ANN.Bull.RILP(1990)No.24, 157-162

ABILITY OF PERCEPTUAL NORMALIZATION OF VOCAL TRACT SIZE IN YOUNG CHILDREN

Keiko Sakata*, Toshisada Deguchi*, Akiko Hayashi and Shigeru Kiritani

In recent years, it has been shown by several studies that human infants exhibit auditory capacities which conform to the phonemic categories in languages [1]. Infants (even pre-linguistic infants) are capable of differentiating speech sounds according to phonetic categories (at least for some phonetic categories). These studies have been mainly concerned with the so-called phenomena of categorical perception. More specifically, an adult's auditory perception generally shows an enhanced discrimination for pairs of sounds that straddle phonetic boundaries. Studies on infants have focused upon examining whether infants show similar perceptual characteristics.

Another important feature of adults' speech perception is the ability at perceptual normalization of wide variations in the acoustic characteristics of given phonemes in natural utterances. Such variations are introduced through differences among speakers and/or through contextual effects. It is quite interesting and important to ask whether such perceptual normalization is observed in infants.

One of the most dominant and systematic factors contributing to variations in speech signals due to speaker difference is the effect of vocal tract size. Children generally have smaller vocal tracts than adults and, thus, higher formant frequencies for vowels.

These shifts in formant frequencies are observed not only for the first and second formants which are most important for distinguishing vowels, but also for the third and higher formants. At the same time, children generally have higher fundamental frequencies than adults. It has been demonstrated by Fujisaki et al. [2] that both of these acoustic factors contribute to the perceptual normalization of vowel formants. The present study aims at investigating whether a similar perceptual normalization of vocal tract size exists in infant's and children's auditory perception.

This paper reports the results of perceptual experiments on children aged 3-5 years, which were performed as a preliminary step toward experiments on infants.

Experimental Procedure

Synthetic vowels were produced using a computer program simulating a terminal-analog speech synthesizer. A series of

*Faculty of Education, Tokyo Gakugei Univ.

vowels which varied from /o/ to /a/ and from /u/ to /e/ were synthesized by changing the first and second formant frequencies as follows.

Series /o/-/a/

F_1=400, 450, 500,........, 900Hz

$F_2=1.1F_1 + 350$Hz

Series /u/-/e/

$F_1=F_2/3$

F_2=1100, 1200, 1300,....., 2300Hz

Each series of vowels was synthesized under three different conditions of fundamental frequency (f_0) and higher formant frequencies. For a given condition of f_0, the frequencies of the higher formants were varied concurrently.

F_0	100, 180, 220, 260Hz
F_3	$6.3(f_0+270)$Hz
F_4	$1.4F_3$
F_5	$1.8F_3$

These synthetic vowels were presented to the subjects to evaluate shift in the vowel boundaries due to the change in fundamental frequency and the higher formant frequencies. The range of parameter values were essentially the same as in Fujisaki et al's experiment [2)], which were selected based on an analysis of natural utterances by a group of subjects containing children, adult females and adult males.

Subjects were young Japanese children aged 3, 4 and 5. The number of subjects in each age group was 12, 16 and 14 respectively. Reference data from the adults were obtained from a group of 12 subjects. The stimulus sounds were presented in random order. Each stimulus sound was presented 10 times. The subjects were asked to respond orally whether a given stimulus was /o/ or /a/ (or /u/ or /e/). For some of the children, it was rather difficult to make an oral response by pronouncing isolated vowels. In case of these children, a picture card related to each vowel category was prepared and responses were obtained of the children pointing at these cards. The experiments with the children were conducted using stimulus sounds with fundamental frequencies of 100Hz and 220Hz. For the adult experiments, stimulus sounds with all four fundamental frequencies were used.

Results and Discussion

Fig.1 shows an example of the identification curve for the synthetic vowels of /u/-/e/ series obtained for a child aged 3. In order to evaluate quantitatively the pattern of the identification curves, the curves were analyzed through an approximation by a cumulative normal distribution, and, for each curve, the position of the vowel boundary and the accuracy of identification were estimated. The vowel boundary was determined as the mean of the normal distribution, and the accuracy of identification was determined as the standard deviation of the normal distribution. Fig.2 shows the mean vowel boundaries estimated for each of the 4 subject groups and for the /o/-/a/ series and the /u/-/e/ series. It can be seen in the figure that all the subject groups show a similar shift in the boundary of the formant frequencies with changes in the fundamental frequency and higher formants. It was noted, however, that the amount of boundary shift appeared to be greater for young children than for adults. Fig.3 compares the accuracy of identification for the four subject groups. It was confirmed that children aged 3-5 already have almost the same accuracy of identification as adults.

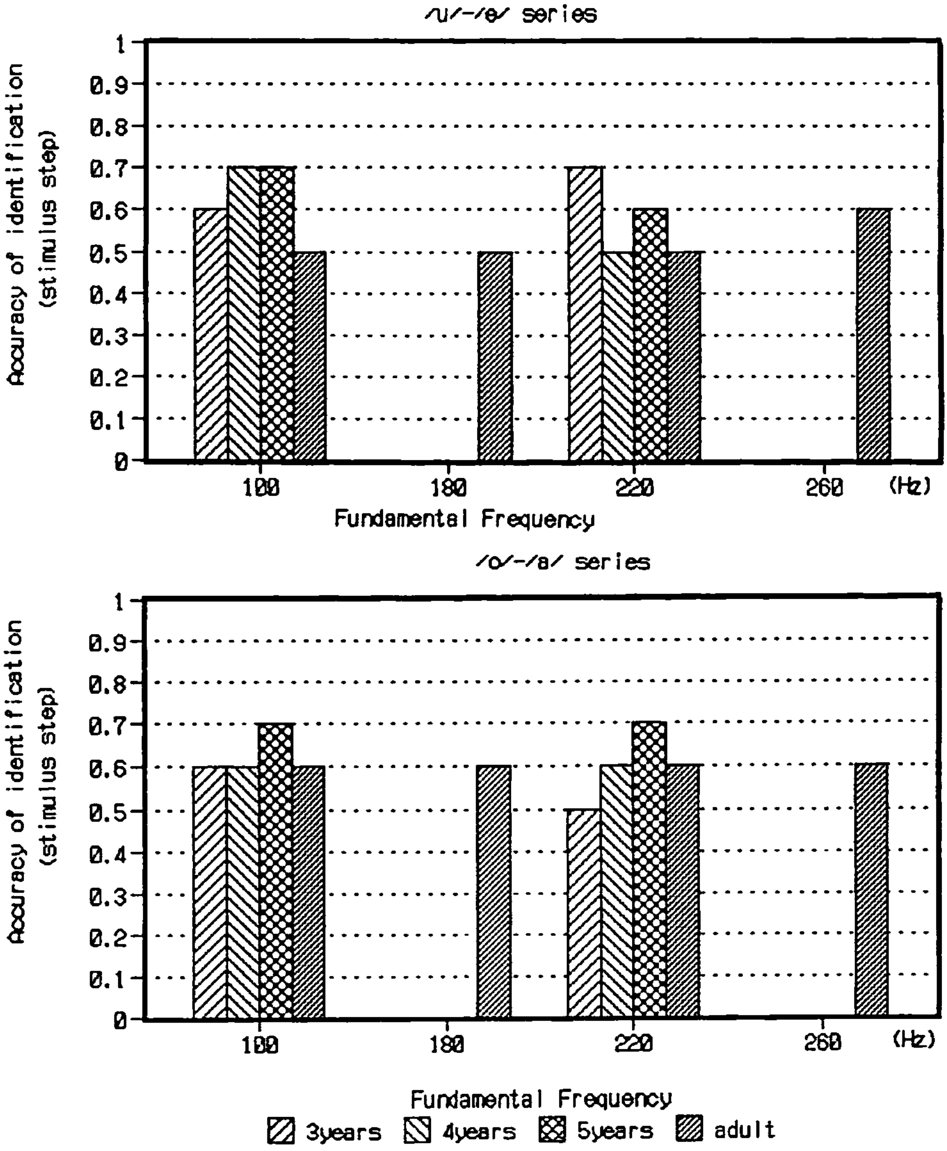

Fig. 3 Accuracy of identification for each group.

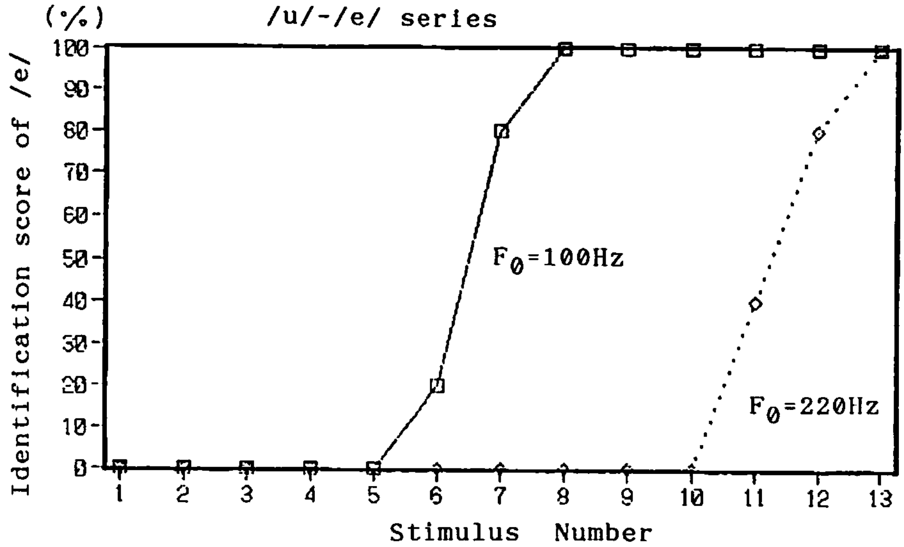

Fig. 1 An example of the identification curves for a child of 3 years.

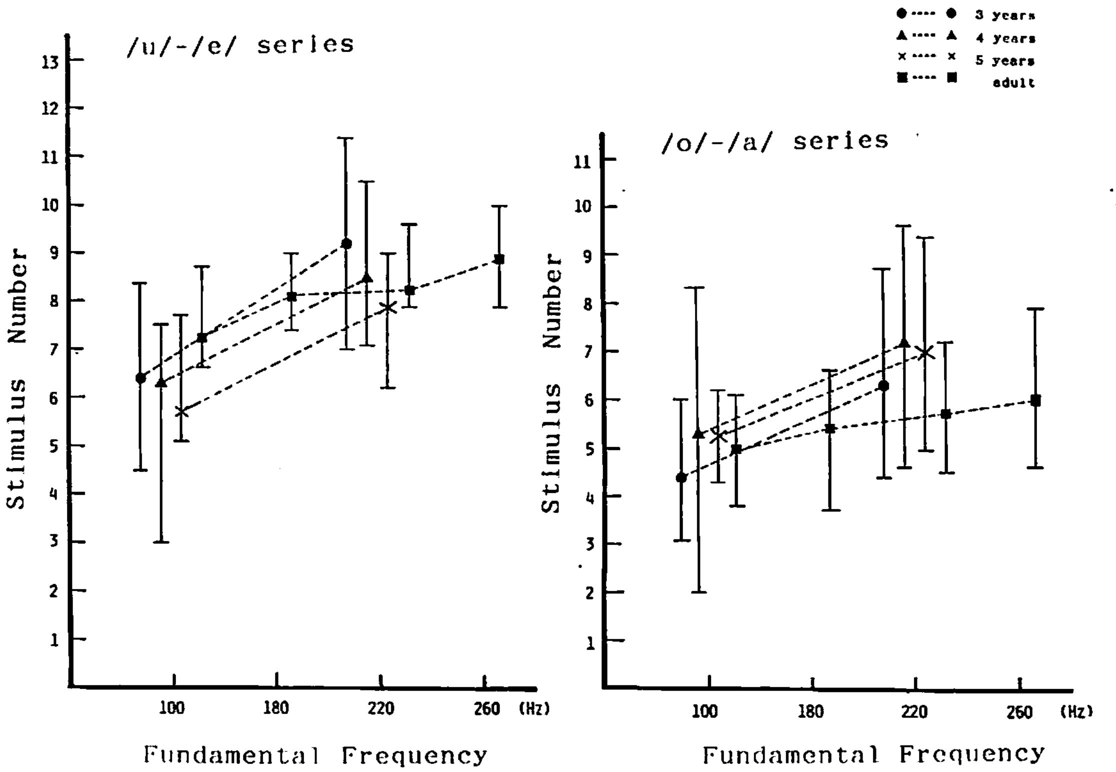

Fig. 2 Relationship between the fundamental frequency and the vowel boundaries for each subject group. The vertical bars represent the ranges.

Fig.4 shows the amount of boundary shift associated with the change in fundamental frequency from 100Hz to 220Hz (with concurrent changes in the higher formant frequencies). It is clear that the children showed a greater boundary shift than the adults. These results suggest that children aged 3-5 show a perceptual normalization of vocal tract size which is similar to that of adults. Furthermore, the data suggest that young children's responses to the change in fundamental frequency are more naive than those of adults. It appears that the effect of fundamental frequency becomes less for adults.

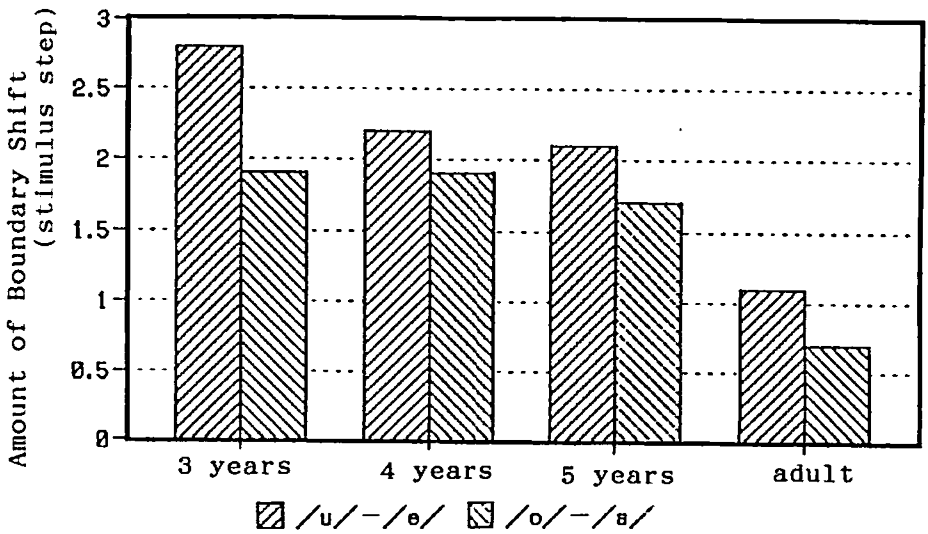

Fig. 4 Amount of boundary shift associated with the change in fundamental frequency from 100Hz to 220Hz.

Further experiments are now being conducted with infants using the technique of conditioned head-turn responses. Preliminary results show that infant can appropriately categorize the synthetic vowels used in the present experiment.

The experimental procedure is as follows. The infant held by the caretaker sit in a booth. A loudspeaker is located at a 90° - angle to the left of the infant and a visual reinforcer (a mechanical animal toy) is placed on the speaker. When activated, the mechanical toy starts to move. 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 the training phase, two synthetic vowels, representative /o/ and /a/, are used as stimulus sounds. One of the two synthetic vowels is presented every one second as the background stimulus. The sound is then changed to another stimulus sound, and, following the change of the stimulus sound, the reinforcer is activated. The infant learns to anticipate the reinforcer and make a head-turn with the change of the stimulus sound. In the second stage of the training phase, the stimulus sound is changed in half of the trials and not changed in the remaining half of the trials. The infant learns to turn only when the stimulus sound is changed.

In the experimental trials, the background stimulus is one of the representative vowels used in the training phase, but it is changed to an intermediate vowel. The probability of making a head-turn for each intermediate vowel is measured. Fig.5 shows an example of the experimental results obtained for a 6-month-old infant. The results show that this 6-month-old infant can categorize these synthetic vowels in an adequate way.

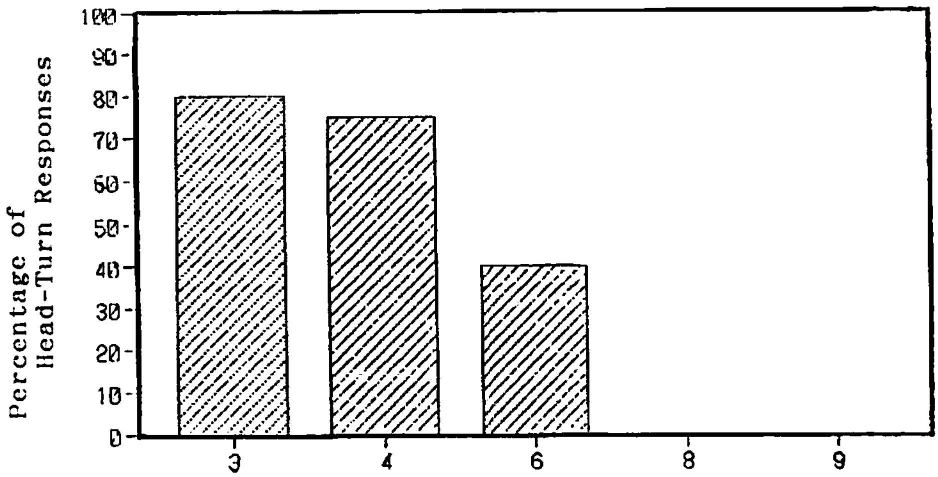

Fig. 5 Relationship between the percentage of head-turn responses and stimulus numbers in the /a/-/o/ series for a 6-month-old infant.

Using this experimental paradigm, further experiments are now being conducted on infants to measure the effect of changes in the fundamental frequency and the higher formants frequencies on the shift in vowel boundaries.

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 pitch and higher-formants in the perception of vowels, IEEE Transactions on Audio and Electroacoustics, AU-16(1), 1968.

3) Kuhl, P.K.: Speech perception in early infancy: Perceptual constancy for spectorally dissimilar vowel categories, J. Acoust. Soc. Am., 66, 1668-1679, 1979.