SPEECH PRODUCTION OF MANDARIN-SPEAKING CHILDREN WITH HEARING IMPAIRMENT AND NORMAL HEARING

Shu-Chuan Tseng

Institute of Linguistics, Academia Sinica, Taiwan
tsengsc@gate.sinica.edu.tw

ABSTRACT

This paper studies speech data of 45 children with hearing impairment and 79 children with normal hearing in terms of their speech intelligibility, vowel production, and tone production. Age played an important role in the intelligibility scores and vowel space size of the normally hearing group. In contrast, only the word comprehension ability had a significant effect on the intelligibility scores of the hearing-impaired group. Tone production was examined by means of F0 linearization and automatic k-means clustering. The results suggest that the normally hearing children produced tone shapes with a wider variety, closer to the adults' results. But differently, the hearing-impaired children with lower speech intelligibility mainly produced high-pitched and extremely monotonous F0 contours.

Keywords: vowel space, speech intelligibility, F0 contour

1. INTRODUCTION

Speech production has been assessed by different methods. One way to represent the degree of speech intelligibility is impressionistic judgments obtained by listener ratings of speech. Peng, et al. [7] showed that the percentage of correctly produced segments was highly correlated with rating-scale intelligibility of speech produced by children with hearing impairment. Acoustic properties can also be used to assess speech, in which acoustic patterns are specified for different speech phenomena. For example, hearing-impaired children with more reduced vowel spaces tended to have lower intelligibility scores, as proposed by Tseng, et al. [8]. Moreover, Bradlow, et al. [2] recruited 200 listeners to assess the intelligibility of 20 normally hearing adults and they found a negative correlation between the degree of vowel space reduction and the overall intelligibility. Liker, et al. [5] found a smaller and more fronted vowel space in hearing-impaired children compared with that of children without hearing impairment. Milder and Liker [6] reported similar vowel space in hearing-impaired and normally hearing children when examining nonsense syllables. In this study, we use both speech intelligibility and vowel space size to assess the speech production of our subjects and analyze their correlations with age, word comprehension ability and degree of hearing loss.

Lexical tones determine the form and meaning of the associated syllables [3]. Taiwan Mandarin has four lexical tones (T1: high level, T2: rising, T3: low and dipping, T4: falling) and one neutral tone (T5). Han, et al. [4] reported that the hearing-impaired children with cochlear implant produced 48% of tones correctly, and the age-matched normally hearing children produced 78% correctly. In the hearing-impaired group, the production of Tone 2 was most severely impaired, followed by Tone 3 and Tone 4. In this study, we adopt the automatic clustering method to analyze the complex F0 contour shapes produced by children with and without hearing impairment.

2. DATA PREPARATION

2.1. Subjects

The hearing-impaired group (the HI group) consists of 45 hearing-impaired children who were receiving regular auditory-verbal therapy in the Children's Hearing Foundation (CHF) at the time of recording. The normally hearing group (the NH group) consists of 79 children from the private Jengo Kindergarten in Taipei. We used the auditory memory level (AML) developed by the CHF to assess the word comprehension ability of the HI children. It is defined as the maximum number of "content words" a child can memorize from the designed sentence spoken to her/him. For subject selection, we only recruited children whose AML above 4 (the highest score is 5+). The AML of the HI group was assessed by the CHF. The NH children were assessed during the recording sessions following the standard procedure provided by the CHF.

2.2. Recording

Both groups of children were recorded in soundproof rooms. Eighteen sentences composed of words preschool children were familiar with were designed to cover all vowels and consonants in Mandarin. Word-picture tests (conducted shortly before or after the recording sessions) showed that all subjects understood the meaning of each target word. It was ruled out that the results presented in this study could be caused by hearing problem. Subsequently, the sentences read by a native female adult were presented by PowerPoint, in which each sound file was played twice with a meaning-matched cartoon picture shown on the screen for the children to repeat. We also recorded the same 18 sentences read by six adult speakers for comparing the tone production of the children with adults.

2.3. Data processing

The data were digitized with a sampling rate of 44.1 kHz, 16-bit, mono-track audio format. Three trained linguists segmented and labeled the syllabic boundaries. Transition discrimination and pause segmentation were the main labeling criteria. Each labeler was responsible for one-third of the data and double-checked the other labelers' data by turns. Data labeled inconsistently were discussed until all three labelers achieved a consensus. As a result, we labeled 4,080 syllables from the HI group and 7,512 syllables from the NH group.

3. METHOD

3.1. Speech intelligibility

Speech intelligibility was scored by adopting a 5-scale system: 1 to 5 to represent the poorest, poor, ordinary, good, and clearest. Three trained linguists gave a score from 1 to 5 to each child, making the minimum score 3 and the maximum score 15.

3.2. Vowel production

The most stable formant pattern is normally observed at the point of maximal intensity, because this point theoretically corresponds to the most sonorous location from the phonological point of view. Formants were therefore extracted for all tokens of /i, a, u/ in open syllables at the point where the maximal intensity was measured. The means of F1 and F2 values of each vowel were taken as the representative formant values for respective vowels. The statistical analysis was done on the normalized z-scores of the original measurements.

The shape and size of vowel space reflect the degree of contrast in the tongue height and position of the articulated vowels. Conventionally, the vowel space size is the area of the triangle determined by /i, a, u/. We adopted a slightly different calculation method. Vowel space size is defined as the area of the quadrangle determined by the maximum F1 and F2 differences, as defined in [8]. In normal cases, the quadrangle area is proportional to the triangle area. This adjustment can better identify the abnormality for problematic vowel space shapes. For example, eight out of our HI subjects produced only minimum contrasts in /i/ and /u/, while four of them even had /i/ and /u/ in the reverse position.

3.3. Acoustic features for modeling F0 shapes

F0 values were automatically extracted by PRAAT [1] and the stylized F0 contour was represented by two lines which optimally model the original F0 contour. Our feature set includes the slopes of the two lines, the duration ratio of each line in the syllable, the range of F0, and the z-score of the initial F0 value. The design of these acoustic features avoids influences from the discrepancy of pitch ranges resulting from individual speaker differences.

4. ASSESSMENT OF SPEECH PRODUCTION

4.1. Factors

We adopted the AML as a perceptual measure for representing the word comprehension ability of our subjects. As we only recruited children whose AML was above 4, the level ranged from 4, 5 to 5+ (represented as 6 in the analysis). The assessment of hearing-loss degrees was based on pure tone average thresholds, which were conducted by the CHF. It was mutually exclusive, ranging from mild to profound. In the statistical analysis, they were transformed as follows, 1: mild, 2: moderate, 3: moderate-severe, 4: severe, 5: severe-profound, 6: profound. Speech intelligibility and vowel space size were used as speech production measures.

4.2. Analysis

We examined the correlation between the two production measures and three factors: age, word comprehension ability, and degree of hearing loss.

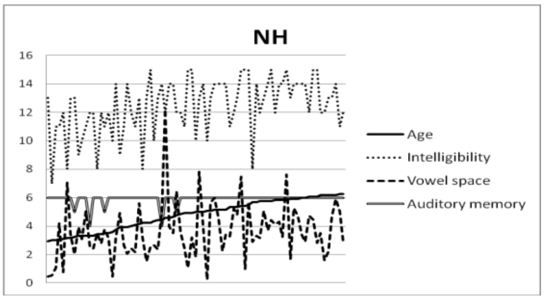
The means of both groups and the results of the 1-tailed Pearson correlation test are summarized in Table1. Please note that the significance level we refer to in this study is 0.05.

Table 1: Correlation results of speech intelligibility and vowel space with age, auditory memory level (AML), and degree of hearing loss (DHL).

	MEAN	Age	AML	DHL
HI (N=45)				
Speech intelligibility	10	.183 p=.115	.636* p=.000	.175 p=.126
Vowel space	2.7	.315* p=.018	.095 p=.268	-.076 p=.309
NH (N=79)				
Speech intelligibility	12.4	.459* p=.000	-.016 p=.444	
Vowel space	3.7	.239* p=.017	-.033 p=.386	

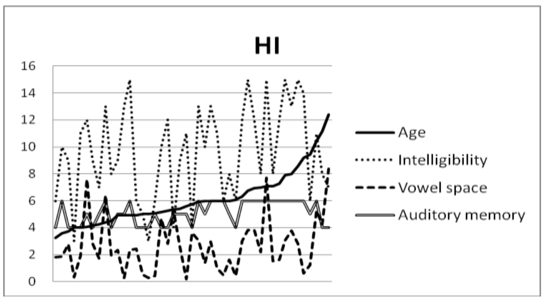
Age was correlated with speech intelligibility and vowel space size in the NH group, as shown in Figure 1. But only age had a significant effect on speech intelligibility in the result of ANOVA test, F(8, 70)=2.773, p=0.01.

Figure 1: Interaction of age, speech intelligibility, and vowel production of normally hearing children.



The HI group was different, as shown in Figure 2. The speech intelligibility was correlated with the AML instead of age. This suggests that the speech production ability of the HI children was related to the assessed comprehension ability. Also approved in the ANOVA test, AML is the only factor that had a significant effect on speech intelligibility in the HI group, F(12,32)=2.293, p=0.03.

Figure 2: Interaction of age, speech intelligibility, and vowel production of hearing-impaired children.



4.3. Vowel production

Intelligibility was used as an objective judgment to assess speech production ability. Does it correlate with the acoustic properties of vowels?

Figure 3: Speech intelligibility vs. vowel space size.

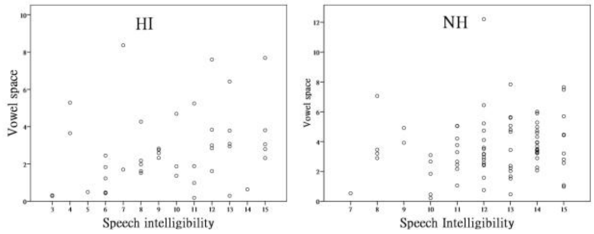


Figure 3 shows the scatterplots of vowel space size and speech intelligibility of both HI and NH groups. Statistically, vowel space size was correlated with speech intelligibility in the HI group (Pearson correlation coefficient=.290, p=.027). But it was not the case in the NH group (Pearson correlation coefficient=.161, p=.078). Nevertheless, the trend of a correlation is clear in both charts. This result may suggest that speech intelligibility and vowel production are two dependent, but different measures.

5. TONE PRODUCTION

5.1. Groups

Using the acoustic features defined in 3.3, we were not able to achieve a good clustering result for the HI subjects as a group, as the result cannot converge to a meaningful number of clusters. This may be due to the fact that the HI subjects form a diverse group in terms of their speech ability. Therefore, we divided the HI group into two subgroups according to their speech intelligibility scores. Those who were scored higher than 9 were classified into the high intelligibility group, the HI-H group. The others belonged to the low intelligibility group, the HI-L group. The final data used for the clustering experiment are summarized in Table 2.

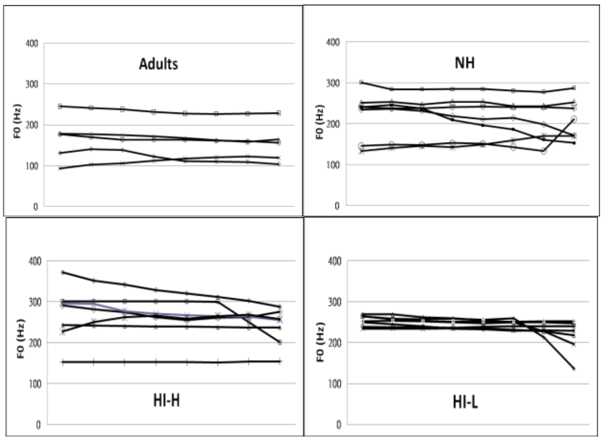
Table 2: Data summary.

	T1	T2	T3	T4	T5	Total
Adults	134	70	259	182	48	693
HI-H	408	215	779	553	149	2,104
HI-L	382	201	735	519	139	1,976
NH	1,433	764	2,798	1,993	524	7,512

5.2. F0 contour pattern

Figure 4 illustrates the F0 contours by plotting the original F0 values of syllable which are closest to each cluster centre. The duration of each syllable was divided into 10 measurement points and only eight out of them were plotted, leaving out the first and the last measurements.

Figure 4: Tone shapes of adults, normally hearing children, and hearing-impaired children.



The results show that the F0 shapes of the six adults were classified into five contours: one high-pitched shape (slightly falling), three mid-pitched shapes (two slightly falling, one slightly rising), and one low-pitched shape (slightly rising). These surface forms seem to reflect the general intonation declination effect, because most of them have a falling tendency. In contrast, the NH children show more diverse F0 shapes. But we still can classify them into one high-, four mid-, and two low-pitched contours. The F0 shapes produced by the HI-H children were slightly higher than the NH group. And the only one low-pitched contour was very flat. The F0 contours were relatively flat, but they were roughly located in the three pitch areas as in the adults and the NH group. However, the HI children with lower intelligibility produced extremely flat contours, mainly centered in the high-frequency area. This preliminary result of tone shape modeling demonstrates that hearing-impaired children with higher intelligibility scores were able to produce similar F0 shapes like the normally hearing children. Only the rising contour was missing. But those with lower intelligibility scores had more problems in producing F0 shapes in mid- and low-pitched areas. The contours were extremely monotonous, lacking falling and rising varieties.

6. CONCLUSION

Mandarin-speaking children with normal hearing and hearing impairment were studied in terms of their speech intelligibility, vowel production, and tone production. Age played a more important role for children with typical speech development. Word comprehension ability was correlated with speech intelligibility of the HI children, suggesting a relationship between speech input and output. Differences in tone shapes were preliminarily identified. Further analysis on the interplay of cluster labels and the associated canonical tones is needed to gain new insights into how tones are produced in authentic speech communication.

7. ACKNOWLEDGMENTS

This study was financially supported by the Children's Hearing Foundation and the National Science Council of Taiwan, under grant NSC 99-2410-H-001-097. We are grateful to the teachers of the Children's Hearing Foundation and the Jengo Kindergarten for their kind help with the data collection. We also thank Yifen Liu for conducting the F0 linearization and clustering.

8. REFERENCES

[1] Boersma, P., Weenink, D. Praat: doing phonetics by computer. Version 5.0.43. *http://www.praat.org/*

[2] Bradlow, A.R., Torretta, G.M., Pisono, D.B. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20, 255-272.

[3] Duanmu, S. 2000. *Phonology of Standard Chinese*. Oxford University Press.

[4] Han, D., Zhou, N., Li, Y., Chen, X., Zhao, X., Xu, L. 2007. Tone production of Mandarin Chinese speaking children with cochlear implants. *International Journal of Pediatric Otohinolaryngolog* 71, 875-80.

[5] Liker, M., Mildner, V., Sindija, B. 2007. Acoustic analysis of the speech of children with cochlear implants: A longitudinal study. *Clinical Linguistics & Phonetics* 21, 1-11.

[6] Mildner, V., Liker, M. 2008. Fricatives, affricates, and vowels in Croatian children with cochlear implants. *Clinical Linguistics & Phonetics* 22, 845-856.

[7] Peng, S., Spencer, L.J., Tomblin, J.B. 2004. Speech intelligibility of pediatric cochlear implant recipients with 7 years of device experience. *Journal of Speech, Language, and Hearing Research* 47, 1227-36.

[8] Tseng, S.-C., Kuei, K., Tsou, P.-C. 2011. Acoustic characteristics of vowels and plosives/affricates of Mandarin-speaking hearing-impaired children. *Clinical Linguistics & Phonetics* (to appear).