Perceptual Normalization of Vocal Tract Size in Infants

Fumi Katoh, Toshisada Deguchi*, Akiko Hayashi* and Shigeru Kiritani

Introduction

Human speech perception shows remarkable abilities to normalize wide variations in speech sounds such as those due to speaker differences and context differences. Among such abilities is perceptual normalization of vocal tract size effects. We have been conducting a series of experiments on the development of this ability.

In our previous paper (K.Sakata et al., 1990), we reported the results of experiments comparing the ability of adults and young children aged 3-5. In this paper, preliminary results on infants aged 5-7 months will be presented.

Experiments on infants have been conducted using the technique of conditioned head-turn responses. Our experiments consist of two parts, Experiment 1 and 2. Experiment 1 was conducted essentially following the procedures developed by Kuhl et al. Based on our experience in Experiment 1, slightly modified procedures were employed in the Experiment 2. At the same time, a control experiment using adult subjects was conducted under experimental conditions similar to used for the infants.

Experiment 1

Experimental Procedure

Synthetic vowels were produced using a computer program simulating a terminal-analog speech synthesizer. A series of vowels which varied from /o/ to /a/ were synthesized by changing the first and second formant frequencies as follows.

F_1=400,450,500,..........,900 Hz
$F_2=1.1F_1+350$ Hz

The parameter values were selected based on the results of the analysis in Fujisaki and Kawashima's study (Fujisaki and Kawashima, 1968). Two set of vowels were synthesized using different fundamental frequencies (F_0=100Hz, 220Hz). For a given F_0, the frequencies of the higher formants were varied concurrently.

F_0 100, 220 Hz
F_3 $6.3(F_0+270)$ Hz
F_4 $1.4F_3$ Hz
F_5 $1.8F_3$ Hz

* Faculty of Education, Tokyo Gakugei University

The effect of the changes in the fundamental frequency and the formant frequencies on the perceived vowel boundaries in infants were measured using the conditioned head-turn response technique.

The experimental procedure essentially follows that developed by Kuhl (1979). An infant is held by a caretaker sitting in a booth. A loudspeaker is located at a 90-degree-angle to the left of the infant and a visual reinforcer (a mechanical animal toy) is placed on the speaker. The stimulus sounds are delivered through the speaker, and the infant is conditioned to make a head-turn response toward the visual reinforcer when the stimulus is changed from one speech sound to another.

In each trial, a series of ten stimulus sounds is presented. The method of stimulus presentation is illustrated in Figure 1. In the training phase, two synthetic vowels, a representative /o/ and /a/, are used as stimulus sounds. One of the two synthetic vowels is presented four times as the background stimulus at intervals of two seconds. The stimulus sound is then changed to the other vowel and presented three times. Following the change in the stimulus sound, the reinforcer is activated. The infant learns to anticipate the reinforcer and make a head-turn with the change in the stimulus sound.

In the second stage of the training phase, the stimulus sound is changed in half of the trials but not in the other half. The infant learns to make a head-turn only when the stimulus sound is changed. The infant remains at the second stage of training, until the performance criterion, nine out of ten consecutive trials correct, is met.

In the experimental phase, the background stimulus is one of the representative vowels used in the training phase, and the test stimuli are the intermediate vowels (See Fig.1 for illustration). The frequency of the head-turns for each intermediate vowel is recorded.

Five infants aged 5 to 7 months served as subjects.

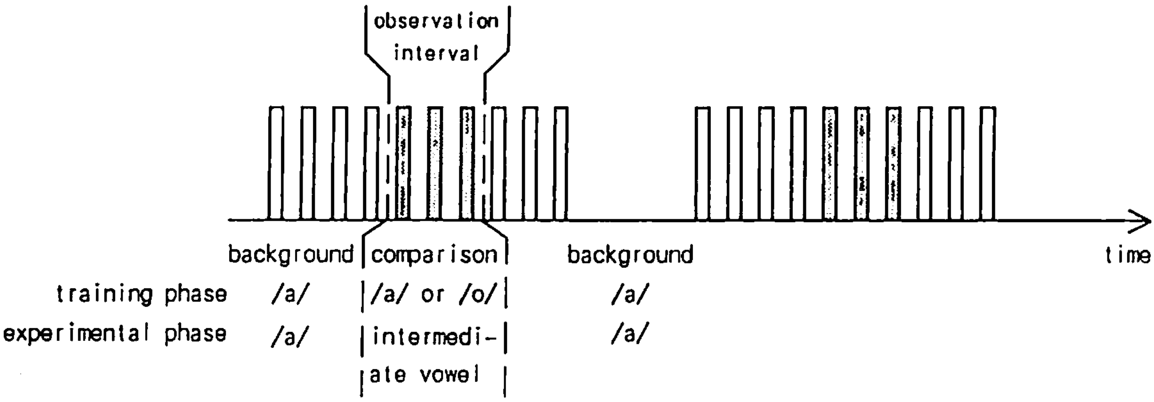

Figure 1. Stimulus presentation format in Experiment 1. Head-turn responses were measured during the observation interval.

Results and Discussion

Figure 2 shows an example of the response pattern obtained from a 6-month-old infant who could categorize the synthetic vowels in this experimental paradigm. Generally, all five infants were able to discriminate between the stimulus vowels, although, as shown in Fig.3, the accuracy of then identification was worse than that of the control adults. The response curves in Figure 2 were approximated by the method of maximum-likelihood, and the positions of the vowel boundary were estimated. Fig.4 shows the vowel boundaries estimated for the 5 infants. As for the shift in the vowel boundary due to the change in the fundamental frequency, there were some inter-subject variations. Subjects 1,3 and 4 showed effects similar to those of shown by the adults, namely, the boundary formant frequency became higher with a higher fundamental frequency. Subject 2 did not show a significant shift in the vowel boundary. Subject 5 showed a vowel boundary shift opposite to that shown by the adults. However, this infant required an exceptionally large number of training trails, compared to the other infants, to be able to discriminate the stimulus vowels, and thus may be considered a special case.

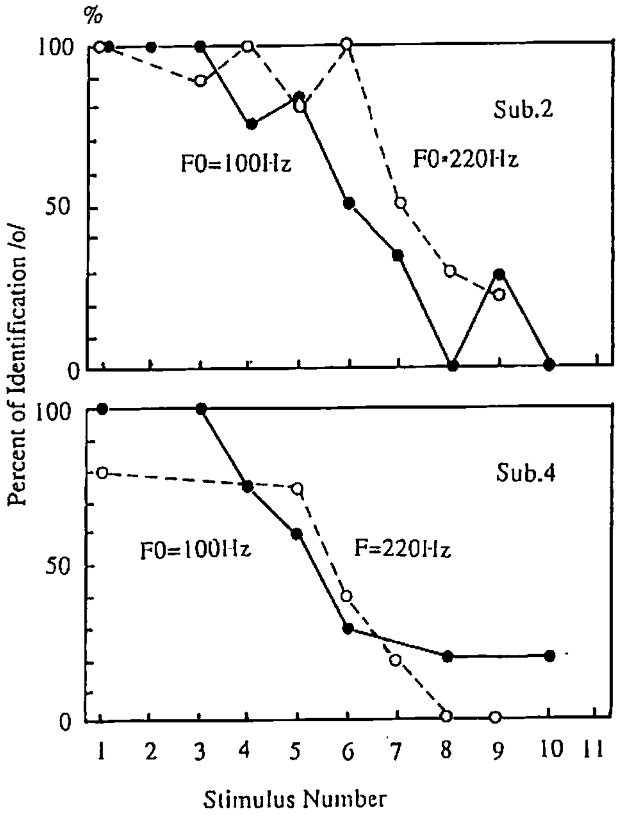

Figure 2. Discrimination curve for infants' head-turn responses.

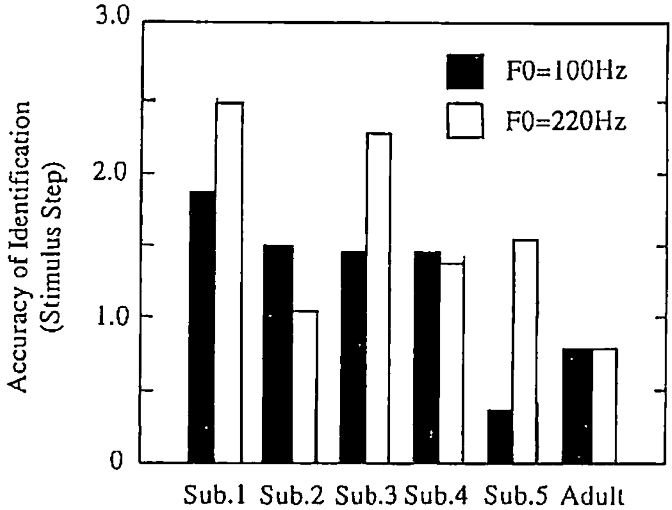

Figure 3. Accuracy of discrimination for the 5 infants in Experiment 1 and the average result for the adults control.

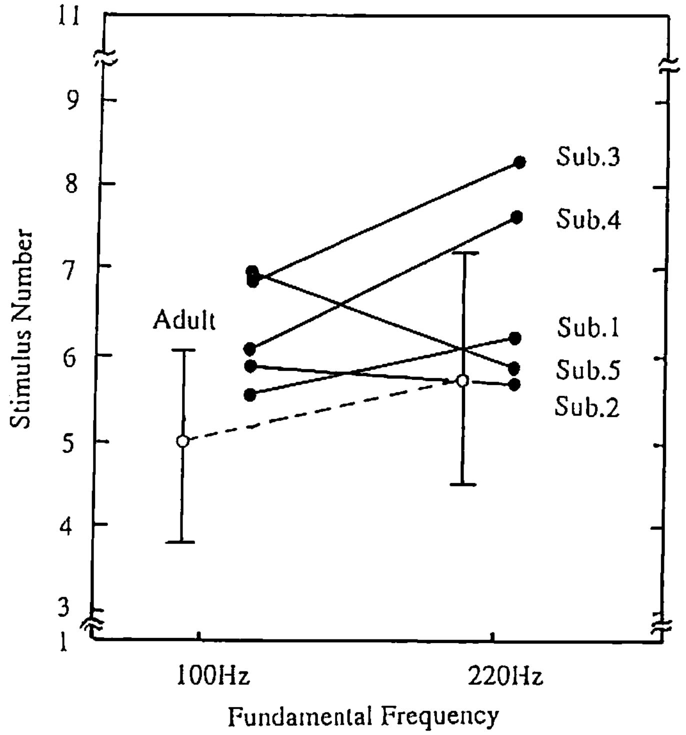

Figure 4. Vowel Boundary for infants(•–)in Experiment 1 and the average result for the adults.

Experiment 2

Experimental Procedure

Based on the results of Experiment 1, Experiment 2 was conducted with slight modifications in the stimulus sounds and the method of stimulus presentation.

The procedure in Experiment 1, where stimulus sounds were synthesized with a constant pitch, appears to have resulted in a somewhat unnatural sound quality. Thus, in Experiment 2, the stimulus sounds were synthesized with pitch inflection as follows.

	0msec	100msec	140msec	500msec
F_0=100Hz	100Hz	118Hz	118Hz	82Hz
F_0=220Hz	220Hz	260Hz	260Hz	181Hz

As can be seen in Fig.5(a), a considerable number of trials were required for the training phase in Experiment 1. With an aim to reduce the number of trials required for

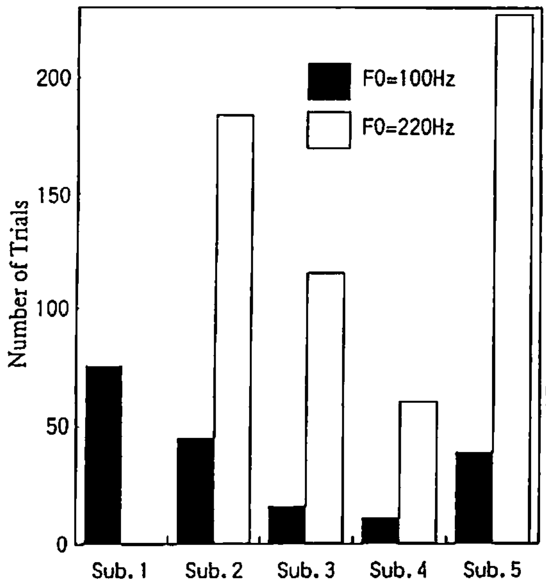

Figure 5(a). Number of trials required to meet the criterion of the training stage in Experiment 1.

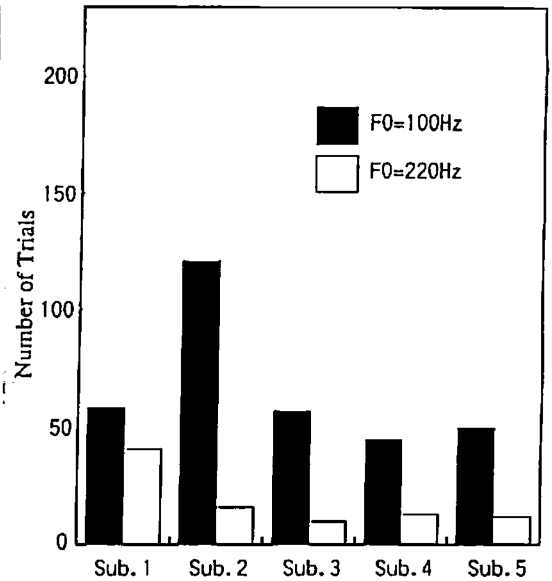

Figure 5(b). Number of trails required to meet the criterion of the training stage in Experiment 2.

the training phase, several modifications in the stimulus presentation were tested, and the following procedure was adopted (Fig.6). The background stimulus was presented constantly at regular intervals until the experimenter judged the infant ready and signaled the computer to start the observation period. Then the control program randomly determined whether or not to present a comparison stimulus. After presentation of the comparison stimulus, the control program went back to the continuous presentation of the background stimulus.

As for the reference data from the adults, these were recollected using new stimulus sounds . At the same time, considering a possible contrast effect of the background stimulus in the infants experiment, the method of the identification test with the adults was modified so that precursor stimuli were presented preceding target stimuli. The precursor stimuli were the same as the background stimuli in the infant experiment and were repeated 4 times.

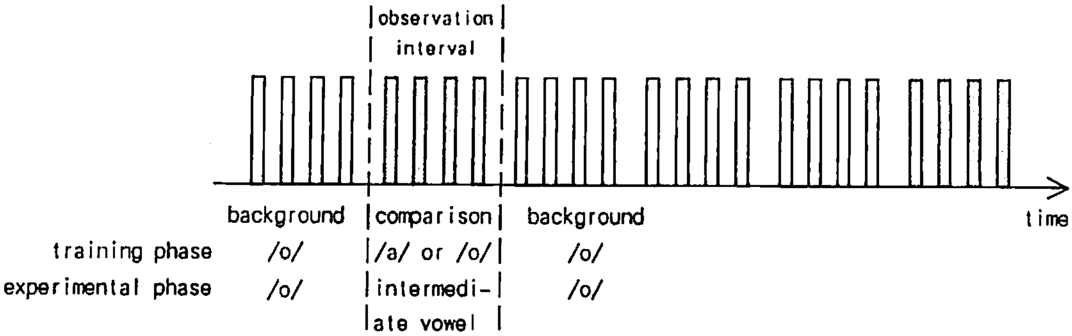

Figure 6. Stimulus presentation format in Experiment 2.

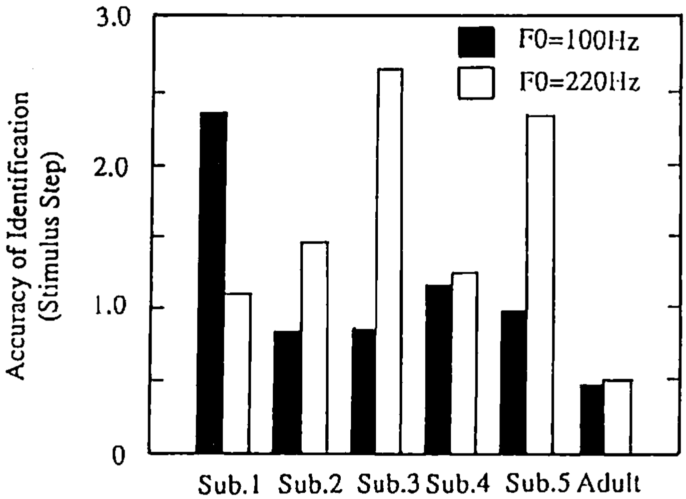

Figure 7. Accuracy of discrimination for the 5 infants in Experiment 2 and the averaged result for 7 the adults controls.

Results and Discussion

Fig.5(b) shows that number of trials in the training phase was greatly reduced with the new method, with nearly the same accuracy of discrimination as that obtained in Experiment 1 (Fig.7). Fig.8 shows the boundary shift associated with the fundamental frequency change obtained for the 5 infants. Average results for the seven adults are also shown in the figure. It can be seen in the figure that on an average the subjects showed a shift in vowel boundary due to the fundamental frequency change, although there were some inter-subject variations.

The above results suggest that infants show an effect of boundary change due to fundamental frequency similar to that shown by adults. However, within the above experimental paradigm, these results may be interpreted as reflecting a change in discriminability. Further experiments are now being considered to confirm the change in the vowel category boundary, including a mixed presentation of vowels with different fundamental frequencies within one experimental session.

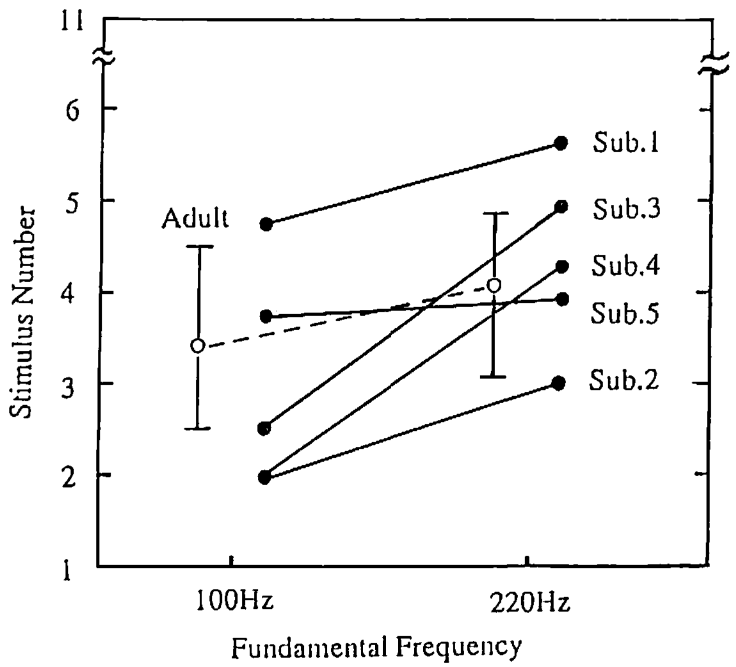

Figure 8. Vowel Boundary for the 5 infants(•–•)in Experiment 2 and the average result for the adults.

References

1) Kuhl,P.K. :Perception of speech and sound in early infancy, In Salapatek,P. and Cohen,L., Handbook of Infant Perception, vol.2, From Perception to Cognition, New York Academic Press, 275-373,1987.
2) Fujisaki,H. and Kawashima,T.:The role of pith and higher-formants in the perception of vowels, IEEE Transactions on Audio and Electroacoustics, AU-16(1),1988.
3) Kuhl,P.K.:Speech perception in early infancy:Perceptual constancy for spectrally dissimilar vowel categories, J. Accost. Soc. Am., 66,1668-1679,1979.
4) Sakata,K. et al.,:Ability of perceptual normalization of vocal tract size in young chil dren, ANN. Bull. RILP,No.24,1990.