Ann. Bull. RILP (1982) No. 16, 81-88

TONAL DIFFERENCE LIMENS FOR SECOND FORMANT FREQUENCIES OF SYNTHESIZED JAPANESE VOWELS

Tatsuo Nakagawa, Shuzo Saito, and Tomoyoshi Yoshino***

Introduction

To evaluate the maximum accuracy necessary in designing a speech transmission system, the frequency difference limens (DL's) for vowel formants were measured by Flanagan[1]. They were on the order of three to five percent of the formant frequencies for both the first and second formants (F1 and F2).

The acoustic features of the synthetic stimuli used in his experiment were not those of natural vowels. For example, to determine the DL's for the F2 frequencies of 1,000, 1,500, and 2,000 Hz F1, F2, and F3 frequencies of 500, 2,500, and 3,550 Hz, respectively, were used. And for all stimuli the fundamental frequency was held constant at 120 Hz.

In this paper, we attempt to investigate whether the above results hold true in the case of determing the DL's for synthesized stimuli with similar acoustic characteristics to natural vowels. Here the results of an experiment measuring the DL's for the F2 frequencies of stimuli synthesized on the basis of the data analyzed by the linear prediction analysis method will be presented.

Method

1. Subjects

7-, 9-, and 11-year-old normal-hearing children and normal adults (age range: 22 to 29 years; mean = 25.2 years) participated in this experiment. Six subjects took part in four groups.

2. Stimuli

The five Japanese vowels /i/, /e/, /a/, /o/, and /u/ were pronounced by three adult male speakers. These speech sounds were low-pass filtered at 5 kHz and sampled at 10 kHz with 11 bits. For each utterance, linear prediction spectra were calculated with 10 poles.

The reference stimuli were prepared as follows. The values of the formant frequencies and bandwidths were chosen from those of the quasi steady-state portions of the analyzed data. The formant frequencies and bandwidths were thus fixed at those values for each synthetic stimulus (Table 1). The duration of each stimulus is shown in Table 2.

* Graduate Program in Special Education, Tsukuba University
** Institute of Special Education, Tsukuba University

Table 1 *The formant frequencies and bandwidths for each reference stimulus*

	SPEAKER	Y					M					S				
	PHONEME	/a/	/i/	/u/	/e/	/o/	/a/	/i/	/u/	/e/	/o/	/a/	/i/	/u/	/e/	/o/
F1	FREQUENCY (HZ)	690	305	335	515	500	710	300	390	540	540	815	270	300	505	480
	BANDWIDTH (HZ)	100	20	50	80	100	125	45	60	30	55	80	50	80	80	60
F2	FREQUENCY (HZ)	1100	2240	1250	1985	800	1140	2100	1200	1850	860	1155	2500	740	2000	730
	BANDWIDTH (HZ)	130	90	120	70	150	130	50	130	80	120	60	90	50	180	150
F3	FREQUENCY (HZ)	2750	3285	2280	2600	2800	2700	2940	2250	2460	2450	2730	3400	2400	2500	2970
	BANDWIDTH (HZ)	180	200	150	230	200	160	150	100	160	200	150	200	250	270	120
F4	FREQUENCY (HZ)	3750	3700	3550	3750	3600	3530	3540	3370	3580	3420	3500	3770	3250	3050	3450
	BANDWIDTH (HZ)	200	200	200	170	200	220	200	150	200	250	180	150	200	140	450
F5	FREQUENCY (HZ)	4050	5500	5500	4100	5500	5500	5500	5500	5500	5500	5500	5500	3600	5500	4200
	BANDWIDTH (HZ)	180	300	300	220	300	300	300	300	300	300	300	300	80	300	150

Table 2 *The duration for each synthesized vowel*

SPEAKER	Y					M					S				
PHONEME	/a/	/i/	/u/	/e/	/o/	/a/	/i/	/u/	/e/	/o/	/a/	/i/	/u/	/e/	/o/
DURATION (MS)	245	230	240	245	240	215	215	210	210	205	235	265	285	225	240

The overall amplitude and the fundamental frequency of speaker Y's /a/ are shown in Fig. 1. As can be seen in this figure, the overall amplitude change and fluctuation in the fundamental frequency were simulated in the analyzed data.

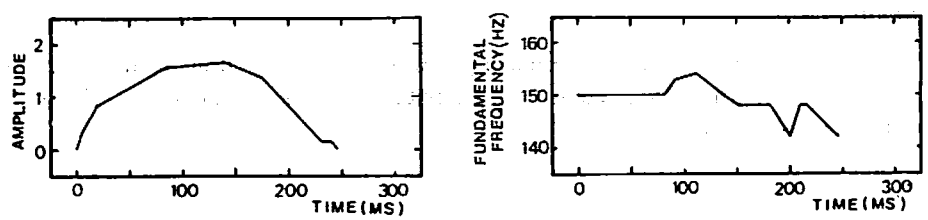

Fig. 1 *The overall amplitude change and fluctuation in the fundamental frequency for the synthesized /a/ by speaker Y.*

On the basis of these synthetic parameters, the stimuli were generated by a computer-simulated cascade-resonance synthesizer.

The test stimuli were prepared as follows. Only the F2 frequencies were shifted up or down. For the adult subjects the relative values defined by the ratio of the F2 frequency deviations to the corresponding reference F2 frequency are shown in Table 3. For the children, F2 was equally shifted in descrete steps on either side of the formant. The minimal shift of the F2 frequency for /i/, /e/, /a/, /o/, and /u/ was 65, 50, 30, 25, and 50 Hz respectively.

Table 3 *Relative F2 frequency deviations (ΔF2/F2) for the test stimuli for the adult subjects. The symbol "+" indicates an upward shift and the symbol "–" indicates a downward shift.*

STIMULUS NUMBER		1	2	3	4	5	6	7
PHONEME	/a/	+0.02	+0.05	+0.08	+0.12	+0.16	+0.2	+0.35
	/i/	-0.02	-0.05	-0.08	-0.12	-0.16	-0.2	-0.35
	/u/ <	+0.04	+0.08	+0.12	+0.16	+0.22	+0.28	+0.4
		-0.04	-0.08	-0.12	-0.16	-0.22	-0.28	-0.4
	/e/	-0.02	-0.05	-0.08	-0.12	-0.16	-0.2	-0.35
	/o/	+0.04	+0.08	+0.12	+0.16	+0.22	+0.28	+0.4

3. *Procedure*

The AB test method was used. The reference stimulus (A) was paired with one of the test stimuli (B) or the reference stimulus (A). The time spacing between stimuli was 0.6 second and between pairs was 6 seconds.

To determine the DL's for each test condition (vowels and F2 shift directions), ten sessions were conducted with the adult subjects and six sessions were conducted with the children. One session consisted of ten randomized pairs, of which seven pairs were different (AB or BA) and three pairs were identical (AA).

The subjects listened monaurally over TDH-49 earphone while seated in a sound-poof room. The most comfortable listening level was determined by presenting the five reference stimuli to each subject and finding the sound-pressure level at which each subject felt best hearing them. This level was used for presentation of the stimuli to each subject. There were no significant differences in the levels between subjects. The mean sound pressure-level for the reference stimulus /a/ was approximately 77.0 dB re. 2×10^{-5} N/m^2.

A "same/different" response paradigm was employed. The subjects were provided with the following instructions: Please listen carefully to the paired sounds. If you think the two sounds are different, please push the red button, and if you think they are same, please push the yellow button.

Prior to the measurement of the DL's for each vowel, two practice sessions were given to each subject. In these sessions the subjects were given feedback on the correctness of their responses.

Each adult subject listened to synthetic stimuli produced by two different speakers. Children listened to stimuli produced by one of the speakers (speaker Y).

Results and Discussion

For the adult subjects' data, the percentage of responses judged as different was calculated separately for each of the vowels and speakers. For the children's data, the percentage was calculated separately for each of the vowels and age groups. Figs. 2 and 3 show the results of the F2 frequency discrimination tests for the adults

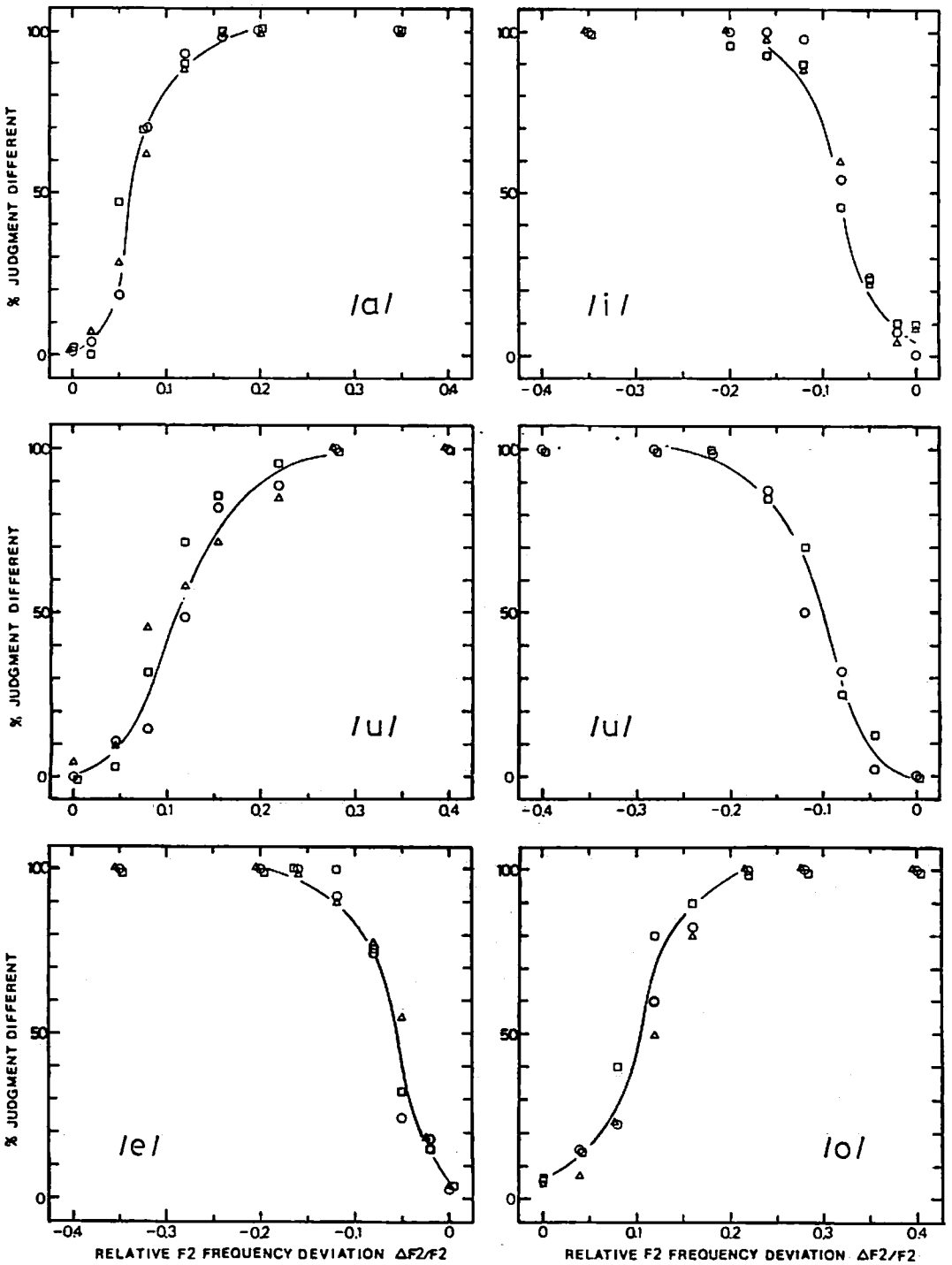

Fig. 2 *Discrimination functions by normal-hearing adults. The open squares, open circles, and open triangles stand for the mean values of the percentage of responses to each synthesized vowel for speakers Y, M, and S, respectively.*

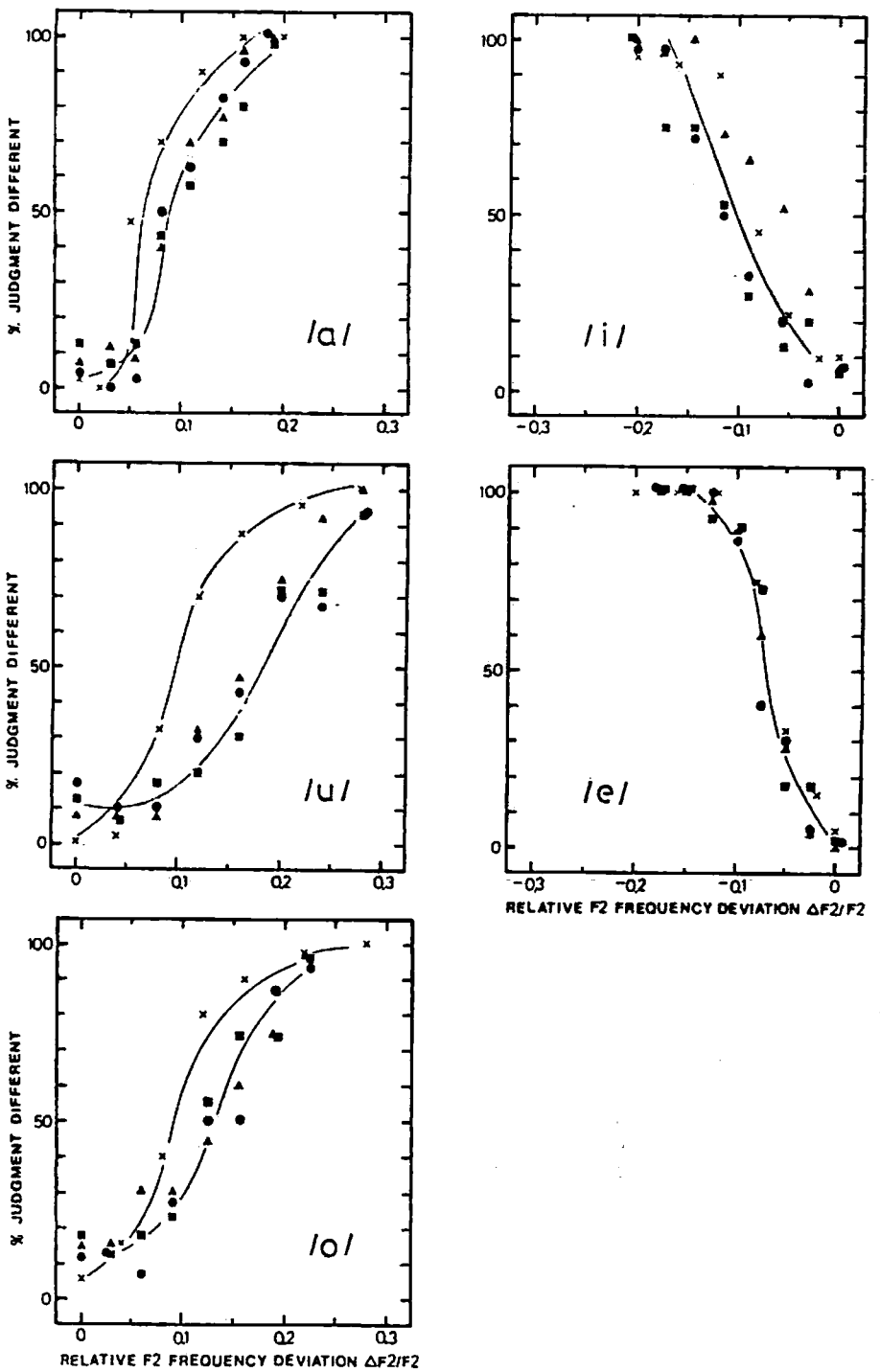

Fig. 3 *Discrimination functions by normal-hearing children. The filled squares, filled circles, and filled triangles stand for the mean values of the percentage of responses for the seven-, nine-, and eleven-year-old groups, respectively. For the sake of comparison, the symbol "x" on the basis of the data from the adult subjects are included in this figure.*

Table 4 *DL's and relative DL's (DL/F2) for each F2 frequency for the adult subjects*

SPEAKER	PHONEME	F2 OF REFERENCE(HZ)	DL(HZ)	DL/F2(%)
Y	/a/	1100	76.2	6.9
	/i/	2240	-165.3	- 7.3
	/u/	1250	146.4	11.7
			-135.1	-10.8
	/e/	1985	-121.8	- 6.1
	/o/	800	70.7	8.8
M	/a/	1140	85.9	7.5
	/i/	2100	-142.9	- 6.8
	/u/	1200	159.5	13.3
			-138.2	-11.5
	/e/	1850	-119.7	- 6.5
	/o/	860	68.1	7.9
S	/a/	1155	89.2	7.7
	/i/	2500	-178.8	- 7.2
	/u/	740	96.8	13.1
	/e/	2000	-125.3	- 6.3
	/o/	730	86.6	11.8

Table 5 *DL's and relative DL's (DL/F2) for each F2 frequency for the children*

AGE	PHONEME	F2 OF REFERENCE(HZ)	DL(HZ)	DL/F2(%)
7years	/a/	1100	105.0	9.5
	/i/	2240	-183.7	- 8.2
	/u/	1250	188.7	15.1
	/e/	1985	-137.7	- 6.9
	/o/	800	88.5	11.1
9years	/a/	1100	109.9	10.0
	/i/	2240	-270.9	-12.1
	/u/	1250	213.8	17.1
	/e/	1985	-134.5	- 6.8
	/o/	800	90.6	11.3
11years	/a/	1100	109.5	10.0
	/i/	2240	-229.3	-10.2
	/u/	1250	209.4	16.8
	/e/	1985	-139.7	- 7.0
	/o/	800	97.4	12.2

and children, respectively.

The DL for the F2 frequency was defined as the point along the F2 frequency continuum at which 50% of the judgments were different and 50% were same. This was estimated by the Müller-Urban process (Tables 4 and 5). Also included in these tables is the relative DL defined by the ratio of the DL to the corresponding reference F2 frequency.

For the adult subjects the discrimination tests gave values of 6.1–13.3% of the formant frequencies, which were larger than the corresponding values of 3–5% reported in Flanagan's study. What could be the explanation for these differences between Flanagan's data and ours?

In judgments of vowel quality, the phonemic decoding process that must follow the acoustic analysis of the stimuli may be involved more in our case than in the case of Flanagan. Because the stimuli used here are characterized by properties similar to those of natural vowels.

There was a general tendency for the relative DL's of the back vowels /u/ and /o/ to be larger than those of the front vowels /i/ and /e/ or the middle vowel /a/. The results also show that there was little effect from speaker variation on the DL for each vowel.

Dellatre et al.[2] have tested the identifiability of synthetic vowels with a single-formant. They found that the single-formant positions which were sufficient to produce a back vowel color were in the region of the first formant. Non-back vowels, except /i/, needed two formants for high identifiability. According to these findings, for back vowels the perceptual dominant formant seems to be the first formant. Therefore, it seems likely that F2 frequency deviations in vowels do not affect the perceptual continuum.

This tendency observed in the F2 frequency discrimination test for the adult subjects was also obtained in the same test for children. It can be seen that the children differed from the adults in the relative DL's of vowels other than front vowels. An age effect for the relative DL's was not obtained among the children.

Considering the age groups studied, one might expect more intersubject variability in the children's data than in the adults' (and likewise, more in the 7-year-olds' than the 11-year-olds'). Thus, we are going to carry out futher experiments for each of these age groups.

Summary

The difference limens (DL's) of second formant (F2) frequencies were measured for five synthetic Japanese vowels. The stimuli were synthesized using a computer-programmed cascade-resonance synthesizer on the basis of data analyzed by the linear prediction analysis method. The results of the discrimination test gave larger values than those reported by Flanagan. The DL's for the back vowels were found to be larger than those for the middle and front vowels. Possible explanations for the results obtained were discussed. Though tentative, F2 frequency discrimination data for children were also presented.

Acknowledgments

We are extremely grateful to Sotaro Sekimoto and Hiroshi Imagawa at the University of Tokyo for their assistance in preparing the synthetic stimuli. Our sincere appreciation is also extended to Hisashi Kado at the Electrotechnical Laboratory for his statistical advice and to Sanetomi Eguchi at Tsukuba University for his useful advice and interest in this study.

References

1. Flanagan, J.L. (1955); "A difference limen for vowel formant frequency," The Journal of the Acoustical Society of America 27, 613-617.
2. Dellatre, P.C., A.M. Liberman, F.S. Cooper, and L.J. Gerstman (1952); "An experimental study of the acoustic determinants of vowel color; Observations on one- and two-formant vowels synthesized from spectrographic patterns," Word 8, 195-210.