Ann. Bull. RILP, No. 11, 51-55 (1977)

THE SIMULATION OF CHILDREN'S SPEECH BY ADULT ACTRESSES*

I. Kirikae**, M. Sawashima, H. Hirose and H. Fujisaki

Introduction:

It has been established that children's speech is generally characterized by a higher fundamental frequency of the glottal source as well as by higher formant frequencies of vowels as compared to speech of normal adults. It is also known that perturbations in fundamental frequency and intensity of the glottal source are more pronounced in children's speech than in adults' speech. These acoustic characteristics of children's speech can be ascribed mainly to the anatomical and physiological properties of their vocal organs. In animation movies or in TV cartoons for young children, however, most of the roles that require children's voices are played by adult actresses who quite successfully simulate speech of young children. The principal purpose of the present study was to investigate the mechanisms by which these skilled actresses simulate children's speech.

Procedures:

Three adult actresses served as subjects in the present study. The ages of the subjects were 24, 28 and 29. Each subject was required to utter the five Japanese vowels /a/, /i/, /u/, /e/ and /o/ in isolation in her own natural voice (Series A) and subsequently in a simulated voice of a pre-school child (Series B). Each vowel sample was recorded for acoustic analysis, in which the sample was computer-processed and the fundamental frequency as well as frequencies of the first four formants (F1 - F4) were extracted using the method reported by Fujisaki [1]. During the production of each vowel sound, a lateral x-ray picture of the head and neck regions of the subject was taken together with a still picture of the anterior view of the face to compare the vocal tract configuration between the two series. For the purpose of comparison, measurements were made on the x-ray films to obtain the pertinent size of the vocal tract in the midsagittal plane according to Heinz and Stevens' method [2] and, eventually, the configuration for each vowel sound was represented by a straight tube.

Results:

1. Acoustic analysis

It was generally observed that the fundamental frequency (F0) and the frequencies of the extracted formants of the simulated children's vowels shifted higher when compared to the subject's natural vowels. Figure 1

---

* Paper to be presented at 17th International Congress of Logopedics and Phoniatrics. Copenhagen 1977.

** Department of Otorhinolaryngology, Jichi Medical School

shows the shift in each acoustic parameter for two different vowels /a/ and /u/. Open circles represent F0 and filled circles indicate F1, F2 and F3 (from bottom to top). For all the subjects examined, it is shown that F0 and all the formants tend to become higher when the subject changed her natural voice to simulate children's voices. In subject I, for example,

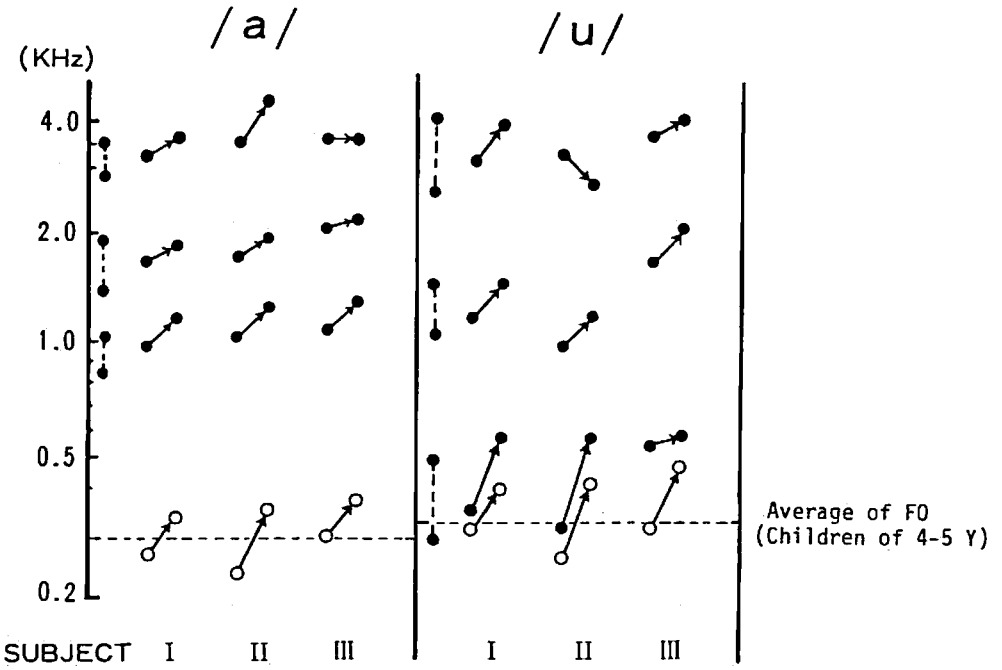

Fig. 1: Changes in acoustic parameters from the subject's natural voice to simulated children's voices. (Vowels /a/ and /u/).

F0 for vowel /a/ moved from 264 Hz to 333 Hz, F1 from 960 to 1160, F2 from 1660 to 1880 and F3 from 3220 to 3590. In each graph, a dotted horizontal line near 300 Hz level indicates the average of F0 for ordinary children of between 4 and 5 years of age. Pairs of filled circles connected by a vertical dashed line at the leftmost part of each graph indicate the average values of formant frequencies of adult female (lower circle), and those of children of betwwen 4 and 5 years (upper circle). These average values for children and adult females were obtained from Okamura [3]. It is noted that the shift in F0 is so marked that the value for simulated children's voice is even higher than the average of ordinary children. Each formant is also very high in simulated children's voices and it appears that the amount of formant shift is almost comparable to the difference in the pertinent formant frequencies between ordinary females and children. A similar tendency was observed for other vowels and in other subjects as well.

2. Configuration of the vocal tract

Figure 2 illustrates the tracing of the vocal tract configuration from the lateral x-ray pictures of the head and neck regions for subject I

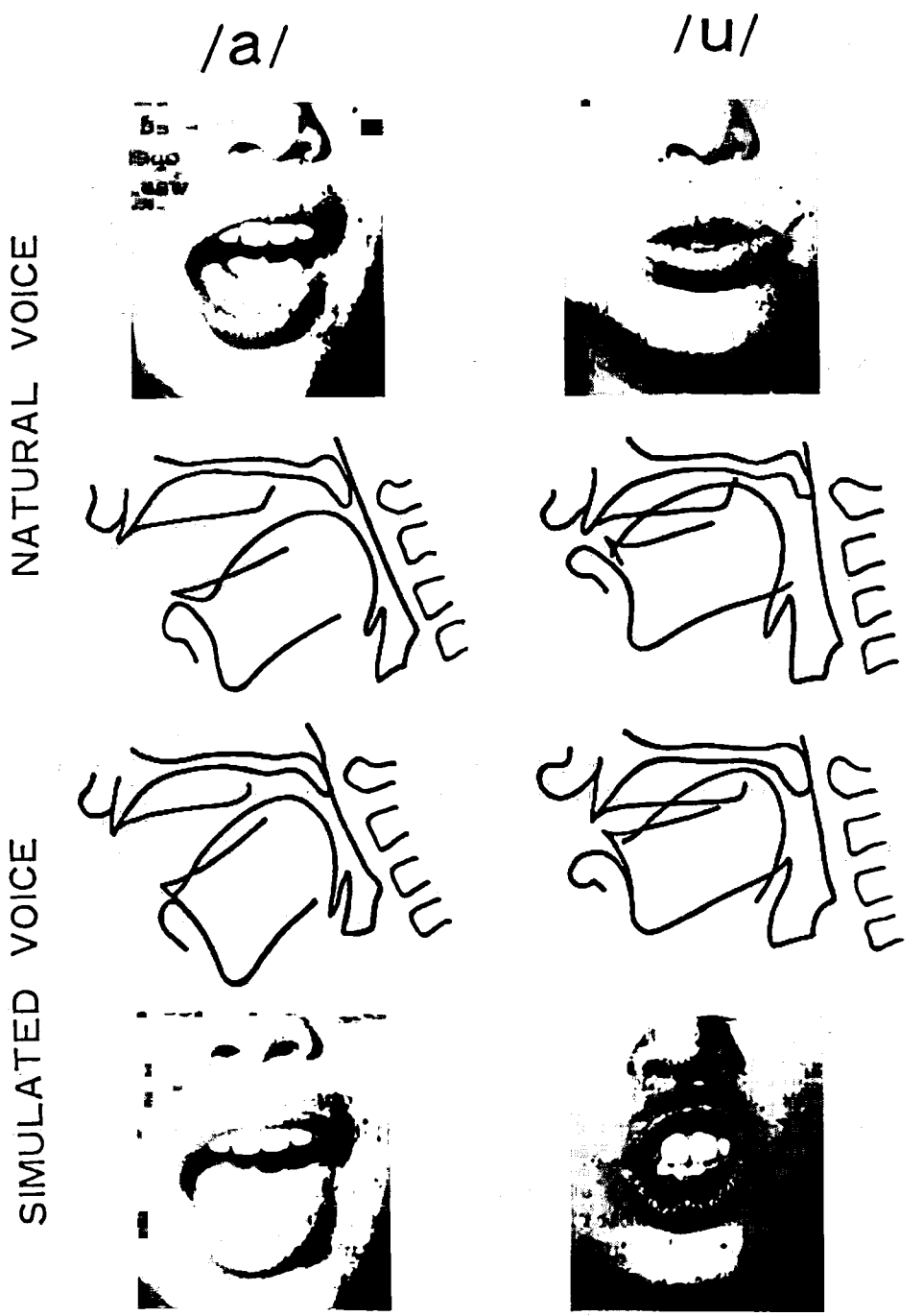

Fig. 2: Comparison of vocal tract configuration and articulatory gestures between natural voice (upper) and simulated voice (lower), (Vowels /a/ and /u/)

producing Japanese vowels /a/ and /u/, in her natural voice (upper) and in a voice simulating a child (lower). The still picture of the anterior view of the face during the production of the same vowel is also illustrated. It is shown that the larynx is elevated and the pharynx is shortened in the production of simulated voice. It is also revealed that the lip opening is wider in simulated voice for the vowel /u/. The same tendency was observed for the vowel /o/ which is not shown here. The elevation of the velum was apparently less marked in the case of simulated children's voices.

Examinations of the vocal tract configuration in the form of a straight tube revealed that the length of the vocal tract was shortened by approximately 1 cm. The point of maximum constriction of the tract shifted considerably toward the oral side, particularly in the vowels /a/ and /o/.

Comments:

The acoustic differences between speech of children and adults are based apparently on differences both in their phonatory and articulatory mechanisms. In particular, differences in the latter mechanism are characterized primarily by shorter vocal tract and larger volume ratio of oral to pharyngeal cavities in children as compared to adults. Simulation of children's speech by adults thus requires considerable adjustments on the part of adults both in the phonatory and in the articulatory processes.

To a skilled female adult with a relatively high mean voice fundamental frequency F0, however, phonatory adjustments for raising her F0 to the mean value for young children seemed to present little difficulty. In fact, all three subjects could raise their F0 well beyond the mean value of pre-school children of 4- and 5-year olds, thus exaggerating the "child-like" character of the simulated speech in this particular aspect.

Articulatory adjustments for simulating children's formant frequencies were found to be accomplished primarily by elevating the larynx so as to shorten the vocal tract and, at the same time, to decrease the size of the pharyngeal cavity. Further shortening of the vocal tract was also found to be accomplished by widening the lip opening particularly in the vowels /u/ and /o/. These articulatory adjustments were generally quite effective for simulating children's formant frequencies except for a few cases, where the shift in the third formant frequency was not as large as in the two lower formant frequencies.

In addition to these adjustments that are more or less related to simulation of static characteristics of children's speech, the subjects were also found to be controlling their articulatory movements to simulate dynamic characteristics of children's speech, i.e., by lowering the overall speech rate and by decelerating articulatory transitions involving certain consonants. Although we have not conducted quantitative analysis on the relative psychological significance of the static and the dynamic aspects, the latter certainly seems to be of considerable importance in imparting a child-like character to the simulated speech.

References

1. Fujisaki, H., N. Nakamura and K. Yoshimune (1970), Analysis, nominalization and recognition of sustained Japanese vowels, Jap. Jour. Acoust. Soc. 26: 152-154.

2. Heinz, J. M. and K. N. Stevens (1964), Derivation of area functions and acoustic spectra from cineradiographic films of speech. QPR, MIT. 74: 192-198.

3. Okamura, M. (1966), Acoustical studies on the Japanese vowel in children - The formant construction and the developmental process - , Jap. Jour. Otol. Tokyo 69: 1198-1214.