Proceedings of DiSS-LPSS Joint Workshop 2010,
25-26 September 2010, Tokyo, Japan, pp. 11-14.

# Assessing sentence repetition and narrative speech data produced by hearing-impaired and normally hearing children

*Shu-Chuan Tseng, Pei-Chen Tsou, Ko Kuei, Chien-Wen Lee*

Institute of Linguistics, Academia Sinica, Taiwan
{tsengsc, amber910, grace05, kiwi1101}@gate.sinica.edu.tw

## Abstract

This paper examines sentence repetition and narrative speech data produced by hearing-impaired and normally hearing children with matched gender, age and level of speech comprehension. We assessed these two kinds of speech styles by talker intelligibility, vowel space, and spike production in plosives. In both speaking styles, normally hearing children performed better in talker intelligibility than their hearing-impaired counterparts. No clear vowel space shrinkage was observed in respect of speech style, hearing impairment, and age group. Surprisingly, the production of the spike in plosives was a useful measure for distinguishing acoustic properties of different speaking styles and hearing ability.

Index Terms: Speech assessment, hearing impairment, speaking style, acoustic properties

## 1. Introduction

Speech production can be assessed by two main criteria: the degree of speech intelligibility (a kind of impressionistic judgment) and the measurement of acoustic properties. Bradlow *et al.* [1] recruited 200 listeners to assess their intelligibility of sentences produced by 20 normally hearing English-speaking adults. They found a negative correlation between the degree of vowel space reduction and the overall intelligibility. Speakers with more reduced vowel spaces tended to have lower intelligibility scores. Consonant production is also correlated with talker intelligibility, as stated in Shriberg *et al.* [2].

Ferguson and Kewley-Port reported in a study of different styles of adult speech that in spite of individual differences, longer vowel duration, expanded vowel space and raised F2 for front vowels were found in clear speech [3]. For preschool children, Redford and Gildersleeve-Neumann [4] examined segmental and supra-segmental features of clear and casual speech elicited from children whose ages were 3, 4, and 5 years old and found that F1 and F2 values were similar regardless of speech style. But Liker *et al.* found a smaller and more fronted vowel space in hearing-impaired children by comparing the vowel space size of children with and without hearing impairment [5]. After receiving an implant, a steady and systematic expansion of vowel inventories was demonstrated in the speech of Cantonese-speaking preschool children [6]. Milder and Liker reported similar vowel space in hearing-impaired and normally hearing children when examining the nonsense syllables which were designed and recorded to investigate formant-defined vowel space [7].

With regard to consonants, Peng and colleagues [8] studied the phonemic inventories of syllable-initial consonants in Mandarin-speaking children with cochlear implants and found that the average correct percentage was highest for plosives (77.8%), followed by nasals, affricates, fricatives, and the lateral approximant. Lee [9] concluded that the voice onset time cannot be used to distinguish whether a plosive is aspirated or not in the speech produced by hearing-impaired children. Khouw and Ciocca studied the voice onset time of plosives of ten Cantonese-speaking hearing-impaired adolescents, too and could not find any significant difference between the voice onset time of aspirated and unaspirated plosives [10].

Thus, we collected, to a certain degree, realistic speech data of two different speaking styles: sentence repetition and narrative, produced by hearing-impaired and normally hearing children. As plosives are the most likely to be produced correctly by Mandarin-speaking hearing-impaired children and the voice onset time does not seem to be useful for assessing plosive production, we will analyze the talker intelligibility, vowel space size and spike production in plosives. The purpose of our study is to examine if the above mentioned measures are useful to assess the speech ability of normally hearing and hearing-impaired children.

## 2. Method

### 2.1. Subjects

Two groups of Mandarin-speaking preschool children aged from 4 to 6 participated in this study with matched gender and age. The hearing-impaired group (hereafter the HI group) contains six hearing-impaired children who were receiving regular auditory-verbal therapy in the Children's Hearing Foundation at the time of recording. The normally hearing group (hereafter the NH group) consists of six age- and gender-matched children with no known hearing problems from the Jengo Kindergarten in Taipei. In order to control the language ability of both groups of children, we adopted the auditory memory level, developed by the Children's Hearing Foundation to assess the comprehension level of their hearing-impaired children. The level of auditory memory is defined as the "content words" a child can memorize from the designed sentence spoken to her/him. We conducted the same test on the NH group of children following the standard procedure instructed by the Children's Hearing Foundation. To ensure that both HI and NH groups of children have a comparable language comprehension ability, only children whose auditory memory level was 5 or 5+ (the highest) were included in this study. Details of the subjects are listed in Table 1.

Table 1. *Subject details.*
(HA: traditional hearing aids, CI: cochlear implants)

| Group | Subject | Age | Sex | Age at test | Auditory memory | Hearing aids |
|---|---|---|---|---|---|---|
| NH_4 | NH_4_M | 4 | M | 4;7 | 5+ | |
| | NH_4_F | 4 | F | 4;11 | 5+ | |
| HI_4 | HI_4_M | 4 | M | 4;5 | 5+ | HA |
| | HI_4_F | 4 | F | 4;11 | 5 | HA |
| NH_5 | NH_5_M | 5 | M | 5;4 | 5+ | |
| | NH_5_F | 5 | F | 5;11 | 5+ | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **HI_5** | HI_5_M | 5 | M | 5;4 | 5 | HA |
| | HI_5_F | 5 | F | 5;11 | 5+ | HA |
| **NH_6** | NH_6_M | 6 | M | 6;0 | 5+ | |
| | NH_6_F | 6 | F | 6;1 | 5+ | |
| **HI_6** | HI_6_M | 6 | M | 6;0 | 5+ | HA |
| | HI_6_F | 6 | F | 6;0 | 5 | CI |

### 2.2. Recording

Both groups of children were recorded in soundproof rooms, with the HI children in their therapy session and the NH children in the phonetics lab at Academia Sinica. Eighteen sentences were designed to contain vowels and consonants in Taiwan Mandarin, composed of words preschool children were familiar with. In addition, we have carried out a word-picture recognition test before recording to make sure that the children understand the meaning of words used in the sentences. 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. As a result, 107 valid sentences of the NH group and 103 of the HI group were used for analysis. For collecting narrative data, we used a set of six picture cards illustrating the story *The Tortoise and the Hare* as material. The children were told to arrange the cards in the correct order before they started telling the story.

### 2.3. Data processing

The data were digitized with a sampling rate of 44.1 kHz, 16-bit, mono-track audio format. We processed the data for phonetic analysis by using PRAAT v.5.0.43 [11]. Three trained linguists segmented and labeled the syllabic and segmental boundaries as well as the articulation phases (the occlusion, the transient, the frication, the aspiration, and the transition to the vocalic part) of plosives in the data. 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 which were labeled inconsistently were discussed until all three labelers achieved a consensus. The summary of the labeled data is given in Table 2.

Table 2. *Labeled data.*

| | **Sentence repetition** | | **Narrative** | |
|---|---|---|---|---|
| | **NH** | **HI** | **NH** | **HI** |
| **# of syllables** | 582 | 566 | 466 | 453 |
| **# of segments** | 1,372 | 1,319 | 680 | 924 |
| **# of plosives** | 186 | 183 | 236 | 328 |

### 2.4. Vowel space

The formant values of quantal vowels /i, a, u/ in open syllables at the maximal intensity point within the syllable were measured. The reason for this measurement is that the point of maximal intensity should be the most sonorous location from the phonological point of view. It is only applicable for monophthongs. The means of F1, F2 values of each vowel were calculated as the representative formant values. Normalized by Z-score ($z=(x-\mu)/\sigma$, where x is the measured value, $\mu$ is mean and $\sigma$ the standard deviation.), the size of vowel space of each children in each speech style was subsequently calculated. Conventionally, the vowel space size is the triangle area determined by /i, a, u/. In this study, in order to display the maximal contrast of the tongue height and position, we multiplied the difference of F2 values of /i/ and /u/ which indicated the tongue position contrast represented by $|F2(i) - F2(u)|$, with the difference between the minimum F1 values of /i/ and /u/ and the F1 value of /a/ which reflects the maximal tongue height contrast represented by $|\min(F1(i), F1(u)) - F1(a)|$. The vowel square emphasizes the maximal F1 and F2 differences and it illustrates better the abnormal vowel positions produced by the HI children [12].

## 3. Results

### 3.1. Talker intelligibility

Acoustic properties of segmental and supra-segmental features reflect the clarity of articulation in terms of measurements of physical signals. For perception, an impressionistic assessment with a 5-scale rating was conducted by three trained linguists to obtain an objective index to score the talker intelligibility of each child. 1 represents the poorest intelligibility; 5 for the clearest. Thus, the maximum score is 15 and the minimum at 3 for each child. The Pearson correlation coefficients among the three linguists were all above 0.9, $p<0.01$.

Figure 1 shows the mean of talker intelligibility scored for sentence repetition and narrative data of each group of children. The NH children outperformed their age-matched HI counterparts in talker intelligibility. We observed a tendency, though not statistically significant, that the speech of the NH children is clearer than that of the HI children regardless of speaking styles. But no clear tendency can be identified in terms of talker intelligibility in the two speaking styles. However, we noticed that the more spontaneous narrative speech seems to be more clearly perceived than the sentence repetition speech. Only in the group of age 4, sentence repetitions were reported more clear than the narratives. We chose our subjects based on the auditory memory level, i.e. the level of language comprehension ability was controlled. This may be the reason why there is no significant difference among the two groups of children.

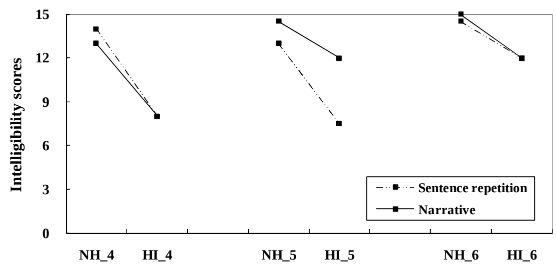



Figure 1: *Results of talker intelligibility.*

### 3.2. Vowel space size

The shape and size of vowel space determined by the acoustic characteristic are supposed to reflect the degree of articulated contrast in the tongue height and position of vowels. The larger the vowel space size is, the clearer the vowel contrast will be. In this current study, the vowel space size of the NH children does not present any superior performance than that of the HI children, as shown in Figure 2. Age and speaking style have no effect on the vowel space size, either. No vowel space shrinkage was identified in the more spontaneous speaking style. Figure 2 illustrates that the vowel space size varies to a great extent and no clear tendency is observed.

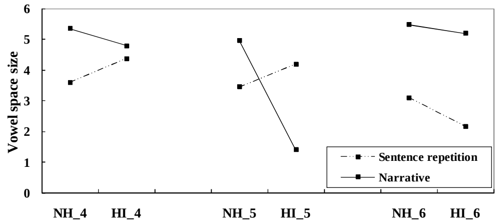
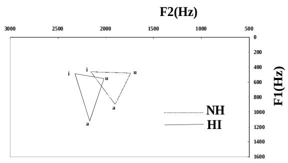



Figure 2: *Results of vowel space size.*

In further observation, Figure 3 shows the vowel space by taking the means of the NH and the HI groups. Fronted vowel space shape is observed in the HI group in both speaking styles, compared with the NH group. This difference is clearer in the sentence repetition data. In the narrative data, the vowels /a, u/ of the HI group were fronted. Certainly, more data are needed to examine the change of vowel space more closely.



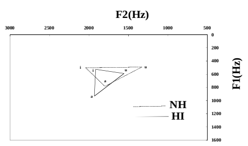




Figure 3a: *Sentence repetition.* Figure 3b*: Narrative*

### 3.3. Articulation phases of plosives

There are six plosives /p, t, k, $p^h$, $t^h$, $k^h$/ in Taiwan Mandarin [13]. Typical production of plosives contains five articulation phases which can be clearly identified in the spectrogram: the occlusion, the transient, the frication, the aspiration, and the transition (to the vocalic part) phases [14, 15]. In our data, these articulation phases were labeled to 933 plosives. Formant patterns and intensity in spectrogram were used as primary cues to identify boundaries. The majority of plosives labeled in our data appear in the following three types:

- Type 1: No clear boundaries within the plosives can be identified except from the transition phase to the vocalic part.
- Type 2: Most phases can be distinguished except the boundary between the frication and the aspiration phases. This is the finest articulation variant of the plosives and can be regarded as an index for proper pronunciation of plosives.
- Type 3: Only the occlusion and the transition phases can be identified.

Table 3. *Articulation types of plosives.*

| | | Type 1 | Type 2 | Type 3 | Other types |
|---|---|---|---|---|---|
| NH_4 | Sentence | 12.7% | 39.7% | 39.7% | 7.9% |
| | Narrative | 10.7% | 26.8% | 62.5% | 0.0% |
| HI_4 | Sentence | 65.6% | 6.6% | 0.0% | 27.9% |
| | Narrative | 76.1% | 0.0% | 6.0% | 17.9% |
| NH_5 | Sentence | 10.0% | 33.3% | 50.0% | 6.7% |
| | Narrative | 22.2% | 17.5% | 47.6% | 12.7% |
| HI_5 | Sentence | 66.1% | 21.0% | 0.0% | 12.9% |
| | Narrative | 62.0% | 3.9% | 21.7% | 12.4% |
| NH_6 | Sentence | 14.1% | 40.6% | 43.8% | 1.6% |
| | Narrative | 23.9% | 41.0% | 25.6% | 9.4% |
| HI_6 | Sentence | 76.3% | 11.9% | 0.0% | 11.9% |
| | Narrative | 76.8% | 1.2% | 13.4% | 8.5% |

Table 3 shows the percentage of the articulation types in both speaking styles. As Type 2 is regarded as the finest articulation variant, we analyzed the percentage of Type 2 for each group of children. First of all, the percentage of Type 2 is consistently higher in the sentence repetition data than in the narrative data. Different from the size of vowel space, Type 2 production of plosives is a solid cue illustrating a clear difference between sentence repetition and narrative speech styles. The greatest difference (17.1%) was performed in HI_5, while there was only a minor difference (0.4%) in NH_6.

With regard to subject groups, the percentages of Type 2 were clearly higher in the NH group than in the HI group. This is also shown in Table 3. Taking the average of the Type 2 percentage of the NH and HI groups in the sentence repetition data, we respectively obtained 37.9% and 13.2%. Similarly, in the narrative data, the average percentage of Type 2 of the NH group was higher (28.4%) than the HI group (1.7%). In summary, the difference between the HI and the NH groups in both speaking styles can be clearly presented in terms of Type 2 production of plosives.

### 3.4. Spike production

After analyzing the types presented in the previous section, we came to a conclusion that the main characteristic of Type 2 production does not solely lie in the fine distinction of pronunciation phases, but the appearance of the spike. Only in Type 2, the spike can be identified. This implies that the indicator for assessing pronunciation clarity can be further simplified to the appearance of the spike. Figure 4 illustrates the percentage of spike occurrences in plosives produced by both groups of children. The higher the percentage is, the better the production of plosives. As shown in Figure 4, NH_6 performed the highest percentage in both speaking styles while the lowest was found in HI_4. The results of the Chi-square test showed that the percentage of spike production of three age groups did show a statistically significant difference (NH $\chi^2(2)=6.48$, $p<.05$ and HI children $\chi^2(2)=9.63$, $p<.01$). The effect of speaking styles was only significant in the HI group $\chi^2(1)=29.07$, $p<.001$. For the NH group, different speaking styles did not result in significant differences in the pronunciation of plosives.

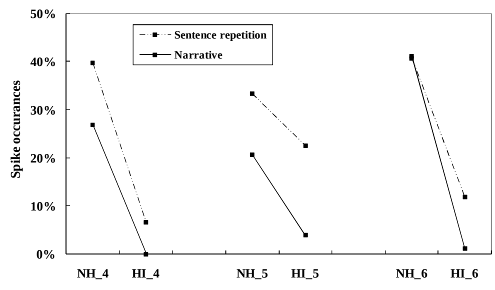



Figure 4: *Results of spike production in plosives.*

## 4. Discussion

We scored talker intelligibility to assess the degree of pronunciation clarity, which is mainly based on the perception of listeners. And we measured the frequency of the spike in plosives by means of acoustic analysis. Both indices reflect differences in age and hearing ability, though to different extents. But vowel production seems to develop quite inconsistently in different stages of child language development, as we found no consistent tendency in terms of the shape and size of vowel space. Also, we did not find clear vowel space shrinkage in the more spontaneous speaking style. It was noted that F2 was not quite stable when children were producing /i/ and this may due to the poor control of the tongue position. From the physiological point of view, it was reported that not until 6-8 years of age, the full human 1:1 vocal tract would be achieved. That is, at the point the vocal tract proportion had been achieved, one would be able to produce the quantal vowels /i, a, u/ with a more stable formant frequency pattern [16]. However, in the process of development, it was observed that various speech production characteristics would not develop at the same rate or in the same fashion among individual children [17].

In the production of plosives, we did not find any token of plosives, in which all five articulation phases can be clearly separated, as we found in the adult speech data. Further studies are needed to clarify whether this is a common phenomenon in children's phonological development. The finest articulation type, in which only the boundary between the frication and the aspiration is merged, makes up around 37.9% and 13.2% of the labeled plosives in the sentence repetition data of the NH and HI children, and 28.4% and 1.7% in the narrative data of the NH and HI groups. These results clearly suggest that this kind of measure is useful for distinguishing different speaking styles and subject groups.

Our finding also shows that in all age groups, the NH group outperformed the HI group in the occurrences of the spike production. The result is in line with previous studies suggesting that children gradually develop fine speech motor control and acquire the phoneme-specific articulator movement differences [18, 19]. In such case, children's articulator movements become more stable and decrease in variability over time. As shown in Figure 4, the difference of the spike production of the NH_4 and NH_5 children in sentence repetition and narrative data was about 10%, but levels of variability leveled off in the NH_6 group. But this tendency is not observed in the HI group. The performance in terms of the spike production distinguishes sentence repetition from narrative speech styles in the HI group.

# 5. Conclusions

Effects of age and hearing impairment were found in terms of talker intelligibility scores and percentage of spike production in plosives. Vowel production is relatively unstable in child speech production, as the size of vowel space does not show a consistent tendency across groups of age or hearing impairment. It is unlikely to use a single measure to assess normally hearing children's speech ability because of individual differences of the children and the different development stages in child language. In the case of including hearing-impaired children for comparison, more indicators are certainly necessary to represent the differences. Our preliminary results suggest that talker intelligibility and spike production may be useful for assessing the performance of speech production of children with and without hearing impairment. Although we have controlled age and language comprehension ability (via auditory memory), the sample of data is small and diverse in terms of different hearing aids worn by the HI children. We are currently processing more data for future studies.

# 6. Acknowledgements

This study was financially supported by the Children's Hearing Foundation and the National Science Council of Taiwan, under grant NSC 98-2410-H-001-066. We would like to thank the teachers and technicians of the Children's Hearing Foundation and Jengo Kindergarten for their help.

# 7. References

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

[2] Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L. and Wilson, D. L., "The percentage of consonants correct (PCC) metric: extensions and reliability data", Journal of Speech, Language, and Hearing Research, 40: 708-722, 1997.

[3] Ferguson, S. H. and Kewley-Port, D., "Talker Differences in clear and conversational speech: Acoustic characteristics of vowels", Journal of Speech, Language, and Hearing Research, 50: 1241-1255, 2007.

[4] Redford, M. A. and Gildersleeve-Neumann, C. E., "The development of distinct speaking styles in preschool children", Journal of Speech, Language, and Hearing Research, 52: 1434-1448, 2009.

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

[6] Barry, J. G., Blamey, P. J. and Fletcher, J., "Factors affecting the acquisition of vowel phonemes by pre-linguistically deafened cochlear implant users learning Cantonese", Clinical Linguistics and Phonetics, 20: 761-780, 2006.

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

[8] Peng, S.-C., Weiss, A. L., Cheung, H. and Lin, Y.-S., "Consonant production and language skills in Mandarin-speaking children with cochlear implants", Archives of Otolaryngology-Head & Neck Surgery, 130: 592-597, 2004.

[9] Lee, P., "Intelligibility of aspirated and unaspirated words by hearing-impaired students", Special Education and Rehabilitation, 9: 133-158, 2001.

[10] Khouw, E. and Ciocca, V., "An acoustic and perceptual study of initial stops produced by profoundly hearing impaired adolescents", Clinical Linguistics and Phonetics, 21: 13-27, 2007.

[11] Boersma, P. and Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 5.0.43 from http://www.praat.org/, retrieved 9 December 2008.

[12] Tseng, S.-C., Tsou, P.-C., and Kuei, K., "Measuring vowel space of Chinese-speaking hearing-impaired children", Paper presented at the 6th Asia Pacific Conference on Speech, Language, and Hearing, Honolulu, USA, 2009.

[13] Ho, D., "Concept and Method in Phonology", Daah Publisher, 1996.

[14] Kent, R. D. and Read, C., "The Acoustic Analysis of Speech", Singular, 2002.

[15] Lieberman, P. and Blumstein, S. E., "Speech Physiology, Speech Perception, and Acoustic Phonetics", Cambridge University Press, 1988.

[16] Lieberman, P. "The evolution of human speech. Its anatomical and neural bases", Current Anthropology, 48: 39-66, 2007.

[17] Smith, B. L. and Kenney, M. K., "An assessment of several acoustic parameters in children's speech production development: longitudinal data", Journal of Phonetics, 26:98-108, 1998.

[18] Grigos, M. I., Saxman, J. H. and Gordon, A. M., "Speech motor development during acquisition of the voicing contrast," Journal of Speech, Language, and Hearing Research, 48: 739-752, 2005.

[19] Cheng, H. Y., Murdoch, B. E. and Scott, D., "Physiologic development of tongue-jaw coordination from childhood to adulthood", Journal of Speech, Language, and Hearing Research, 50: 352-360, 2007.