of the female speaker. Unfiltered [o] of the female speaker was confused 102 times out of which 99 was misidentified as /a/<sup>5)</sup>, and only 3 listeners found it difficult to identify the stimulus as one of the five vowels. Median rating score of the estimation of clearness of correctly identified /o/ and the one misidentified as /a/ in the identification of [o] of the female speaker was, however, consistently '3' and '2' respectively.

To sum up, the unfiltered vowels used in the present experiment were well identified as the intended vowels except the vowel samples of [o] of the female speaker which had significantly lower identifiability than the others.

### *Identification of Filtered Vowels*

Per cent correct identification of each [i], [e], [a], [u] and [o] of the male and female speakers distorted by high- and low-pass filtering is presented in Figures 1 to 20 on pages 227-233, and mean per cent correct identification of the five vowels of each speaker is presented in Figures 21 to 26 on pages 233-235.

*Discontinuous functions of per cent correct identification of the vowels versus filter cut-off frequency.* Significant discontinuous functions of per cent correct identification of the vowels versus filter cut-off frequency were found to be in the identification of [i], [e], [a], and [u] of the male speaker, and [i], [u], and [o] of the female speaker, which were distorted by high-pass filtering, and in the identification of [u] and [o] of the male speaker and [i], [e], [u], and [o] of the female speaker, which were distorted by low-pass filtering. The

---

5) Peterson and Barney (26) and Fairbanks and Grubb (10) have reported similar results of the American English vowel [ɔ] misidentified as [ɑ].



results of chi square tests are presented in Tables V and VI. The *discontinuous functions* appear as follows :

[i] (HP): nearly 100 per cent correct (at 1800 Hz) identification of the vowel of the male speaker drops to less than 30 per cent at 450-560 Hz, and then rises to reach 94 per cent and more at 280 Hz, and the high per cent correct (96 to 90 per cent at 1800 Hz) identification of the vowel of the female speaker also drops to approximately 40 per cent at 560-900 Hz, and rises to reach 92 per cent and more at 280 Hz.

[i] (LP): the correct identification of the vowel of the female speaker starts from approximately 17 per cent at 355 Hz, drops to 2 per cent at 1120 Hz, and then rises to reach 88 per cent and more at 2800 Hz.

[e] (HP): the correct identification of the vowel of the male speaker gradually rises from nearly 0 per cent at 2800 Hz, and reaches 98 per cent and more at 710 Hz, but it decreases to 50-69 per cent at 900-1120 Hz, then it rises again to reach 98 per cent and more at 560 Hz.

[e] (LP): the correct identification of the vowel of the female speaker begins to rise from 10 per cent and less at 560 Hz, reaches 31-40 per cent at 710-1120 Hz, then it drops to 4-19 per cent, rises again, and reaches 90 per cent and more at 2240 Hz.

[a] (HP): the correct identification of the vowel of the malespeaker starts from 1800 Hz at which 4 per cent and less subjects identified the vowel correctly, rises gradually, and reaches 100 per cent at 710 Hz, but it drops to 77-88 per cent at 450 Hz, and again rises to 96 per cent and more around the cut-off frequencies which are lower than 355 Hz.

[u] (HP): the correct identification of the vowel of the male speaker starts to rise from 0 per cent at 2240 Hz, reaches 35-48 per cent at 1400 Hz, then drops to 4-19 per cent at 710 Hz, and again rises gradually to reach 85 per cent and more at 355 Hz, while the correct identification of the vowel of the female speaker begins to rise gradually from 4 per cent and less at 2240 Hz, reaches 58-71 per cent at 1400-1800 Hz, then drops to 25-35 per cent at 710-900 Hz, and again rises gradually to reach 100 per cent at 450 Hz.

[u] (LP): the correct identification of the vowel of the male speaker begins to rise from 42-56 per cent at 224 Hz, reaches 90 per cent and more at 355 Hz, drops to 67-75 per cent at 560 Hz, rises up again to 88-94 per cent, drops a little to 65-92 per cent at 900 Hz, and rises again to 90 per cent and more at 1120 Hz, while the correct identification of the vowel of the female speaker rises from 46 per cent and

Table V

Discontinuous Functions of Per Cent Correct Identification of Vowels  
of the Male Speaker versus Filter Cut-Off Frequency †

Filtered Vowels		Cut-off Frequencies in Hz								
		355	450	560	710	900	1120	1400	1800	2240
[ i ] HP	1	69***27    **    60 ** 88 * 67***96								
	2	88	***	27***65	**	92 ** 67***98				
	3	52	***	13***80 *	58***90 *	71***99				
[ e ] HP	1	90 ** 67    *    86    86 * 63								
	2	98 ** 73    63 * 83    N.S.    71*** 2								
	3	100***54    50***83    **    58*** 2								
[ a ] HP	1	98	N.S.	88	N.S.	100	N.S.	88*** 4		
	2	96	*	78	*	100	N.S.	92***27		
	3	98	*	80	*	100	N.S.	92*** 4		
[ u ] HP	1	86    ***    19       **    44    ***    0								
	2	67    ***    13       ***    48       ***    4								
	3	46    ***    4       ***    35       ***    2								
[ u ] LP	1	48***90    *    71       *       90								
	2	52***90    N.S.    75    *    92								
	3	42***100    ***    67 ** 90 ** 65 ** 90								
[ o ] LP	1	46       ***       81 ** 48       **       81								
	2	2       ***       88N.S.71       *       92								
	3	6       ***       81       N.S.    71    *    88								

† Values in per cent correct.

Filtering: 1, once.

2, twice.

3, three times.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .



less at 180 Hz, the lowest cut-off frequency which is below the fundamental voice frequency of the vowel, reaches 73-79 per cent at 355-450 Hz, drops to 46-65 per cent at 560 Hz, rises to 60-81 per cent at 710 Hz, then drops again to 52-60 per cent at 900 Hz, and rises again gradually to reach 92 per cent and more at 1120 Hz.

[o] (HP): the correct identification of the vowel of the female speaker begins to rise gradually from 2 per cent and less at 1800 Hz, reaches 85-92 per cent at 355 Hz, but it drops to 54-60 per cent at the lowest cut-off frequency, 180 Hz.

[o] (LP): the correct identification of the vowel of the male speaker rises gradually from 6 per cent and less at 280 Hz, reaches 81-88 per cent at 900 Hz, then drops with several variations, one of which is statistically significant, while the correct identification of the vowel of the female speaker begins to rise at 280 Hz, where the 4 per cent of the listeners correctly identified the vowel, reaches approximately 60 per cent at 560 Hz, 900 Hz, and 1120 Hz, then drops to 25-38 per cent at 1800 Hz, and it rises again to 48-58 per cent at 2800 Hz. It should be noted that quite low per cent of correct identification of the vowel of the female speaker was obtained under the whole 15 cut-off frequency conditions.

*Significant differences in per cent correct between the identification of the vowels of the male speaker and those of the female speaker.* Statistically significant differences in per cent correct were obtained between the identification of each of the five vowels of the male speaker and that of the female speaker. Also was made the comparison between mean per cent correct identification of the five vowels of the male speaker and that of the female speaker. The results of chi square tests are presented in Tables VII and VIII and are summarized as follows:

[i] and [e]: the vowels of the female speaker were identified significantly better than those of the male speaker around the lower cut-off frequencies under both high-pass and low-pass filtering conditions ([i], 355, 450, 560, 710 Hz in HP, 355 Hz in LP; [e], 710, 900 Hz in HP, 710, 900, 1120 Hz in LP), while the vowels of the male speaker were identified better than those of the female speaker around the higher cut-off frequencies ([i], 710, 900, 1120 Hz in HP, 1800 Hz in LP; [e], 1120, 1400, 1800 Hz in HP, 1400, 1800 Hz in LP).

[a] and [u]: the vowels of the female speaker were identified better than those of the male speaker under high-pass filtering conditions ([a], 450, 1400, 1800 Hz; [u], 450, 560, 710, 900, 1120, 1400, 1800 Hz), but under low-pass filtering conditions were the vowels of the male speaker identified better than those of the female speaker ([a], 710

Table VII

Significant Differences between the Identification of the Vowels of the Male and Female Speakers Distorted by High-pass Filtering †

Intended Vowels	Male > Female			Female > Male		
	1	2	3	1	2	3
[ i ]		710* 900* 1120*	710* 1120*	450* 560*** 710***	450*** 560*	355*** 450*** 560**
[ e ]	1400* 1800**	1120* 1400* 1800*	1120*** 1400* 1800*	900**	710*	710***
[ a ]				1800***	450** 1400***	450* 1400***
[ u ]				450* 560*** 710** 1800***	450*** 560** 710**	450*** 560** 710** 900* 1120* 1400**
[ o ]	180*** 224* 280**	180***	180*** 224*	710*** 900*	710***	560** 710***
All Combined				450* 560*** 710***	450*** 560* 710**	355*** 450*** 560*** 710**

† Values in cut-off frequencies.

Filtering: 1, once.

2, twice.

3, three times.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

Table VIII

Significant Differences between the Identification of the Vowels of the Male and Female Speakers Distorted by Low-pass Filtering †

Intended Vowels	Male > Female			Female > Male		
	1	2	3	1	2	3
[ i ]	1800***			355*		355*
[ e ]	1400*** 1800*	1400* 1800***	1800***	710** 1120**	710** 900**	710*** 900** 1120*
[ a ]	710***	710***	710*			
[ u ]	900*	180*** 355* 560** 710*** 900***	180* 355**			
[ o ]	355*** 450** 710** 900** 1400*** 1800*** 2240*** 2800*** 3550* 4500***	450* 900** 1120* 1400*** 1800*** 2240*** 2800*** 3550*** 4500***	450*** 560*** 710*** 900*** 1120* 1400*** 1800** 2240** 2800** 3550*** 4500***			
All Combined	1400*** 1800*** 2800*** 4500**	560* 710* 1400** 1800*** 2240** 2800*** 3550** 4500***	450* 1800*** 2240* 3550** 4500**			

† Values in cut-off frequencies.

Filtering: 1, once.

2, twice.

3, three times.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

Hz; [u], 180, 355, 560, 710, 900 Hz).

[o]: the vowel of the male speaker was identified better than that of the female speaker around the lowest cut-off frequencies under high-

pass filtering conditions and was also identified better at almost all cut-off frequencies under low-pass filtering conditions (180, 224, 280 Hz in HP, 355, 450, 560, 710, 900, 1120, 1400, 1800, 2240, 2800, 3500, 4500 Hz in LP), while the vowel of the female speaker was identified better than that of the male speaker around some cut-off frequencies under high-pass conditions (560, 710, 900 Hz).

All combined: mean per cent correct identification of the five vowels of the female speaker was consistently higher than that of the vowels of the male speaker around the lower cut-off frequencies under high-pass filtering conditions (355, 450, 560, 710 Hz), while under low-pass filtering conditions was the mean per cent correct identification of the vowels of the male speaker consistently higher than that of the vowels of the female speaker around the lower cut-off frequencies (450, 560, 710 Hz) and around almost all higher cut-off frequencies (1400, 1800, 2240, 2800, 3550, 4500 Hz).

The consistently lower mean per cent correct identification of the five vowels of the female speaker than that of the vowels of the male speaker around the higher cut-off frequencies under low-pass conditions is, however, largely due to the much lower identifiability of [o] of the female speaker distorted by low-pass filtering than that of the male speaker around the cut-off frequency 1400 Hz and above. It should be, therefore, more reasonable to consider that the important differences in mean per cent correct were found to be around the lower filter cut-off frequencies (450, 560, 710 Hz) under both high-pass and low-pass filtering conditions.

*Effects of filter sharpness upon the identification of the vowels.* Significant differences were obtained between the identification of the vowels distorted by different sharpness of filtering as shown in Table IX. Differences were found to be in the filter cut-off frequency region 355 Hz to 2800 Hz where the variation of the correct identification was marked. Effects of filter sharpness upon the correct identification of the filtered vowels were consistent with the degrees of attenuation of the filters used. The correct identification of the vowels distorted

Table IX

Significant Differences between the Identification of the Vowels  
Distorted by Different Sharpness of Filtering †

Speakers and Intended Vowels		High-pass Filtering			Low-pass Filtering		
		1 vs. 2	2 vs. 3	1 vs. 3	1 vs. 2	2 vs. 3	1 vs. 3
Male	[ i ]	450**	355*** 450**	355*** 450*** 710***	1800* 2240**	2800*	1800*** 2240*** 2800*
	[ e ]	710* 2240*		710*** 1800** 2240**	1400***	1400***	1120* 1400***
	[ a ]	1400***	1400***	1400***		710* 900***	560** 710** 900***
	[ u ]	450*	450*	355* 450*** 560*			
	[ o ]	560**		560*** 710*	355*** 1120*	560*	355*** 1120**
Female	[ i ]	355* 560*		450* 560***	2240***	2240**	2240***
	[ e ]	2240***		900* 1800* 2240***	900* 1800*	1400* 1800*	1800***
	[ a ]	1400** 1800***		1400*** 1800***		900*	900*
	[ u ]	560*** 1800***		560*** 1800***	710*	710*	
	[ o ]			560* 710* 900* 1400*	560**	2240*	560**

† Values in Hz.

Filtering: 1, once.

\*  $p < .05$ .

2, twice.

\*\*  $p < .01$ .

3, three times.

\*\*\*  $p < .001$ .



by sharper filtering was usually significantly lower than the correct identification of the vowels distorted by less sharp filtering as shown in Figures 3, 4, 6, 7, 9, 10, 11, 13, 17, and 18. Sharper filtering sometimes indicated the discontinuous functions of per cent correct identification of the vowels versus filter cut-off frequency more clearly than the less sharp filtering did as shown in Figures 1, 2, 5, 8, and 9.

*Some results of the estimation of clearness of correctly identified vowels distorted by high- and low-pass filtering.*

There were statistically significant differences between the estimation of clearness of filtered vowels which were correctly identified by almost all the subjects while their per cent correct identification (100 per cent to 88 per cent) shows no significant difference between them. Such results verify that the vowel spectra around those filter cut-off frequencies still contributed to the clearness and/or naturalness of the vowels as Japanese vowels. The results obtained are summarized as follows:

[i] (HP): all the estimation of clearness of the vowels of the male and female speakers distorted at 224 Hz was significantly higher than that of the vowels of the speakers distorted at 280 Hz. The cut-off frequency region 224 Hz to 280 Hz includes the spectrum portions just below the centres of the first formants of the vowels of the speakers.

[i] (LP): the estimation of clearness of the vowels of both speakers distorted at 3550 Hz was consistently higher than that of the vowels distorted at 2800 Hz. The frequency region 2800 Hz to 3550 Hz includes the third formants of the vowels of the speakers.

[e] (LP): the estimation of clearness of the vowel of the male speaker distorted at 2240 Hz and that of the vowel of the female speaker distorted at 2800 Hz was consistently higher than that of the vowel of the male speaker distorted at 1800 Hz and that of the vowel of the female speaker distorted at 2240 Hz respectively. The vowel spectrum of the male speaker in the frequency region 1800 Hz to 2240 Hz and that of the female speaker in the region 2240 Hz to 2800 Hz were, therefore, active in contributing to the estimation of clearness of

correctly identified /e/. The spectrum portions in those cut-off frequency regions are the ones which are between the centres of F2 and F3 of each of the vowels of the speakers.

[a] (HP): the estimation of clearness of correctly identified [a] of the male and female speakers at the cut-off frequency of 450 Hz was significantly lower than the estimation at each adjacent cut-off frequency. The results imply that the *discontinuous functions* are at least latent in the identification of [a] of both the male and female speakers, though per cent correct identification does not always show a significant drop at the cut-off frequency, 450 Hz. Significant differences were also found to be between the estimation of clearness of the vowels of the male and female speakers distorted at 450 Hz and 710 Hz and at 710 Hz and 1120 Hz. The spectrum portion in the frequency region 450 Hz to 1120 Hz is the one which includes F1 of each of the vowels of the speakers. The results indicate that the spectral shape below the central frequency of F1 of each of vowels still contributed to the evaluation of the quality of the vowels.

[a] (LP): the estimation of clearness of the vowel of the male speaker distorted at 1400 Hz and that of the vowel of the female speaker distorted at 1400 Hz was significantly higher than the estimation of clearness of the vowel of the male speaker distorted at 900 Hz and that of the vowel of the female speaker distorted at 900 Hz and/or 1120 Hz respectively. The spectrum portions in the frequency region 900 Hz to 1400 Hz consist of the second formants of the vowels.

[u] (HP): the estimation of clearness of correctly identified [u] of the male and female speakers distorted at 280 Hz was significantly higher than that of the vowels of the speakers distorted at 355 Hz and/or 450 Hz. The frequency region 280 Hz to 355 Hz or to 450 Hz consists of the vowel spectrum portion around the first formant of each of the vowels of the speakers.

The results of the estimation of clearness of correctly identified [u] distorted by low-pass filtering are rather complicated, and it is difficult for the present author at the moment to find a simple consistent tendency.

### CONCLUDING REMARKS

The three hypotheses previously made were all accepted.

Significant discontinuous functions of per cent correct identification of the vowels versus filter cut-off frequency obtained in the identification of all the five Japanese vowels have verified the first hypothesis of the present study. The results include the *discontinuous functions* in the identification of [u] distorted by high- and low-pass filtering suggested by Chiba and Kajiyama (6), indicating that the cut-off frequency regions where *discontinuities* appeared are essentially in accordance with those described by the authors. The fact that in spite of some differences in acoustic structure of stimuli and filter sharpness between the experiments done by Carterette and Møller (3) and by the present author there is a close resemblance between the results of the experiments concerning the appearance of *discontinuities* in the identification of vowels suggests that those discontinuous functions of per cent correct identification of vowels versus filter cut-off frequency are universal probably mainly due to the physiological function of our auditory organs which supplement it by learning.

The second hypothesis has been verified by the significant differences between the identification of the vowels of the male speaker and the identification of those of the female speaker. In general, the vowels of the male speaker were identified better than those of the female speaker under low-pass filtering conditions and the vowels of the female speaker were identified better than those of the male speaker under high-pass filtering conditions. In this experiment the lower central formant frequencies of F1 and F2 of the vowels of the male speaker made it easier to identify them under low-pass conditions than those of the female speaker did, while the higher central frequencies of the first two formants of the vowels of the female speaker made it easier to identify the vowels of the female speaker under high-pass conditions than those of the male speaker did. The results of mean per cent correct identifi-

cation of the vowels of the male and female speakers obtained in the present experiment do not conflict with those reported by Epstein and Ulrich (9) who used short passages recorded by male and female adult Americans.

Some effects of filter sharpness were demonstrated comparatively easily, though the exact calculation of the frequency selectivity in vowel perception is left for future investigations. The limitation of the present experiment is that the obtained cut-off frequencies of three different degrees of attenuation are not rigorously the same throughout under these three filter sharpness conditions. The filter cut-off frequency defined as 3 dB point below the highest frequency response usually moves positively to a certain extent in frequency under high-pass conditions when a filtered sample is filtered again with the same high-pass filter, and it moves negatively in frequency under low-pass conditions when the repetition of filtering goes on, owing to general characteristics of gliding decrease in frequency response around the cut-off frequencies of the filters. Some deviation of the cut-off frequency from the one obtained in the first filtering must be, therefore, crucial when the cut-off frequency is in the frequency region where the percent correct identification of vowels changes in a greater degree from cut-off frequency to cut-off frequency. The limitation of controlling filter cut-off frequencies makes it unimportant to calculate the precise frequency selectivity in the perception of vowels in this experiment. Another difficulty of controlling experimental conditions in carrying out an experiment of the effect of filter sharpness upon the vowel perception such as this is the problem of formant band widths. There are some contradictions between designing the different sharpness of filtering and controlling the formant band widths under the same cut-off frequency condition. Designing different sharpness of filtering sometimes introduces simultaneously the

change of the formant band width at the cut-off frequency where the effect of filter sharpness is to be calculated. There are variations in the most preferred formant band widths which correlate positively with known characteristics of vowel articulation as reported by House (16). The difference in per cent correct between the identification of vowels distorted by different sharpness of filtering may, therefore, include both the effect of filter sharpness and the effect of formant band width upon the perception of vowels. The future investigations of frequency selectivity in vowel perception might be, in this respect, carried out also by the analysis-by-synthesis techniques, in which the results of the experiment of real vowels distorted under a large number of filter cut-off frequency conditions and the results of the experiment of synthesized vowels are to be compared.

### ACKNOWLEDGEMENTS

The author wishes to express his thanks to Professors Terumi Nakano and Mitoji Nishimoto of the International Christian University for giving generously the helpful advice and fruitful discussion on every phase of the study. Thanks are also due to Mr. Yoshio Takeuchi of the Faculty of Medicine of the Kyoto University, Professor Kunihiro Imai of the Tokyo Metropolitan University, and Professor Hiroshi Suzuki of the Chiba University, from whom the author has gained many valuable ideas which he has drawn on freely in the preparation of the experiment. The kind permission of Professor Kazuo Amano of the Chiba University to use the recording studio and the language laboratory of his English Division in the preparation of the stimuli used in the experiment, and the permission of Mr. Kazuyoshi Akiyama of the Tokyo University of Foreign Studies to use the Rion sound spectrograph for the formant measurement of the vowels used in the experiment are also acknowledged with gratitude.

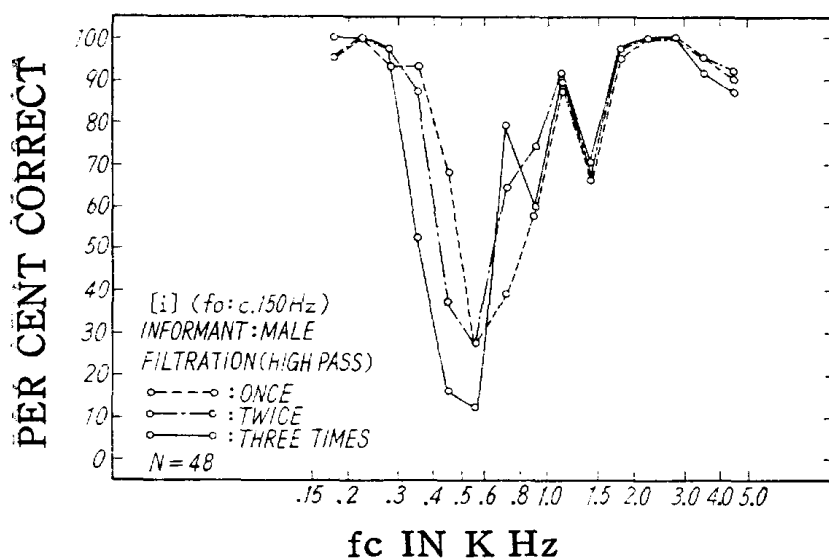


Figure 1. Per cent correct identification of [i] of the male speaker distorted by high-pass filtering.

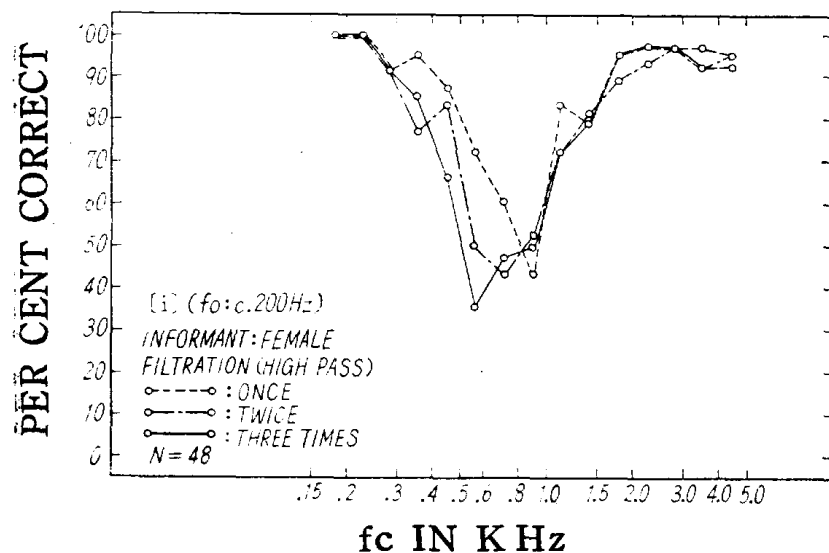


Figure 2. Per cent correct identification of [i] of the female speaker distorted by high-pass filtering.

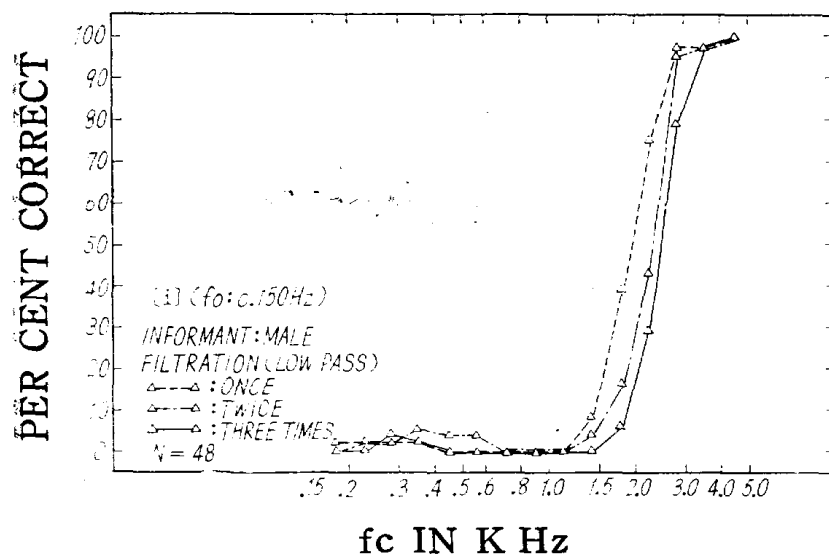


Figure 3. Per cent correct identification of [i] of the male speaker distorted by low-pass filtering.

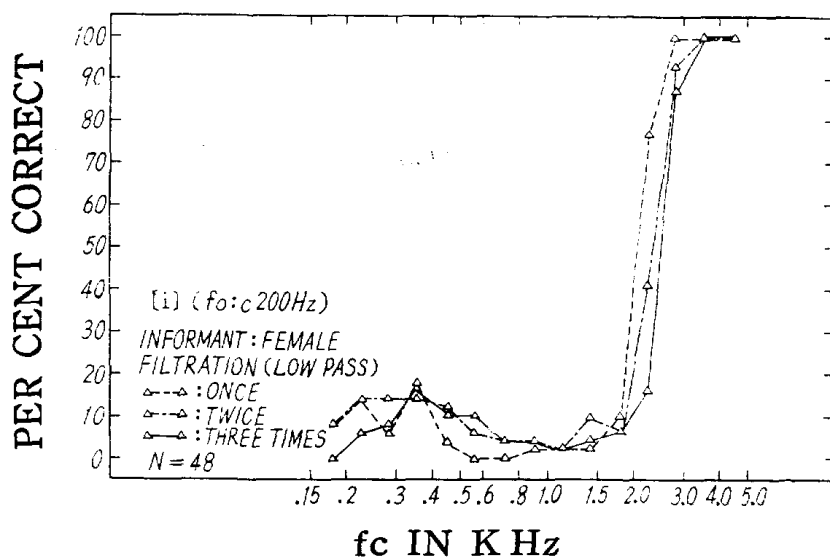


Figure 4. Per cent correct identification of [i] of the female speaker distorted by low-pass filtering.

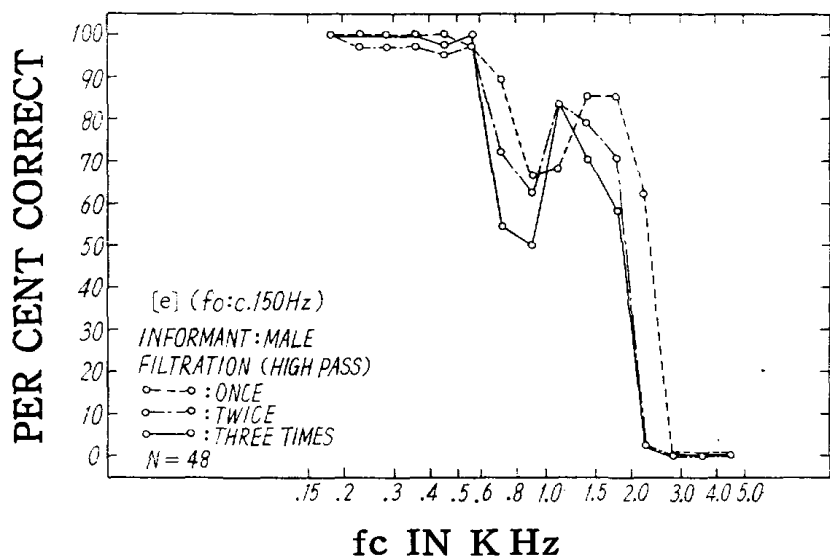


Figure 5. Per cent correct identification of [e] of the male speaker distorted by high-pass filtering.

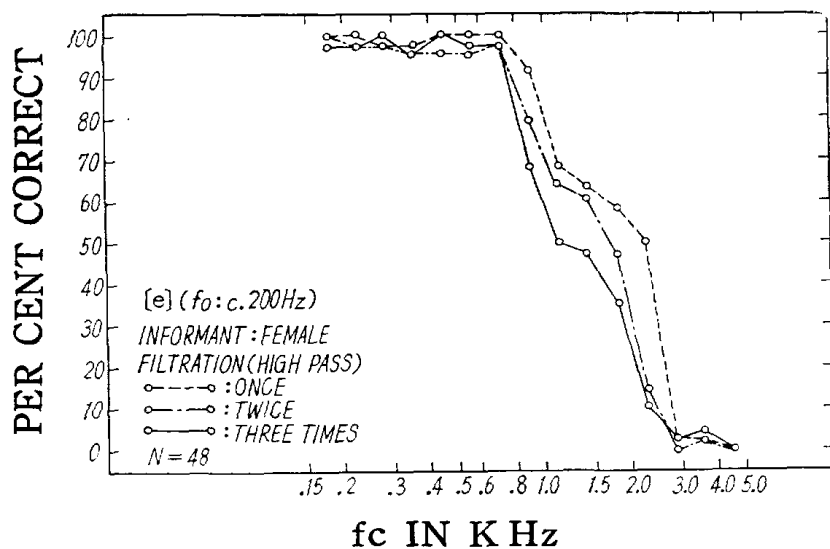


Figure 6. Per cent correct identification of [e] of the female speaker distorted by high-pass filtering.

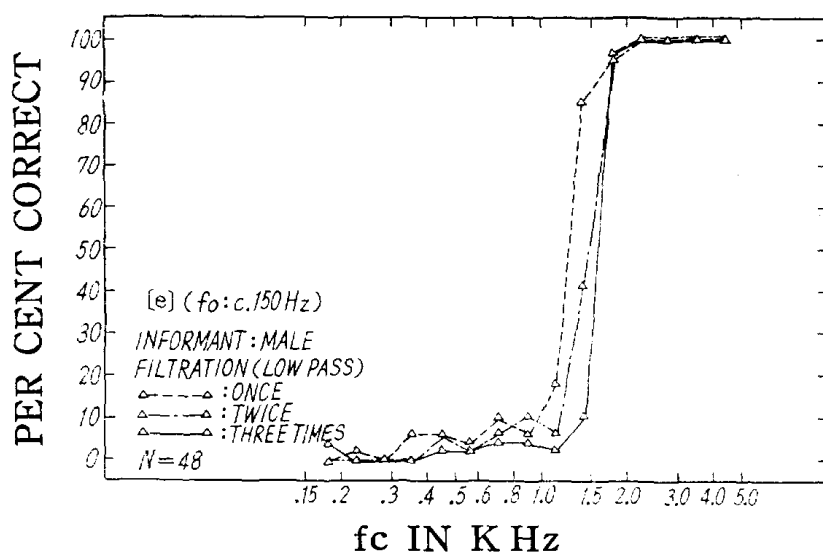


Figure 7. Per cent correct identification of [e] of the male speaker distorted by low-pass filtering.

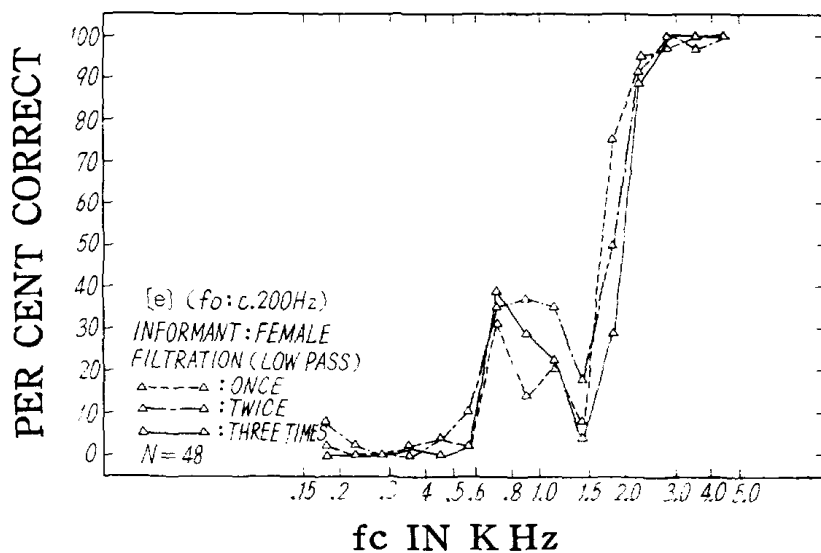


Figure 8. Per cent correct identification of [e] of the female speaker distorted by low-pass filtering.

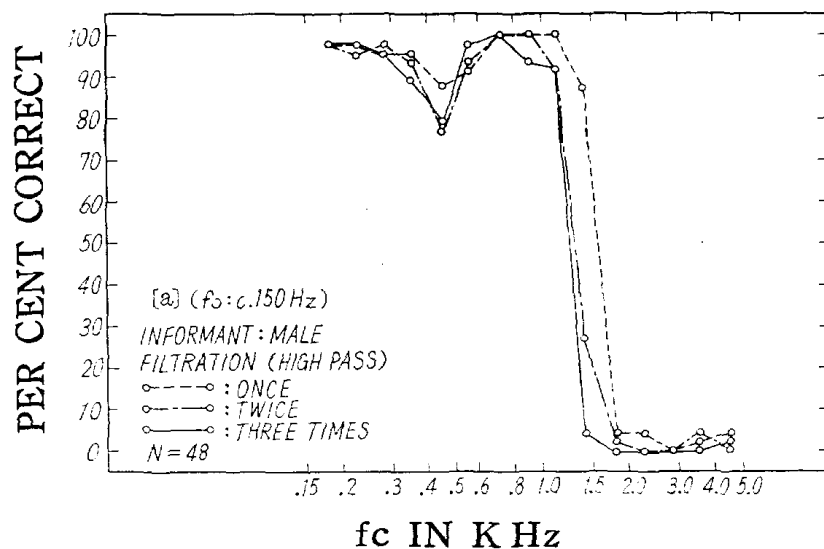


Figure 9. Per cent correct identification of [a] of the male speaker distorted by high-pass filtering.



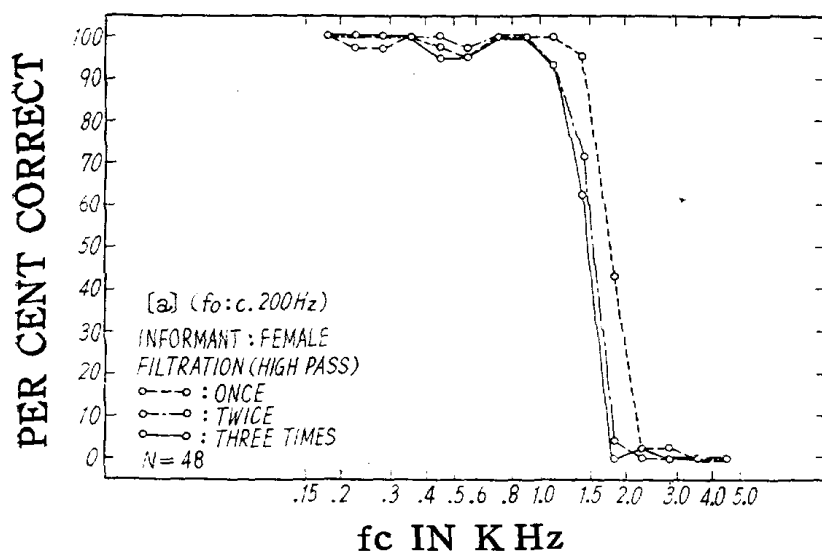


Figure 10. Per cent correct identification of [a] of the female speaker distorted by high-pass filtering.

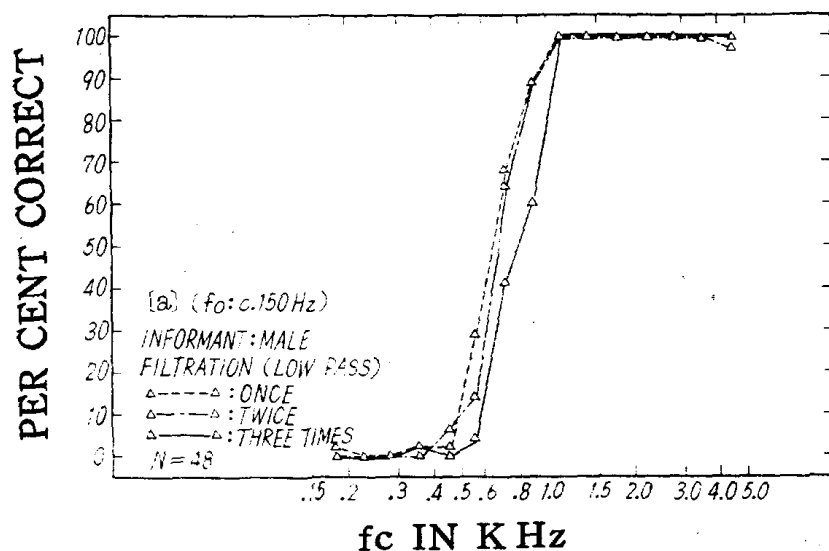


Figure 11. Per cent correct identification of [a] of the male speaker distorted by low-pass filtering.

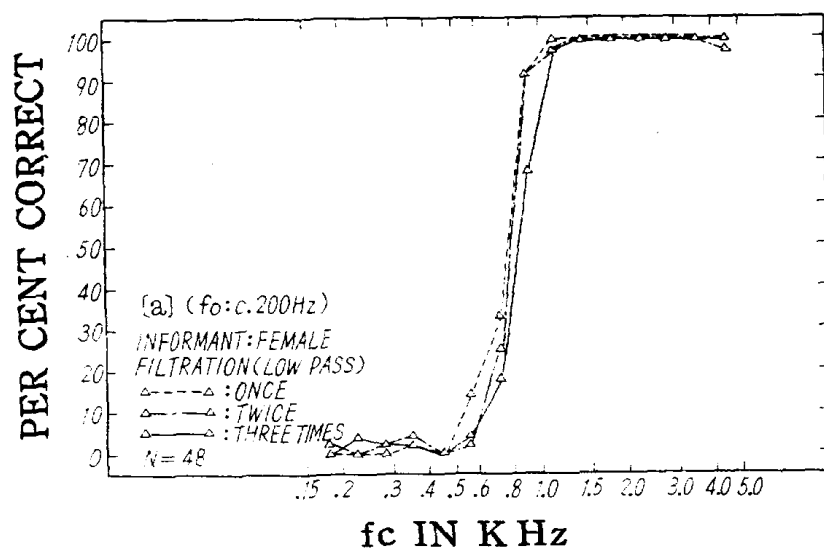


Figure 12. Per cent correct identification of [a] of the female speaker distorted by low-pass filtering.

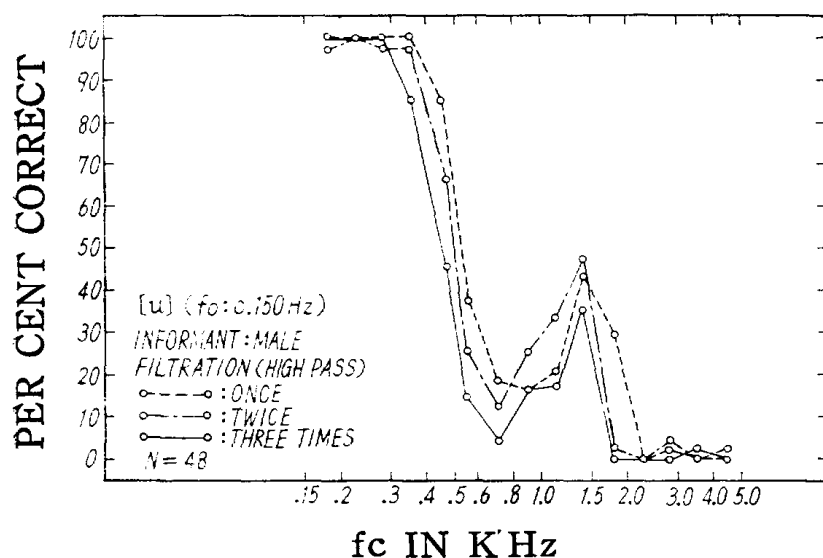


Figure 13. Per cent correct identification of [u] of the male speaker distorted by high-pass filtering.

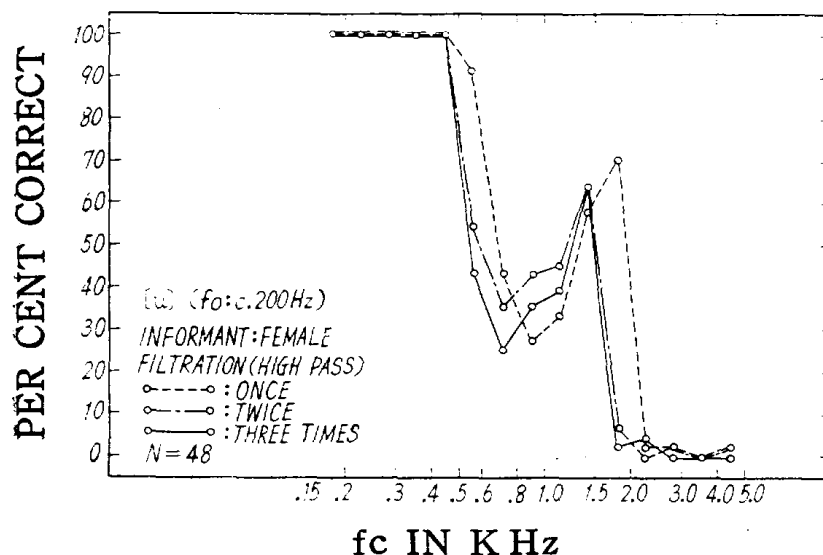


Figure 14. Per cent correct identification of [u] of the female speaker distorted by high-pass filtering.

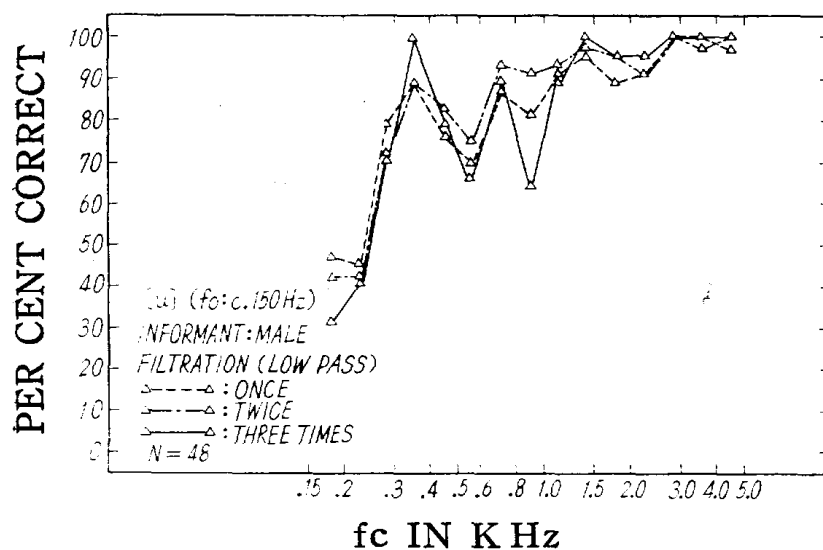


Figure 15. Per cent correct identification of [u] of the male speaker distorted by low-pass filtering.

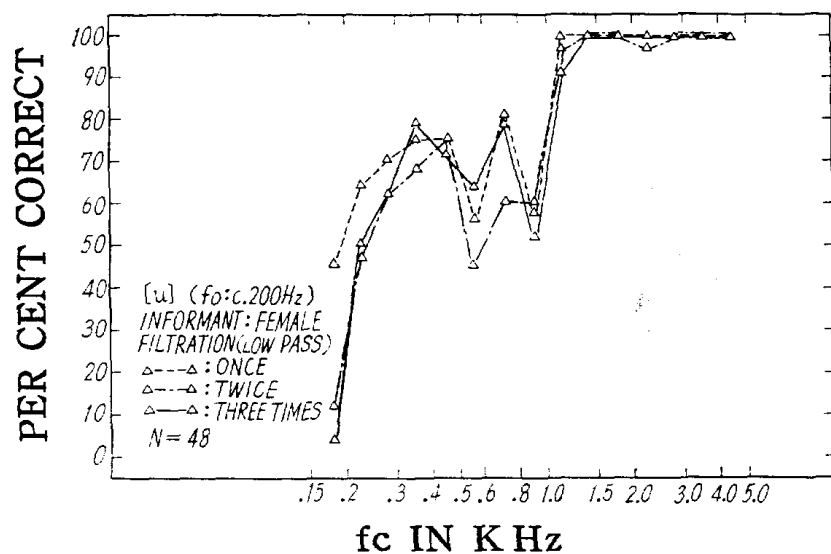


Figure 16. Per cent correct identification of [u] of the female speaker distorted by low-pass filtering.

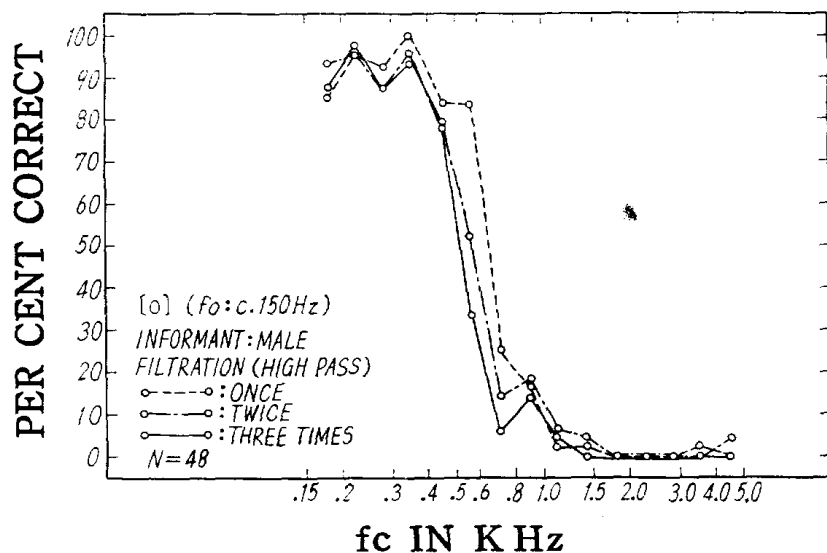


Figure 17. Per cent correct identification of [o] of the male speaker distorted by high-pass filtering.

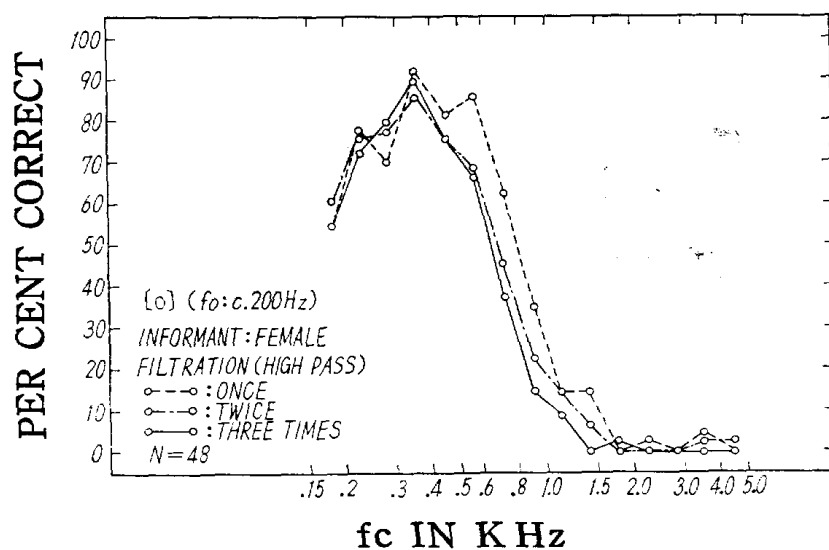


Figure 18. Per cent correct identification of [o] of the female speaker distorted by high-pass filtering.

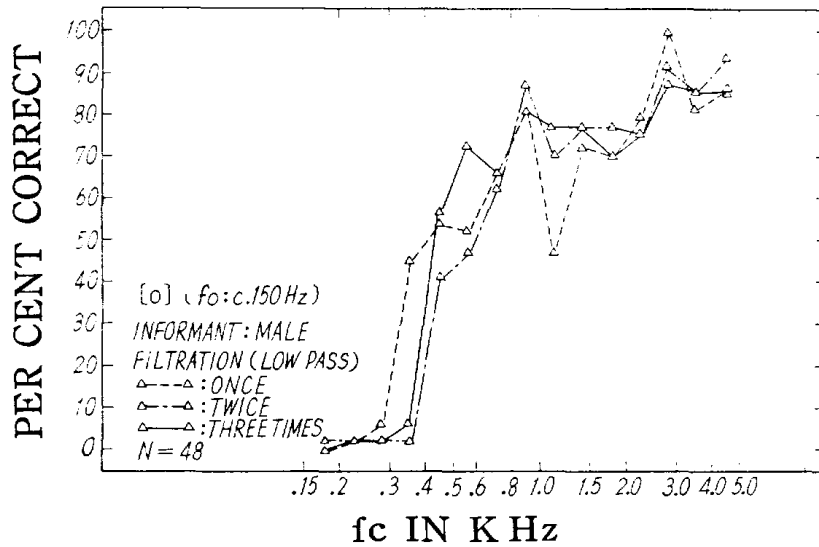


Figure 19. Per cent correct identification of [o] of the male speaker distorted by low-pass filtering.

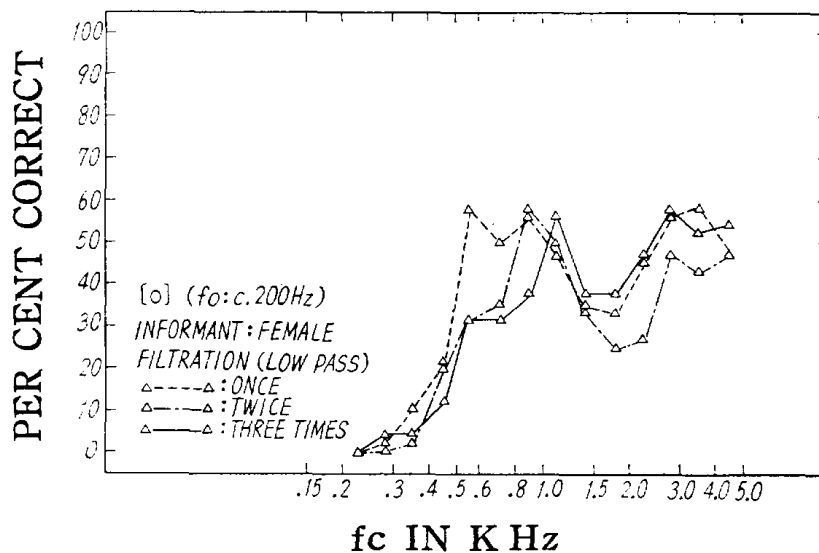


Figure 20. Per cent correct identification of [o] of the female speaker distorted by low-pass filtering.

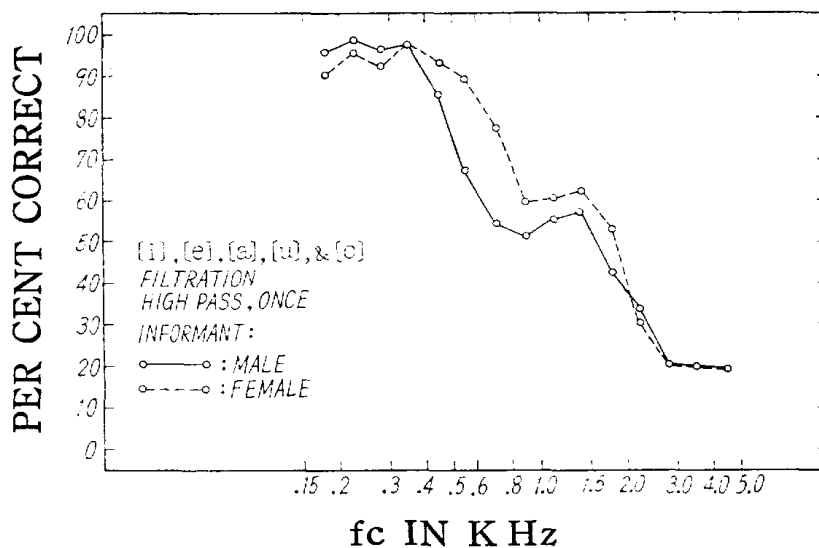


Figure 21. Mean per cent correct identification of 5 Japanese vowels [i], [e], [a], [u], and [o] of the male and female speakers distorted by high-pass filtering.  
(Attenuation: approximately 55 dB/octave.)

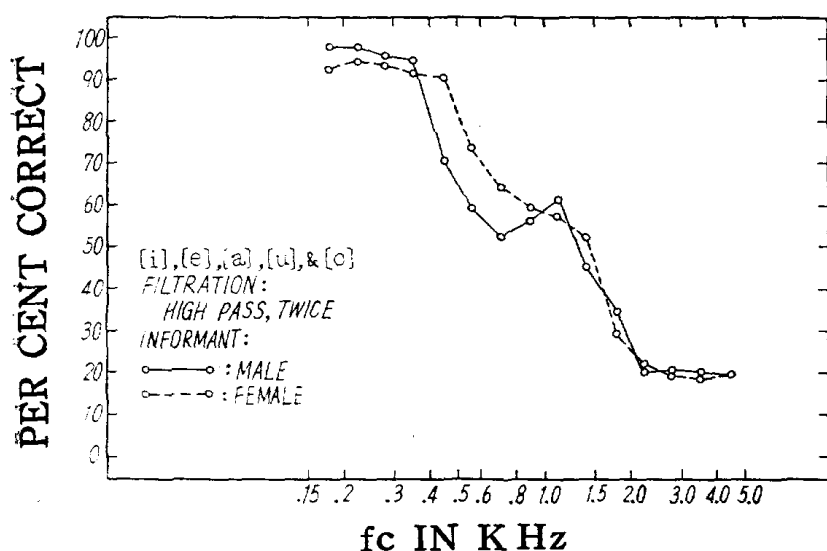


Figure 22. Mean per cent correct identification of 5 Japanese vowels [i], [e], [a], [u], and [o] of the male and female speakers distorted by high-pass filtering. (Attenuation: approximately  $2 \times 55$  dB/octave.)

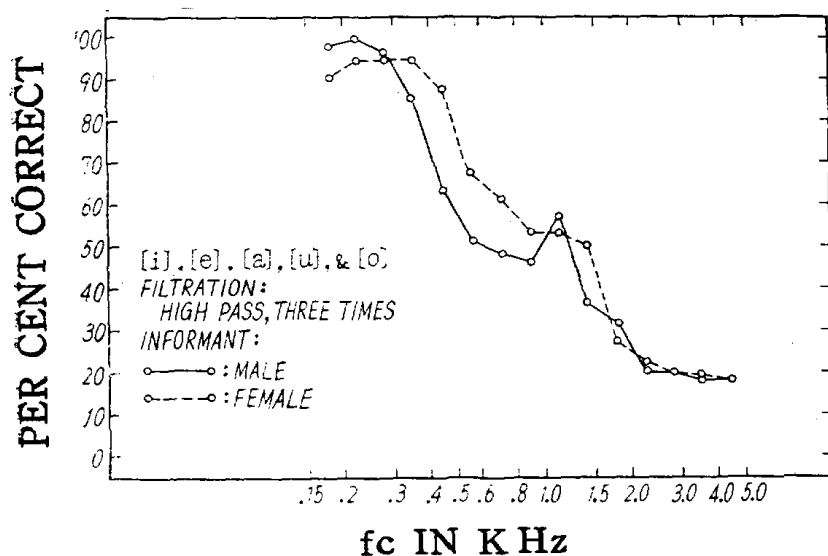


Figure 23. Mean per cent correct identification of 5 Japanese vowels [i], [e], [a], [u], and [o] of the male and female speakers distorted by high-pass filtering. (Attenuation: approximately  $3 \times 55$  dB/octave.)

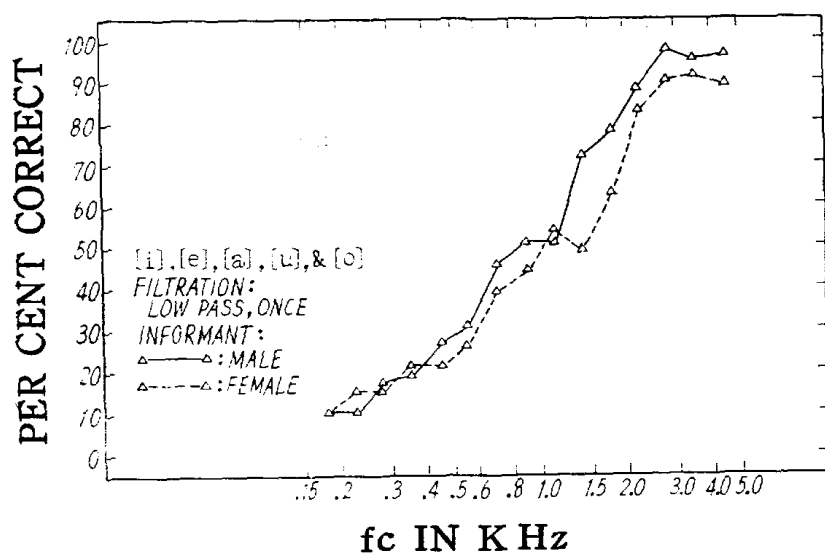


Figure 24. Mean per cent correct identification of 5 Japanese vowels [i], [e], [a], [u], and [o] of the male and female speakers distorted by low-pass filtering. (Attenuation: approximately 55 dB/octave.)

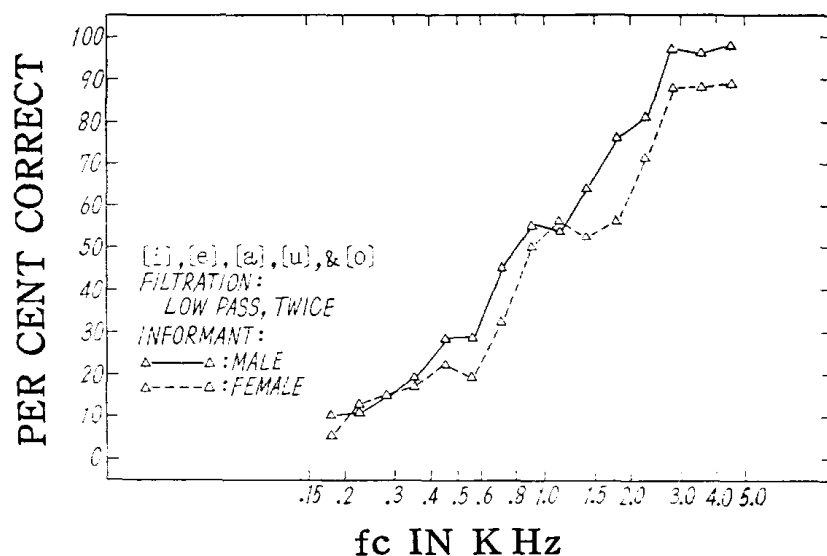


Figure 25. Mean per cent correct identification of 5 Japanese vowels [i], [e], [a], [u], and [o] of the male and female speakers distorted by low-pass filtering.

(Attenuation: approximately  $2 \times 55$  dB/octave.)

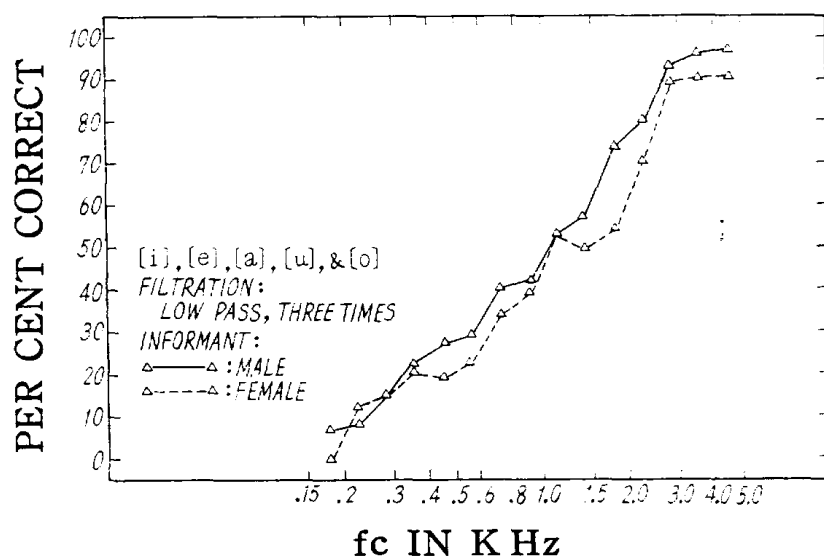


Figure 26. Mean per cent correct identification of 5 Japanese vowels [i], [e], [a], [u] and [o] of the male and female speakers distorted by low-pass filtering.

(Attenuation: approximately  $3 \times 55$  dB/octave.)

## SUMMARY

Japanese vowels [i], [e], [a], [o], and [u] recorded by an adult male and an adult female were identified under various conditions of high- and low-pass filtering.

1. Discontinuous functions of per cent correct identification of the vowels versus filter cut-off frequency were found to be in the identification of [i] (HP), [e] (HP), [a] (HP), [o] (LP), and [u] (HP and LP) of the male speaker, and in the identification of [i] (HP), [e] (LP), [o] (HP and LP), and [u] (HP and LP) of the female speaker. The results indicate that the *discontinuities* exist in more cases than those

suggested by Chiba and Kajiyama.

2. Difference was found to be between the correct identification of vowels of the male speaker and those of the female speaker reflecting their vowel spectrum envelopes. In general, vowels of the male speaker were identified better than those of the female speaker under low-pass filtering conditions, while the vowels of the female speaker were identified better than those of the male speaker under high-pass filtering conditions.
3. Difference was found to be between the identification of vowels distorted by different sharpness of filtering. (Attenuation: 55 dB/octave,  $2 \times 55$  dB/octave, and  $3 \times 55$  dB/octave.)
4. There was no statistically significant difference in per cent correct between the identification of the unfiltered vowels of both speakers intend for /i/, /e/, /a/, and /u/ and the unfiltered [o] of the male speaker, which had high identifiability (per cent correct: 100 to 9 except for [o], 98 to 80). The per cent correct identification of unfiltered [o] of the female speaker (per cent correct: 85 to 46.) was significantly lower than that of the other four vowels.
5. There were sometimes significant differences between the estimation of clearness of correctly identified vowels although the per cent correct identification of the vowels did not show any significant difference between them. The results imply that there are some frequency regions in which the vowel spectra still contribute to the phonetic quality of Japanese vowels, which cannot be revealed by merely phonemic categorization of the stimuli.

## REFERENCES

1. von Békésy, G., *Experiments in Hearing*. New York: McGraw-Hill, 1960.
2. Black, J. W., "The quality of a spoken vowel", *Arch. of Speech*, 2, 1937, 7-27 [not seen, quoted in Castle (4), q. v.].
3. Carterette, E. C., and Møller, A., "The perception of real and synthetic vowels after very sharp filtering," *Proceedings of the Speech Communication Seminar*, session d. Stockholm: The Royal Institute of Technology, 1963.
4. Castle, W. E., *The Effect of Selective Narrow-band Filtering on the Perception of Certain English Vowels*. The Hague: Mouton, 1964.
5. Catford, J. C., "Phonation types: the classification of some laryngeal components of speech production," *In Honour of Daniel Jones*, edited by D. Abercrombie, et al., 26-37. London: Longmans, 1964.
6. Chiba, T., and Kajiyama, M., *The Vowel: Its Nature and Structure*. Tokyo: Tokyo-Kaiseikan, 1941.

7. Crystal, D., and Quirk, R., *Systems of Prosodic and Paralinguistic Features in English*. The Hague: Mouton, 1964.
8. Doi, K., *Nihon-onsei no Jikken-teki Kenkyu* [An Experimental Study of Japanese Phonetics]. Tokyo: The Iwanamis, 1955.
9. Epstein, A., and Ulrich, J. H., "The effect of high- and low-pass filtering on the judged vocal quality of male and female speakers," *The Quarterly Journal of Speech*, 52, 1966, 267-272.
10. Fairbanks, G., and Grubb, P. A., "A psychophysical investigation of vowel formants," *Journal of Speech and Hearing Research*, 4, 1961, 203-219.
11. Fant, C. G. M., "On the predictability of formant levels and spectrum envelopes from formant frequencies," *For Roman Jakobson*, edited by M. Halle, *et al.*, 109-120. The Hague: Mouton, 1956.
12. Feldtkeller, R., and Zwicker, E., *Das Ohr als Nachrichtenempfänger*. Stuttgart: S. Hirzel, 1956.
13. Fletcher, H., *Speech and Hearing in Communication*. Princeton, N. J.: D. Van Nostrand, 1961.
14. French, N. R., and Steinberg, J. C., "Factors governing the intelligibility of speech sounds," *Journal of the Acoustical Society of America*, 1947, 90-119.
15. Greenwood, D. D., "Auditory masking and the critical band," *Journal of the Acoustical Society of America*, 33, 1961, 484-502.
16. House, A. S., "Formant band widths and vowel preference," *Journal of Speech and Hearing Research*, 3, 1960, 3-8.
17. Katsuki, Y., "Neural mechanism of auditory sensation in cats," *Sensory Communication*, edited by W.A. Rosenblith, 561-583. Cambridge, Mass.: The M. I. T. Press, 1962.
18. Lehiste, I., and Peterson, G. E., "The identification of filtered vowels," *Phonetica*, 4, 1959, 161-177.
19. Lindblom, B., "Accuracy and limitations of Sona-Graph measurements," *Proceeding of the 4th International Congress of Phonetic Sciences*, 188-202. The Hague: Mouton, 1962.
20. Nakata, K., *A Synthetic Study of Japanese Speech, Review of the Radio Research Laboratories*, supplement 4, Dec. 1961.
21. Ochiai, Y., and Fukumura, T., "Timbre study of vocalic voices viewed from subjective phonal aspect, Part I: Preliminary studies on naturalness and articulation qualities actually and directly measured with respect to band-eliminating distortion," *Mem. Fac. Engng. Nagoya*, 8, 1956, 1-18.
22. Ochiai, Y., "Phoneme and voice identification studies using Japanese vowels," *Language and Speech*, 2, 1959, 132-136.
23. Ohta, F., Yanagihara, N., and Hosoda, I., "Binaural fusion test with band-pass filters," *Audiology*, 7, 1964, 29-34.



24. Palva, A., *Filtered Speech Audiometry I: Basic Studies towards the Creation of Method for the Diagnosis of Central Hearing Disorders*, *Acta Oto-Laryngologica*, supplement 210. Turku: University of Turku, 1965.
25. Peterson, G. E., "Vowel formant measurements," *Journal of Speech and Hearing Research*, 2, 1959, 173-183.
26. Peterson, G. E., and Barney, H. L., "Control methods used in a study of the vowels," *Journal of the Acoustical Society of America*, 24, 1953, 75-84.
27. Schafer, T. H., and Thompson, P. O., "The frequency selectivity of ear as determined by masking experiments," *Journal of the Acoustical Society of America*, 22, 1950, 490-496.
28. Sekiguchi, S., "Icho-ritsu niyoru boin spekutoramu no kosatsu [Evaluation of vowel spectrum: an experimental study of confusion of filtered Japanese vowels]," *Records of the Radio Research Laboratories*, no. 443, 1956.
29. Steinberg, J. C., "Effects of distortion on speech and music," *Electrical Engineer's Handbook*, edited by H. Pender and M. McIlwain [not seen, quoted in Castle (4), q. v.]. New York: John Wiley, 1936.
30. Tiffany, W. R., "Threshold reliability of recording sustained vowels," *Journal of Speech and Hearing Disorders*, 18, 1953, 379-385.