ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Revisiting Consonant Acquisition in Typically Developing Chinese-Speaking Children With Insights Into a Multiword Data Set of Hearing and Deaf/Hard of Hearing Children

Shu-Chuan Tseng[a]

[a] Institute of Linguistics, Academia Sinica, Nankang, Taipei, Taiwan

ARTICLE INFO

Article History:
Received February 19, 2025
Revision received June 28, 2025
Accepted January 8, 2026

Editor-in-Chief: Jessica E. Huber
Editor: Leah Catherine Fabiano

https://doi.org/10.1044/2026_JSLHR-25-00124

ABSTRACT

Purpose: This study aimed to advance the understanding of consonant acquisition with quantitative and qualitative evidence from various groups of Chinese-speaking children. Normative patterns of phonological development of consonants were affirmed by utilizing phoneme transcription and perceptual judgment of a single-word normative data set, followed by analyses of comparable characteristics of a multiword data set of hearing and deaf/hard of hearing children.

Method: The single-word normative data set comprised 798 typically developing Chinese-speaking children, whereas the multiword data set consisted of 79 normal hearing and 45 deaf/hard of hearing children. The percentage of consonants correct (PCC) was derived from phonemes transcribed by automatic alignment and human verification. Perceptual acceptability/intelligibility ratings include the percentage of correctly produced words (AccWord) in the normative data set and the intelligibility scores (IntScore) in the multiword data set. Distribution and correlation of PCC and AccWord/IntScore, as well as consonant error patterns, were examined and compared.

Results: Developmental patterns and phonological aspects of consonant acquisition in Chinese-speaking children were thoroughly reported. PCC was significantly correlated with AccWord/IntScore across all subject groups in both single-word and multiword data sets. This finding suggested that PCC can indicate speech performance above the phoneme level. In all subject groups, stopping errors occurred more frequently than frication, the accuracy rates of retroflex sounds were low, and there was a mixed use of /n, l, ʐ/.

Conclusions: The current study featured developmental growth curves, error analysis, and possible clinical applications of a wordlist-based normative data set as reference standards. The fact that PCC is correlated with acceptability/intelligibility ratings across data sets and subject groups supports its efficacy as a quantitative indicator of child speech assessment.

A wide variety of phonetic and phonological features can be used to assess the ability to articulate conceptualized, formulated speech in verbal communication (Levelt, 1989; Shriberg et al., 1997). For children, usage-based empirical metrics should reflect their performance in segment production, word formation, and prosodic organization. From a methodological point of view, transcription and acceptability or intelligibility judgments regarding phonemic and suprasegmental properties are essential, but still, research outcomes derived from specific annotation tasks may be affected by discourse context, phonological environment, and biases arising from recording quality and instructions, and so forth (Lockart & McLeod, 2013; Shriberg et al., 2020; Wong et al., 2005). Developmental patterns describing spoken language abilities in typically

Correspondence to Shu-Chuan Tseng: tsengsc@gate.sinica.edu.tw. ***Disclosure:*** *The author has declared that no competing financial or nonfinancial interests existed at the time of publication.*

developing children in designated age groups are commonly used as normative references, for example, for Arabic (Amayreh & Dyson, 1998), Cantonese (So & Dodd, 1995; To et al., 2013), Dutch (Priester et al., 2011), English (Dodd et al., 2003; Sander, 1972), Korean (Kim et al., 2017), Vietnamese (Lee et al., 2024), Putonghua (Beijing Mandarin; Hua & Dodd, 2000; Wu et al., 2020), and for recent reviews of Mandarin and English phonological development studies (Crowe & McLeod, 2020; Li & To, 2017). The data sizes of these normative data sets varied. Concerning participant number and annotation granularity, the normative data set presented in this article would exceed those of Mandarin (including Beijing and Taiwan) mentioned in Li and To (2017).

Annotated child speech data sets as used for acquisition research are needed for studies related to assistive technology, for example, computational analysis and models (Räsänen, 2012; Romani et al., 2017), child speech recognition (Jain et al., 2023), classification tasks (Tseng et al., 2023), and speech assessment tools (Usha & Alex, 2023; Zhang et al., 2020). This study aimed to present linguistically motivated metrics and data processing methodologies for preparing a single-word data set of typically developing children and a multiword data set of hearing and deaf/hard of hearing (DHH) children. Serving as references, developmental patterns can be constituted by a large, well-annotated normative data set. Studies of this kind that explicitly compare the phonological development of different groups of children, such as the longitudinal study presented by Iyer and colleagues (Iyer et al., 2017), are informative for both clinical and educational purposes. By focusing on the intersection of assessment purposes and quantitative indicators, the development of consonant acquisition could be a computable and applicable predictor of speech development delay (Shriberg et al., 1997). While speech data types (single-word vs. multiword) may feature varying speech characteristics, assessing single-word data is a much easier task (collection and processing) than multiword data. It was also one of the main purposes of this study to validate whether children in the normative data set (single-word) and those in the hearing data set (multiword) showed similar patterns in terms of consonant accuracy rates and acceptability/intelligibility ratings.

This study's main research purposes are listed below:

- Describing developmental patterns of consonant acquisition in typically developing Chinese-speaking children by consonant accuracy rates and acceptability ratings.
- Comparing hearing and DHH children's consonant production data to the normative patterns and specifying their characteristics in a multiword data set.
- Validating the applicability of consonant accuracy rate, later specified as percentage of consonants correct (PCC), as a quantitative indicator for assessing single-word and multiword speech data.
- Analyzing consonant error patterns in hearing and DHH children by referring to the normative patterns.

Consonant Production in Children

Transcription-Based Research on Consonant Acquisition

Clinical and phonological considerations are central to speech acquisition research. The accuracy rates of transcribed phonemes serve as empirical indices that underpin our understanding of phonological development patterns. In particular, consonant acquisition highlights patterns of phonological processes that help identify possible developmental delays (Bauman-Waengler, 2019; Bernthal et al., 2016; Shriberg et al., 2020). As phonemes, including consonants, glides, and vowels, are abstract units of speech that create minimal contrasts relative to the sounds within a given language's inventory, transcribing phonemes accurately, both narrowly and broadly, can be challenging. This task is labor intensive for single-word speech data and more so for continuous speech, for example, conversational discourse.

Sander (Sander, 1972) outlines the process of consonant acquisition in English by identifying the age at which speech sounds emerge in all word positions (initial, medial, final), recommending an average threshold of over 50% correct production for emergence. Additionally, a 75% threshold has also been noted in the literature on speech acquisition (Poole, 1934; Templin, 1957; Wellman et al., 1931). For determining mastery or age of acquisition, a commonly accepted threshold is 90% correct production (Hua & Dodd, 2000; G. B.-G. Lin & Lin, 1994; So & Dodd, 1995). For Chinese-speaking children, phonological processes involving syllable-initial consonants include de-/retroflexion, de-/aspiration, fronting/backing, velarization, lateralization, labialization, stopping, affrication, frication, and gliding, among others (Cho, 2008; Dodd et al., 2003; Jeng, 2011; Li, 2018). However, a comprehensive quantitative report on consonant acquisition based on well-annotated phoneme transcription is still lacking.

Intelligibility and PCC

When evaluating speech abilities, intelligibility as perceived by listeners serves as a crucial indicator of evaluating children's speech performance. This concept focuses not only on the clarity of pronunciation but also on the alignment between pronunciation and word meaning.

Decoding word meanings is regarded as one of the essential features of impressionistic intelligibility. For speech assessment, a number of terminologies are often used without general consensus, for example, acceptability, intelligibility, and comprehensibility, and so forth. For an in-depth questionnaire study that surveys the consensus on "intelligibility" and "comprehensibility," please refer to Pommée et al. (2022). Moreover, no matter the objective (acoustic properties) or subjective (perceptual judgment) measures of speech assessment, the speech materials (e.g., single words or continuous speech), the grading criteria (e.g., accuracy concerning pronunciation or acceptability concerning comprehension), and the purposes of evaluation (e.g., screening for speech sound disorders or longitudinal records of intervention progress) may all lead to varying outcomes. Therefore, speech intelligibility growth curves in typically developing children may differ across data sets. To take Hustad et al.'s (2021) report as an example, children aged 49–87 months reached the 75% intelligibility threshold for single-word stimuli. For multiword production, 75% intelligibility was observed in children aged 46–61 months. Although Hustad et al. and this study both employed multiword data, different utterance lengths across languages may yield diverse results.

To evaluate phonological development delays, the PCC was proposed as a quantitative measure based on evidence supporting its correlation with speech intelligibility ratings (Shriberg & Kwiatkowski, 1982). Furthermore, the Intelligibility in Context Scale (ICS; McLeod et al., 2012) rated by parents was also found to be positively correlated with PCC in a sample of 183 children whose parents expressed concerns about their language and speech performance (McLeod et al., 2015). Although PCC appears to correlate with intelligibility ratings, to our knowledge, no large-scale study of PCC and intelligibility-related ratings has been reported for Chinese-speaking children. The current study aimed to validate the correlation. If the correlation is verified, automatic phoneme recognition models that deliver instant PCC are empirically justified for the purpose of early-stage speech performance assessment.

Mandarin Chinese

Given a language, the results of consonant production studies may differ in terms of the size, setting, transcription standards of the data, and the dialectal varieties of the subjects. Various phonological development data for Chinese-speaking children (Beijing and Taiwan Mandarin) have been reported (Hua & Dodd, 2000; Li, 2018; Li & Thompson, 1977; G. B.-G. Lin & Lin, 1994) based on slightly different sound systems. Due to the limitations of resources and technology in previous studies, large speech data sets with signal-aligned annotations were not explicitly made available. In the current study, an automatic phone aligner, the ILAS Phone Aligner[1] (Liu et al., 2016; Tseng, 2019), was used to assist phone boundary labeling. The phoneme inventory implemented in the ILAS Phone Aligner resembles most Chinese phoneme inventories (Duanmu, 2007; Y.-H. Lin, 2007). The onset consonants include six plosives /p, pʰ, t, tʰ, k, kʰ/; six fricatives /f, s, ʂ, ɕ, x, ʐ/; six affricates /ts, tsʰ, tʂ, tʂʰ, tɕ, tɕʰ/; two nasals /m, n/; and one lateral /l/, or it can be vacant. Only /n/ or /ŋ/ is allowed in the coda position. There are two glides /j, w/ and 15 vowels including mono- and diphthongs /i, ɨ, ɯ, u, y, a, o, ə, e, ɚ, ai, ei, ao, ou, ye/. Mandarin Chinese has four lexical tones, that is, high-level, rising, contour, and falling tones and an unstressed, neutral tone. Tones are reported to be acquired at an earlier age than vowels and consonants. As mentioned, dialectal and regional differences may impact the outcomes of phonological development research. To take the two varieties of Taiwan and Beijing Mandarin as an example, G. B-G. Lin and Lin (1994) reported that /p, pʰ, m, n, l, k, kʰ, x, tɕ, tɕʰ/ are acquired before the age of 3 years; /ɕ, ts/, before age 4; /tʰ, tsʰ/, before age 4;6 (years;months); /ʂ/, before age 5 years; /f, tʂʰ, ʐ, s/, before age 5;6; and /tʂ/, after age 6 years. Hua and Dodd (2000) reported similar results in their Beijing Mandarin study, but with /f, ɕ/ acquired earlier and /tɕ, tɕʰ/ later than in the Taiwan Mandarin study presented in G. B.-G. Lin and Lin. The current study would supply additional empirical evidence for consonant acquisition by Taiwan Mandarin–speaking children.

Data and Method

Two data sets were used in this study. One is a normative data set consisting of single-word recordings from 798 typically developing preschool children. Another data set contains sentence-level multiword speech recordings from 79 (reportedly) normal-hearing (NH) children and 45 DHH children. Regarding dialectal and regional influences, these two recording projects were conducted primarily in metropolitan areas of northern Taiwan. It is anticipated that dialectal variation among this group of Taiwan Mandarin speakers will have a minimal effect on the current study.

Single-Word Normative Data Set: Typically Developing Children[2]

The normative data set was recorded in kindergartens in Taipei and New Taipei City. The kindergartens were officially contacted and asked if they agreed to

[1]https://aligner.ling.sinica.edu.tw/

[2]The recording project of the normative data set (2017–2020) was approved by the Institutional Review Board on Humanities and Social Science Research at Academia Sinica AS-IRB-HS07–107079.

participate the recording project. None of the subjects had known or diagnosed diseases related to language, hearing, or cognitive development. All subjects had verified normal hearing in both ears, as measured by passing a pure-tone audiometric screening at 1, 2, and 4 kHz at 20 dB. Appendix A lists the *Sinica Child Balanced Wordlist* used to elicit speech. The list includes 70 multisyllabic words with a balanced phoneme and tone design. All onset consonants appear at least once in the first and second positions. A total of 798 children named 70 pictures that matched the meaning of the words, totaling 55,860 recorded words. The speech data were initially processed by the ILAS Phone Aligner and then underwent three levels (word, syllable, and tone) of perceptual judgment of acceptability rounds conducted independently by two annotators who had previous experience recording young children's speech and were trained by the author (Tseng, 2024a, 2024b).

Single-Word Normative Data Set: Acceptability Rating and Transcription

In the normative data set, acceptability rating was based on perceptual judgments concerning whether words could be understood. At the word level, a tertiary annotation scheme was adopted: correct, acceptable, and incorrect. A word was annotated as "correct" if it could be understood without difficulty and showed no noticeable pronunciation deviations. "Acceptable" words were understandable but contained minor pronunciation deviations. Words whose meaning was not understandable in pronounced phonetic forms were annotated as "incorrect" words. Please note that the word-level annotation aimed to reflect mature adults' perception of word meaning, so no standard or dialectal pronunciation was specified for reference. Syllable- and tone-level annotation schemes were binary: acceptable and unacceptable. The acceptability judgment of syllables and tones relied on the completeness of syllabic segments and tone contours. The weighted Cohen's kappa and the agreement rate for word acceptability annotation are 0.586 and 80.9%, respectively. For syllables in the first, second, and third positions, the Cohen's kappa are 0.597, 0.635, and 0.623, respectively, and the agreement rates are 89.9%, 92.0%, and 92.1%, respectively. For tones associated with syllables in the three positions, the Cohen's kappa and agreement rates are 0.369, 0.464, 0.447 and 98.1%, 98.5%, 98.8%. The normative data set is highly skewed, as the majority of the consonant production is correct. Nevertheless, the two annotators showed moderate to substantial agreement with each other (Landis & Koch, 1977).

A hybrid procedure of automatic and manual processing was designed for phoneme transcription of the normative data set. For automatic transcription, words annotated as "correct" at the word level and "acceptable" at both the syllable and tone levels were converted into canonical phoneme sequences. The remaining 18,694 words were manually transcribed by a phonetician who had long-years' experience labeling adult and child speech, aided by Praat visualizations of waveforms and spectrograms (Boersma & Weenink, 2022; Tseng, 2024a). In order to verify the validity of automatic transcription, one word was randomly selected for each child from the words that were automatically transcribed. Subsequently, the 798 words were verified by two additional phoneticians who had more than 2 years' experience with children's speech. As a result, 34 words (4.26%) were noted as not completely correct. However, none of them would lead to an incorrect word meaning association. The author listened to these words and found that only eight words were phonetically reduced or with an omitted nasal coda.

Table 1 summarizes the transcription results of syllable-initial consonants. Phonemes in Mandarin Chinese are listed in rows, where columns list the transcribed phonemes and their percentages in the corresponding cells. Insertion and distortion were combined in the "ID" column, where distortion referred to consonant productions that cannot be accurately transcribed using the phoneme inventory mentioned above. The manner of articulation was grouped by blocks, indicating plosives, fricatives, affricates, and nasals. Due to rounding, the sums may not always equal one for each consonant in the row. This arrangement format provides a clear overview of consonant transcription summaries for typically developing children. Similar summaries may be made for tracking phonological development of a child or used for different subgroups by age, gender, or other specifications.

As each Mandarin Chinese syllable consists of at most four segments: an onset, a glide, a vowel, and a nasal coda, the accuracy rates of these segments were summarized by onset, glide, vowel, and coda, respectively. In addition, we had a separate rhyme category. For a syllable, the glide, vowel, and coda must all be correct to be counted as a correctly produced rhyme. Acceptability ratings were reported at levels of syllable, tone, and word, where both correct and acceptable words were included in the acceptability rating of words. Figure 1 shows the growth curves of accuracy rates and acceptability ratings. The accuracy rates of onset consonants/rhymes and the acceptability ratings of words/syllables appeared to better reflect the phonological development trends than the glide accuracy rates and the tone acceptability ratings. Despite different metrics derived from transcription and perceptual judgment, Figure 1 shows a correlation among the metrics that will be discussed in depth later.

Single-Word Normative Data Set: PCC and AccWord

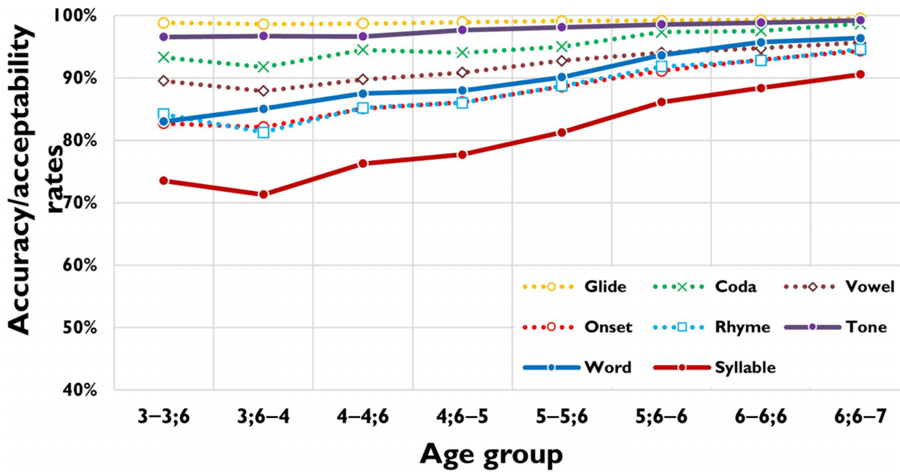
Based on the transcribed phonemes, PCC was calculated by the number of correct consonants over the total

Table 1. Transcription results of syllable-initial consonants in normative data set.

	Transcribed phonemes																						
Phonemes	_	p	pʰ	t	tʰ	k	kʰ	f	s	ʂ	ɕ	x	ʐ	ts	tsʰ	tʂ	tʂʰ	tɕ	tɕʰ	m	n	l	ID
_	**1.00**																						
p		**0.99**																					
pʰ	0.01	0.04	**0.94**																				
t	0.01			**0.97**		0.01																	
tʰ				0.02	**0.95**		0.01												0.01				
k	0.03			0.03		**0.94**																	
kʰ	0.01				0.04	0.02	**0.93**					0.01											
f	0.01	0.01	0.01					**0.94**				0.02											0.01
s				0.01	0.01				**0.89**	0.04				0.02	0.01								0.01
ʂ				0.01					0.35	**0.56**				0.04	0.02								0.01
ɕ									0.01		**0.92**							0.03	0.02				0.01
x	0.06					0.01	0.02	0.01				**0.91**											
ʐ	0.03			0.01									**0.52**									0.42	0.01
ts	0.01			0.02										**0.89**	0.01	0.04		0.01					0.01
tsʰ					0.03				0.01					0.02	**0.90**		0.02		0.01				
tʂ				0.02										0.40	0.01	**0.55**		0.01					0.01
tʂʰ					0.04				0.01					0.03	0.32		**0.57**		0.01				
tɕ																		**0.97**	0.01				
tɕʰ					0.01						0.01							0.03	**0.93**				
m	0.01																			**0.97**			
n	0.04																				**0.95**	0.01	
l	0.03			0.03									0.01								0.01	**0.92**	0.01

Note. The “ID” column stands for insertion and distortion. A segment is considered “distorted” when it cannot be transcribed by any phoneme in the phoneme inventory. Boldfaced numbers indicate correct phoneme production.

Figure 1. Accuracy rates of transcribed phonemes (dotted lines) and acceptability ratings based on judgment (solid lines) in normative data set.



number of produced consonants for each child. Acceptability rating was represented by the proportion of the number of words annotated as "correct" and "acceptable" out of the 70 words in the wordlist for each child, denoted AccWord. Table 2 reports PCC and AccWord summary statistics including their means and standard deviations; Figures 2A and 2B illustrate their distribution in box plots. A clear, rising trend was observed for both PCC and AccWord in the normative data set. Figures 2A and 2B show that there are more outliers using the acceptability metric (AccWord) than with the transcription-based metric (PCC).

Multiword Data Set: Hearing and DHH Children[3]

Subjects participating in the recording project of sentence repetitions included 79 NH children and 45 DHH children. Speech data were recorded at Jengo Kindergarten and the Children's Hearing Foundation in Taipei. We did not conduct any hearing tests but relied on the information provided by Jengo Kindergarten and the Children's Hearing Foundation. The DHH children's group comprised 15 children with cochlear implants (CIs) and 30 with hearing aids (HAs). They were all receiving auditory–verbal therapy training sessions at the time of recording (Tseng, 2019; Tseng & Liu, 2021). For detailed subject information about the CI and HA children, please refer to Tseng et al. (2011). Appendix B lists the 18 sentences used to record 11,575 syllables. The sentences were designed to cover all phonemes in Mandarin Chinese. Although it lacked a balanced design in terms of syllable position and tone, all consonants in Mandarin Chinese were recorded in the data set for each child. This allowed comparative analyses with the normative data set for our exploratory study.

Multiword Data Set: Transcription and Intelligibility Rating

Intelligibility in the multiword data set was rated by the pronunciation and fluency of the recorded sentences for each child by three annotators independently. The degree of intelligibility was annotated on a scale from 1 to 5, representing very low, low, neutral, high, to very high intelligibility levels (Tseng et al., 2011). Results of the weighted Cohen's kappa, pairwise among the three annotators, were 0.458, 0.381, and 0.529 for NH, 0.747, 0.580, and 0.775 for HA, and 0.779, 0.703, and 0.741 for CI. The agreement rates may not be particularly high. However, overall, 65% of the NH children's data had only one scale point discrepancy across the three annotators; for the CI and HA children's data, they were 60% and 43%. Only two HA children were annotated with a discrepancy more than one scale point, making up 7% of the data.

Phonemic transcription of all 11,575 syllables was completed by two phoneticians trained by the author to label children's speech data independently in 2008. To ensure transcription quality meets the same standard used for the normative data set, two phoneticians verified the data in 2023. The agreement rates between the original

[3]The DHH children's project was financially supported by the Children's Hearing Foundation. The hearing children project received financial support from the National Science Council of Taiwan. There were no IRB regulations at the time of recording; however, consent forms were collected for each child who participated in the recording project (2008–2012).

Table 2. Percentage of consonants correct (PCC) and AccWord summary statistics in normative data set.

	Female	PCC		AccWord		Male	PCC		AccWord		
Age (years/ years;months)	*n*	*M*	*SD*	*M*	*SD*	*n*	*M*	*SD*	*M*	*SD*	Total
3–3;6	21	0.846	0.060	0.838	0.102	10	0.787	0.097	0.814	0.129	31
3;6–4	52	0.837	0.074	0.862	0.102	40	0.801	0.116	0.836	0.166	92
4–4;6	58	0.853	0.069	0.880	0.101	64	0.849	0.092	0.871	0.124	122
4;6–5	64	0.872	0.090	0.887	0.131	55	0.849	0.100	0.872	0.127	119
5–5;6	53	0.889	0.076	0.911	0.105	65	0.882	0.085	0.894	0.136	118
5;6–6	62	0.922	0.054	0.945	0.067	64	0.899	0.072	0.928	0.078	126
6–6;6	60	0.926	0.068	0.951	0.074	79	0.931	0.053	0.955	0.067	139
6;6–7	29	0.947	0.038	0.960	0.037	22	0.936	0.054	0.969	0.023	51
Total	**399**					**399**					**798**

transcription and the two annotators' verifications were 86.2% and 89.6%. Among them, 80.3% of the transcription was agreed upon by all three annotators. The final transcription version was achieved through discussions led by the author. Tables 3, 4, and 5 summarize the transcription results of syllable-initial consonants in the CI, HA, and NH data sets.

Multiword Data Set: PCC and IntScore

For serving as a quantitatively comparable score for AccWord in the normative data set, the proportion of the total scores given to each child out of the maximum value of 15 was calculated, denoted IntScore. Table 6 presents PCC and IntScore summary statistics, broken down by subgroup and gender. Figures 3A and 3B present the box plot distribution of PCC and IntScore.

Phonological Development in Normative Data Set

This section presents developmental patterns of consonant acquisition and their phonological characteristics, followed by reports of frequent consonant errors in the normative data set.

Mandarin Chinese Phoneme Acquisition Chart

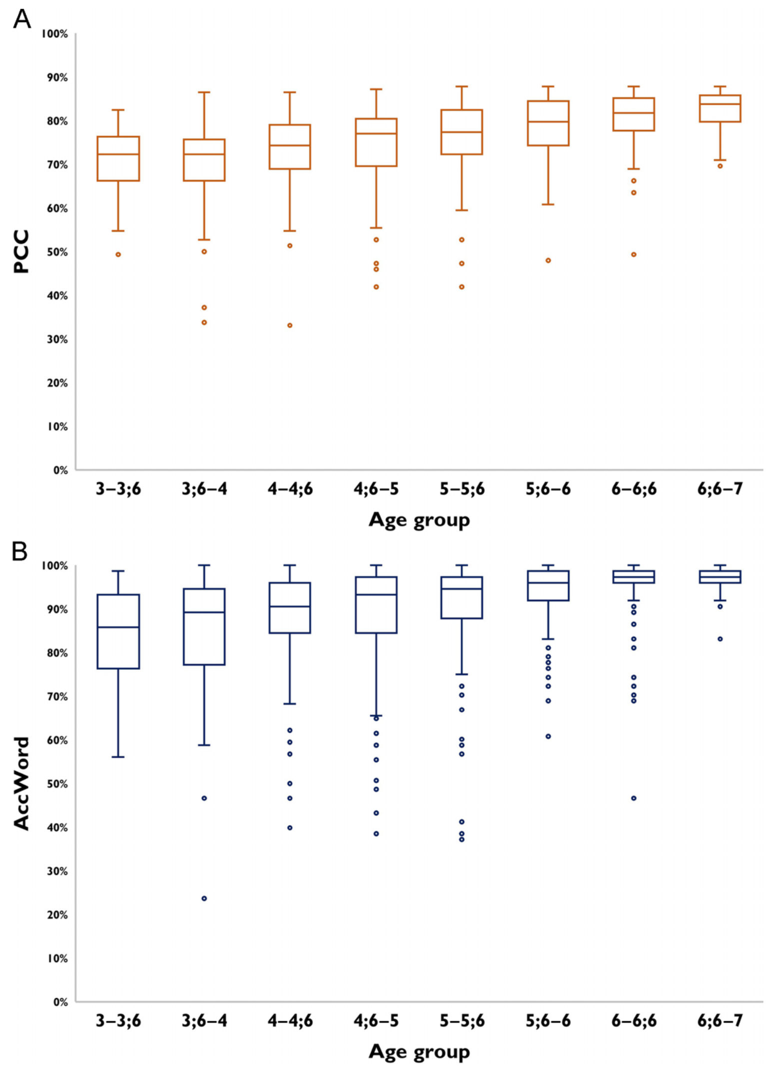
Based on the phoneme transcription of the 798 children's data in the normative data set, Figure 4 illustrates the Mandarin Chinese phoneme acquisition chart. The yellow span marks the ages at which 75% of the consonant production are correct. A black bar indicates the acquisition age upon reaching 90%. Moreover, /o/ is the only speech sound that does not reach the 90% threshold until seven. Please note that although the *Sinica Child Balanced Wordlist* has a balanced design in terms of onset, rhyme, and tone, our single-word normative data set does not comprehensively account for contextual and linguistic factors related to position, syllable structure, and inter-/intrasyllabic phonological environments, among others. Nevertheless, the normative data set presented in this study is, to our knowledge, the largest high-quality annotated data set of Mandarin Chinese–speaking children.

Furthermore, word position was reported to influence phoneme production performance (Sander, 1972). As the *Sinica Child Balanced Wordlist* ensures that all onset consonants appear in both the first and second word positions, it is eligible to explore the effect of positions. Figure 5 reports consonant acquisition data for the first and second positions separately. Compared to Figure 4, in which all consonant positions are considered, Figure 5 shows diverse results, suggesting that stimulus selection by word position affects the outcomes.

Manner and Place of Articulation

To study the role of phonological features in consonants, each consonant was converted into a three-way phonological feature representation: manner of articulation, place of articulation, and aspiration. Concerning the manner of articulation, as shown in Table 7, plosives and onset nasal sounds were acquired earlier than fricatives, affricates, and nasal coda. Regarding the place of articulation, bilabial and alveolar sounds appeared to be mastered before velar ones. Retroflex sounds were the most difficult to produce. Unaspirated sounds were either acquired at the same age as their aspirated counterparts, that is, /k-kʰ/, /ts-tsʰ/, and /tʂ-tʂʰ/ or before, that is, /p-pʰ/, /t-tʰ/, and /tɕ-tɕʰ/. These results may be considered evidence from speech acquisition research, indicating that aspiration may be phonologically a marked feature in Mandarin Chinese.

Figure 2. (A, B) Box plots of percentage of consonants correct (PCC) and AccWord in normative data set, with the line in the box representing the median, the top and bottom of the box representing the 25th and 75th percentiles, the whiskers representing the 10th and 90th percentiles, and the dots representing the outliers.



Frequent Errors

To examine consonant error types in typically developing children, Table 8 summarizes the results for non-empty onset consonants in the normative data set. Please note that due to our three-way feature representation, two error tokens of ʂ- > ʐ were produced with all three features correct, accounting for 0.02% of the errors. These can be considered substitution errors caused by voicing. Aside from this anomaly, errors related to the place of articulation comprised about 85% of the total errors. Among them, 60% involved only place-of-articulation errors, while 18% also involved aspiration. Approximately 7% of consonant errors were made only with articulation manner; and 4.6%, only with aspiration. Overall, errors related to the manner of articulation made up 16% of the total errors. There were only 3.44% of errors for which all three phonological features were incorrect.

Figure 6 illustrates the developmental progress of each age group in the normative data set. Despite slightly different proportions within each group, similar trends as shown in Table 8 are observed across all eight age groups in Figure 6.

Assessing Multiword Data Set of Hearing and DHH Children

Referring to the normative patterns, consonants in the multiword data set were further examined using PCC,

Table 3. Transcription results of syllable-initial consonants in cochlear implant (CI) data set.

	Transcribed phonemes																					
Phonemes	_	p	p^h	t	t^h	k	k^h	f	s	ʂ	ɕ	x	ʐ	ts	ts^h	tʂ	tʂ^h	tɕ	tɕ^h	m	n	l
_	**0.96**	0.01		0.01		0.01						0.01								0.01	0.01	
p	0.05	**0.90**	0.03	0.03																		
p^h			**0.82**		0.18																	
t	0.02			**0.88**	0.01	0.06							0.01								0.01	0.01
t^h	0.01	0.01			**0.93**		0.03					0.01										
k		0.02	0.01	0.12		**0.85**																
k^h	0.02			0.01	0.01		**0.93**					0.02										
f	0.08							**0.92**														
s				0.07					**0.66**	0.12	0.04	0.01		0.04	0.03						0.01	
ʂ	0.04								0.16	**0.67**	0.05	0.04	0.02				0.04					
ɕ	0.02			0.02							**0.86**	0.01						0.03	0.06			
x	0.06					0.02						**0.91**									0.01	
ʐ	0.08												**0.62**								0.08	0.23
ts	0.15								0.04					**0.81**								
ts^h												0.07			**0.86**		0.07					
tʂ	0.14										0.07			0.14		**0.57**			0.07			
tʂ^h	0.07													0.04	0.22		**0.59**		0.04		0.04	
tɕ	0.11			0.04							0.04							**0.74**	0.07			
tɕ^h	0.04				0.04						0.04	0.04							**0.85**			
m																				**0.98**	0.03	
n				0.07																0.04	**0.89**	
l	0.12				0.02																0.02	**0.84**

Note. Boldfaced numbers indicate correct phoneme production.

Table 4. Transcription results of syllable-initial consonants in hearing aid (HA) data set.

	Transcribed phonemes																					
Phonemes	_	p	pʰ	t	tʰ	k	kʰ	f	s	ʂ	ɕ	x	ʐ	ts	tsʰ	tʂ	tʂʰ	tɕ	tɕʰ	m	n	l
_	**0.98**					0.01															0.01	0.01
p	0.02	**0.89**	0.01	0.04		0.01						0.01						0.01		0.01	0.01	
pʰ			**1.00**																			
t	0.02	0.01		**0.92**	0.01	0.03							0.01					0.01			0.01	
tʰ	0.03		0.01	0.01	**0.94**							0.01							0.01			
k		0.01		0.07		**0.90**	0.01					0.01		0.01				0.01				
kʰ					0.02	0.02	**0.89**					0.05			0.02							
f								**1.00**														
s	0.01			0.07		0.03			**0.49**	0.05	0.01		0.03	0.17	0.04	0.06	0.01	0.02				
ʂ	0.06							0.02	0.13	**0.49**	0.02	0.05		0.03	0.01	0.11	0.06	0.01	0.02			
ɕ	0.04			0.02							**0.69**	0.02						0.15	0.07			
x	0.05					0.01	0.01	0.01				**0.91**										
ʐ	0.10												**0.28**								0.03	0.59
ts	0.02					0.02			0.04				0.04	**0.54**	0.05	0.11	0.02	0.16			0.02	
tsʰ			0.04		0.04		0.07					0.11			**0.71**		0.04					
tʂ					0.04	0.08								0.17		**0.58**	0.04	0.08				
tʂʰ	0.05				0.04	0.02	0.04					0.02		0.02	0.21	0.04	**0.46**	0.04	0.05		0.02	
tɕ	0.04										0.02			0.02				**0.90**	0.02			
tɕʰ	0.02				0.04						0.02	0.08							**0.79**		0.04	0.02
m	0.01												0.01							**0.93**	0.05	
n	0.02																			0.04	**0.92**	0.02
l	0.04		0.01	0.03	0.02												0.01	0.01	0.01	0.01	0.05	**0.82**

Note. Boldfaced numbers indicate correct phoneme production.

Table 5. Transcription results of syllable-initial consonants in normal hearing (NH) data set.

	Transcribed phonemes																					
Phonemes	_	p	pʰ	t	tʰ	k	kʰ	f	s	ʂ	ɕ	x	ʐ	ts	tsʰ	tʂ	tʂʰ	tɕ	tɕʰ	m	n	l
_	**0.99**											0.01										
p		**0.96**	0.03																			
pʰ		0.03	**0.96**																	0.01		
t	0.02			**0.95**		0.01												0.01				0.01
tʰ	0.03			0.01	**0.90**		0.01					0.03							0.01			
k				0.04		**0.94**	0.01															
kʰ					0.02	0.02	**0.96**															
f		0.01						**0.95**				0.04										
s				0.01		0.01			**0.87**	0.04				0.04	0.01		0.01					
ʂ	0.01					0.01		0.01	0.58	**0.35**			0.01		0.01	0.01	0.01					
ɕ	0.01			0.01							**0.92**	0.01						0.03	0.03			
x	0.08						0.01					**0.91**										
ʐ	0.05			0.01									**0.41**									0.53
ts	0.01					0.01			0.01				0.01	**0.70**	0.03	0.07		0.15	0.01			
tsʰ				0.01	0.03		0.01					0.01		0.04	**0.86**		0.04					
tʂ	0.03													0.54	0.01	**0.37**		0.05				
tʂʰ						0.01	0.01		0.01				0.01	0.03	0.70	0.01	**0.22**		0.01			
tɕ	0.03			0.01	0.01									0.01		0.01		**0.93**	0.01			
tɕʰ	0.02		0.01		0.03		0.01				0.03	0.01						0.01	**0.88**			
m	0.01																			**0.98**		
n	0.03																			0.01	**0.97**	
l	0.03			0.01								0.01	0.01								0.01	**0.92**

Note. Boldfaced numbers indicate correct phoneme production.

Table 6. Percentage of consonants correct (PCC) and IntScore summary statistics in cochlear implant (CI), hearing aid (HA), and normal hearing (NH) data sets.

	Female	PCC		IntScore		Male	PCC		IntScore		
	N (age)	*M*	*SD*	*M*	*SD*	*N* (age)	*M*	*SD*	*M*	*SD*	Total
Deaf/hard of hearing children (CI/HA)	**21**					**24**					**45**
CI											
Severe						1 (5;1)	0.768		0.400		1
Severe–profound						1 (5;7)	0.708		0.733		1
Profound	5 (3;7–11;2)	0.711	0.082	0.760	0.198	8 (4;0–12;5)	0.667	0.131	0.525	0.299	13
HA											
Mild	1 (3;9)	0.564		0.600		2 (6;0–7;3)	0.774	0.023	0.833	0.047	3
Moderate	4 (3;3–7;11)	0.767	0.058	0.783	0.263	2 (5;3–9;5)	0.674	0.047	0.600	0.283	6
Moderate–severe	4 (4;11–8;0)	0.612	0.180	0.667	0.144	4 (4;0–6;4)	0.714	0.095	0.850	0.114	8
Severe	1 (10;5)	0.750		0.733		3 (4;11–7;1)	0.658	0.215	0.644	0.315	4
Severe–profound	3 (4;2–6;0)	0.554	0.041	0.533	0.116	2 (6;11–7;1)	0.678	0.056	0.667	0.189	5
Profound	3 (4;3–6;1)	0.497	0.033	0.400	0.067	1 (5;4)	0.426		0.267		4
NH children	**37** (2;11–6;3)	0.728	0.093	0.827	0.140	**42** (3;0–6;3)	0.734	0.067	0.832	0.123	**79**

Figure 3. (A, B) Box plots of percentage of consonants correct (PCC) and IntScore in multiword data set. CI = cochlear implant; HA = hearing aid; NH = normal hearing.

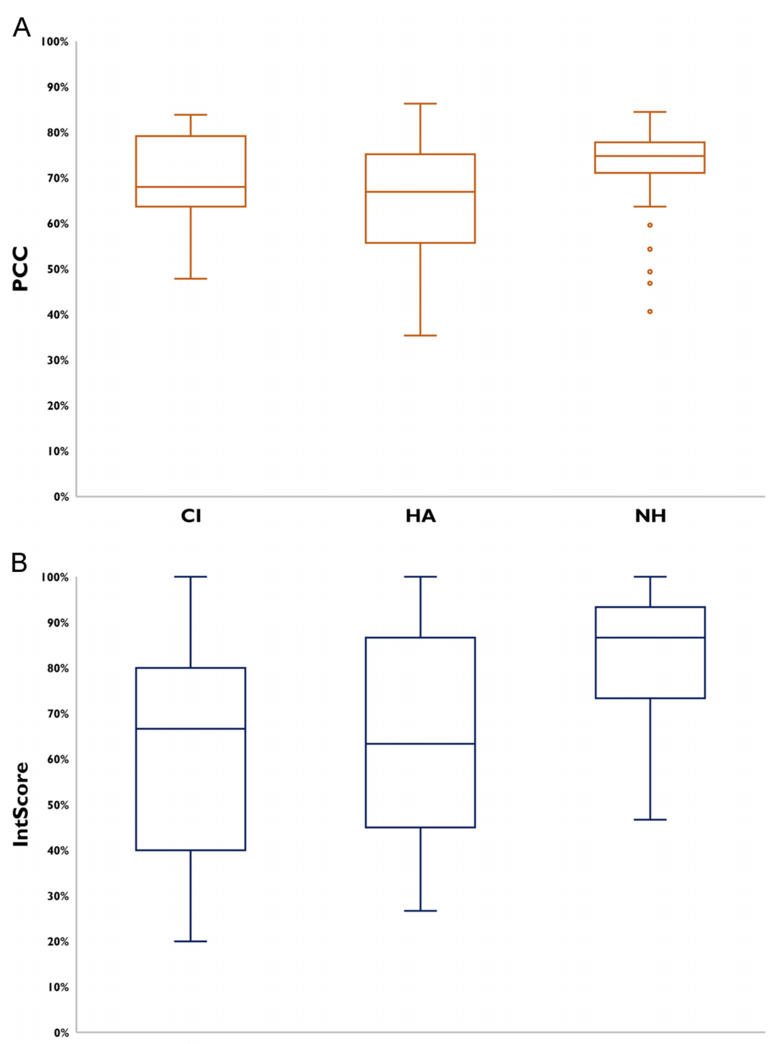
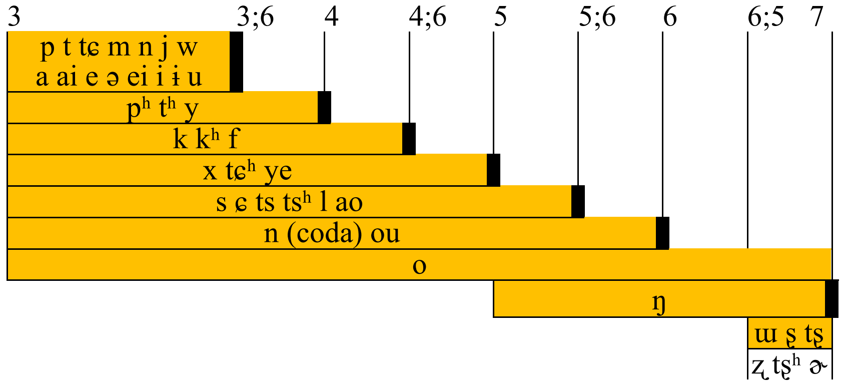


Figure 4. Mandarin Chinese phoneme acquisition chart.



intelligibility rating, and consonant error types. The multiword data set was not originally designed for direct comparison. However, it is by far the most comprehensive resource with transcription and intelligibility ratings available to us. Age and hearing experience, defined as biological age and the duration of experience with hearing devices, were also explored to see how they interacted with PCC and intelligibility ratings.

PCC Is Correlated With Acceptability/Intelligibility Ratings

Shriberg and Kwiatkowski (1982) proposed a classification of phonological development delays based on consonant production and speech intelligibility. Going on a similar research line, this study investigated the correlation between PCC and perceptual judgments of AccWord and IntScore. To illustrate their patterns across all subject groups, Figures 2 and 3 were collapsed into Figure 7 that visualizes a group-wise box plot of PCC and IntScore/AccWord for children in CI, HA, NH, and the normative data sets. Please note that only syllable-initial consonants were considered, and syllables with zero onset were excluded from the calculations. PCC and AccWord were positively correlated for all age groups except for the age group 6;6 to 7 years. Statistically significant, the Pearson correlation coefficients (r) are .64, .8, .7, .81, .79, .7, and .63, $p < .01$, respectively. For the multiword data set, the correlation coefficients between PCC and IntScore are .85, .8, and .75, $p < .01$, for CI, HA, and NH. Moreover, Figure 8 presents a subject-wise scatter plot of PCC and IntScore/AccWord for children in CI, HA, NH, and the normative data sets. The green regression line marks the correlation trend among children in the normative data set. Despite statistically significant overall correlations between PCC and IntScore, hearing and DHH children showed varying correlation patterns. Only CI and HA children's data were positioned in the lower right area of Figure 8. That is, intelligibility was scored high, but with low consonant accuracy rates. This difference may be further elaborated by examining vowel and tone production. Or for future studies, if child-level variables (e.g., degree of hearing loss) can

Figure 5. Onset consonant acquisition in first/second positions.

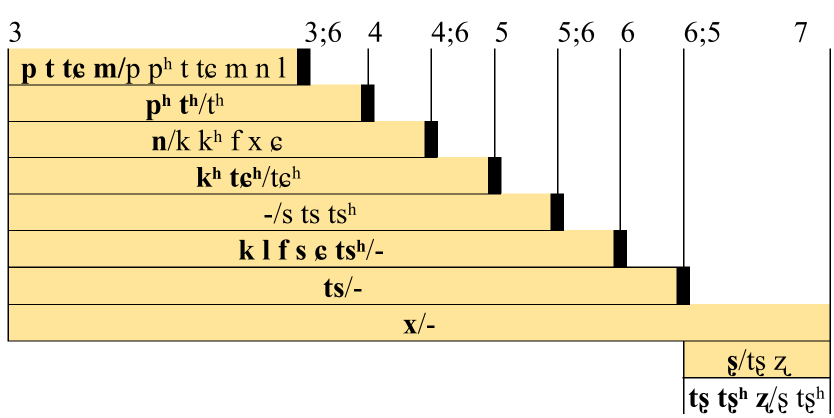


Table 7. Phonological development by manner and place of articulation.

Manner of articulation		3–3;6	3;6–4	4–4;6	4;6–5	5–5;6	5;6–6	6–6;6	6;6–7	> 7
	Plosive	p t	p^h t^h	k k^h						
	Nasal	m n					n (coda)		ŋ	
	Lateral					l				
	Fricative			f	x	s ɕ				ʂ ʐ
	Affricate	tɕ			tɕ^h	ts ts^h				tʂ tʂ^h
Place of articulation		3–3;6	3;6–4	4–4;6	4;6–5	5–5;6	5;6–6	6–6;6	6;6–7	> 7
	Bilabial	m p	p^h							
	Labiodental			f						
	Dental-alveolar	n t	t^h			l	n (coda)			
	Alveolar					s ts ts^h				
	Alveolar-palatal	tɕ			tɕ^h	ɕ				
	Velar			k k^h	x				ŋ	
	Retroflex									ʂ ʐ tʂ tʂ^h

be more strictly controlled, correlation patterns across subject groups or within-subject developmental changes may be comprehensively studied.

The Role of Age

This section examined how children's consonant production varied with age and hearing status in hearing and DHH children. Please note that the age range and hearing loss degree were imbalanced and diverse in the two data sets. Due to limited subject numbers, only children aged 3–7 years were selected. Table 9 presents the numbers of participants in each age group in CI, HA, NH, and the normative data sets. As shown in Figures 9A and 9B, both PCC and intelligibility rating for NH children were lower than their age-matched counterparts in the normative data set. Although we did not conduct hearing screenings ourselves, the NH children were reported without hearing or cognitive impairments. In an intelligibility development study, Hustad et al. (2021) stated that the 90% threshold for single-word intelligibility was achieved later than for multiword intelligibility in typically developing children. In contrast, the single-word normative group showed higher PCC and AccWord than the PCC and IntScore in the multiword NH group in the current study. It was also observed that the difference in PCC was wider than that in IntScore/AccWord. This may stem from continuous speech being phonetically more reduced than words spoken in isolation. These results may not be considered affirmative conclusions yet, but requiring future studies with controlled settings of subjects such as age and gender.

Table 8. Distribution of syllable-initial consonant errors by phonological features; + correct, – incorrect.

Manner	Place	Aspiration	%
+	+	+	0.02%
+	+	–	4.60%
+	–	+	61.20%
+	–	–	17.81%
–	+	+	7.25%
–	+	–	2.90%
–	–	+	2.78%
–	–	–	3.44%

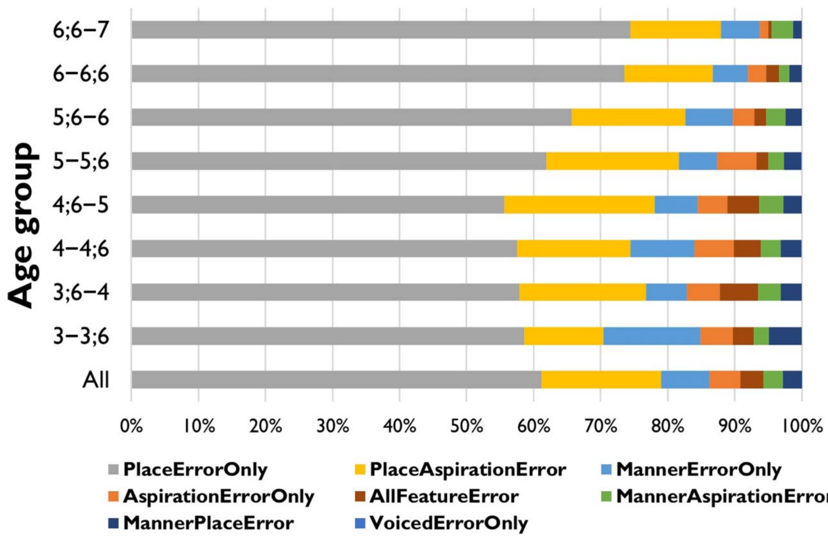
Figures 9A and 9B illustrate developmental results in terms of PCC and IntScore/AccWord. The trend of growth by age in PCC and IntScore/AccWord was observed in the normative and NH data sets. However, it was not explicitly observed in the CI and HA data sets. When considering the duration of experience with hearing device instead of biological age, a more consistent pattern was revealed, as shown in Figures 10A and 10B. Despite the limited size of data, our results showed that PCC and similar intelligibility measures may be useful for representing longitudinal progress in hearing rehabilitation of DHH children.

We have noted a peculiar case involving a child with a CI. This child, who was aged 6;9 at the time of recording, had her profound hearing loss diagnosed at 4 months. She received her CI at 6 months. She achieved the highest intelligibility score, and her PCC even surpassed the NH average. This supports the notion that early detection of hearing loss and intervention benefits speech development in DHH children (May-Mederake, 2012).

Visualizing PCC Summaries by Referring to Normative Patterns

A statistically significant correlation between PCC and IntScore/AccWord was identified earlier in this study. This suggests that PCC may serve as a quantitative assessment measure that reflects perceived intelligibility. In this section, a visualization format that concisely summarizes the transcription results of consonants was proposed for

Figure 6. Developmental patterns of consonant error types.



norm-referenced comparisons and developmental changes. Figures 11A and 11B present consonant acquisition in female and male subjects from the normative data set. Each layer marks the percentage of correctly produced consonants in each age group. The increasing trend from light to dark colors indicates consonant acquisition progress. Figure 11C shows the percentages of correct consonant production in the CI, HA, and NH groups. For instance, the low accuracy rates of retroflex sounds across all subject groups can be clearly visualized. Overall, the NH pattern resembled the normative data set, while the CI and HA subgroups showed distinct differences from the normative patterns. This diagram is currently accommodated with automatic phoneme recognition models, aiming to provide visualized assessment results for tracking longitudinal progress (Y.-S. Lin et al., 2024).

Comparative Error Patterns

In the normative data set, errors involved one or more features in articulation manner, articulation place, aspiration, and voicing. Among the errors associated with

Figure 7. Group-wise box plot of percentage of consonants correct (PCC) and IntScore/AccWord. CI = cochlear implant; HA = hearing aid; NH = normal hearing.

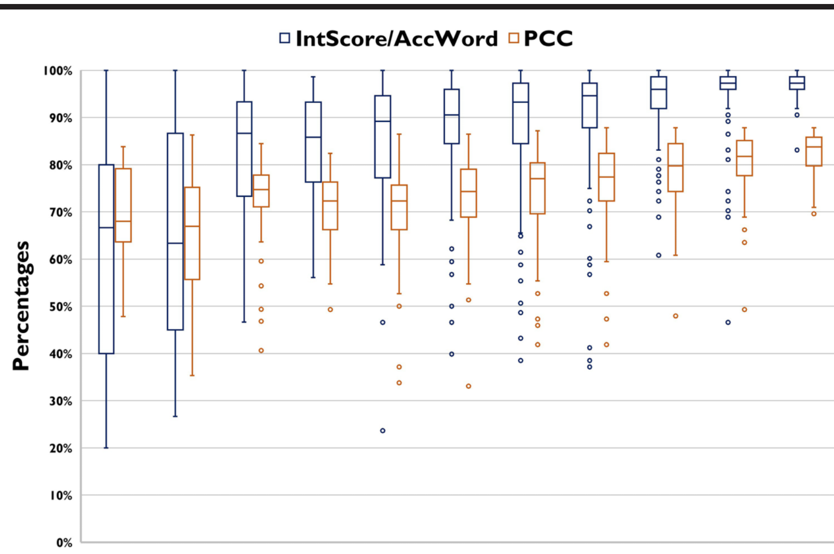
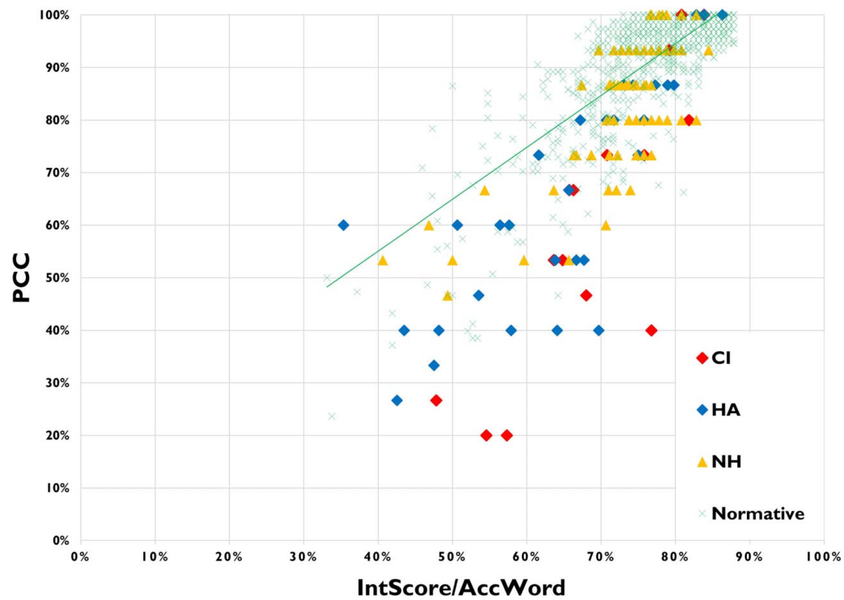


Figure 8. Subject-wise scatter plot of percentage of consonants correct (PCC) and IntScore/AccWord. CI = cochlear implant; HA = hearing aid; NH = normal hearing.



articulation manner, stopping errors comprised 5.31%, while frication and affrication errors were 3.26% and 1.07%, respectively. De-aspiration errors for the aspirated-unaspirated pairs /p, p^h, t, t^h, k, k^h, ts, ts^h, tʂ, tʂ^h, tɕ, tɕ^h/ accounted for 1.76%, compared to 0.31% for aspiration errors. Errors involving retroflex sounds occurred particularly frequently, with rates of 56.36%, 52.26%, 54.51%, and 57.34% for /ʂ, ʐ tʂ, tʂ^h/, respectively. Lastly, /n-l/ and /ʐ-l/ errors were observed in the normative data set, mirroring patterns often found in adults' daily speech in Taiwan Mandarin.

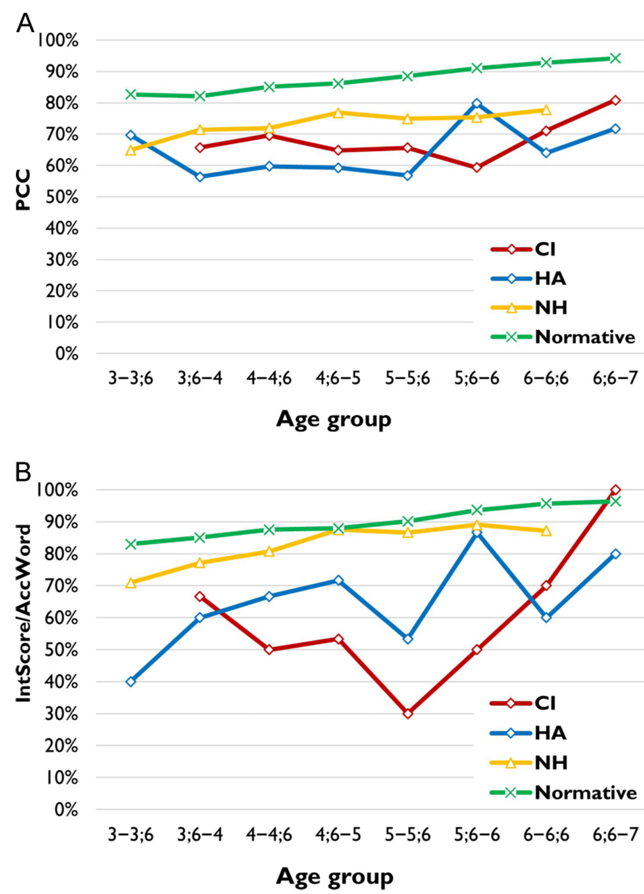
The findings regarding error patterns in the normative data set were summarized and examined whether they were also observed in the multiword data set. Three tendencies were observed in CI, HA, and NH data sets: Stopping errors occurred more frequently than frication, the accuracy rate for retroflex consonants was low, and there was a mixed use of /n, l, ʐ/. The confusion between /n, l, ʐ/ may result from the colloquial usage of speaker ethnic groups in Taiwan with different dialect backgrounds. In contrast to the single-word normative data set, affrication errors were produced more often than stopping and frication errors in the multiword data set. As shown in Table 10, de-aspiration errors were more common than aspiration errors in the DHH subgroups (CI and HA). This tendency differs from the hearing subgroup (NH). Another peculiar phenomenon is that errors involving the

Table 9. Number of subjects whose age and hearing experience are under 7 years old.

	Age group								
	3–3;6	**3;6–4**	**4–4;6**	**4;6–5**	**5–5;6**	**5;6–6**	**6–6;6**	**6;6–7**	**Total**
By age									
CI		1	2	1	2	2	2	1	11
HA	1	1	4	4	5	1	5	1	22
NH	14	7	9	8	14	14	12		78
Normative	31	92	122	119	118	126	139	51	798
By hearing experience									
CI	3	2	3	1			1		10
HA	4	2	3	4	1	1		1	16

Note. CI = cochlear implant; HA = hearing aid; NH = normal hearing.

Figure 9. (A, B) Percentage of consonants correct (PCC) and IntScore/AccWord by age group. CI = cochlear implant; HA = hearing aid; NH = normal hearing.



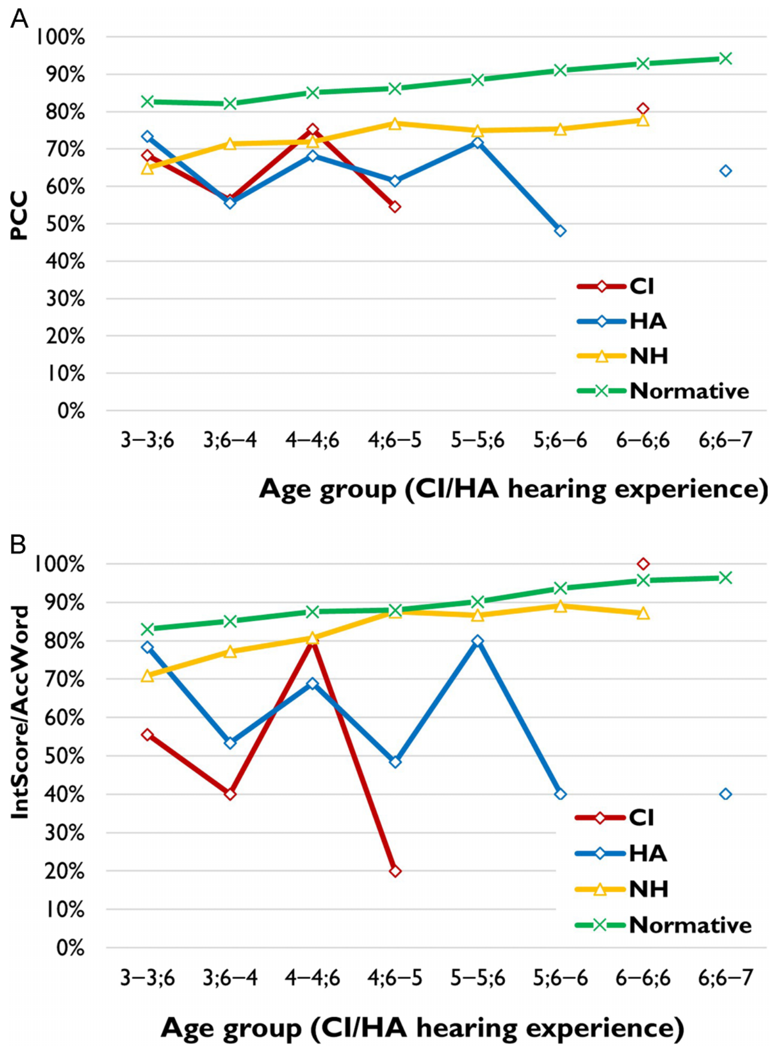
articulation place occurred less frequently in the HA group, for which we have no explanation at this stage.

General Discussion

Limitations and Methodological Considerations

The current study presented normative patterns of consonant production in typically developing Chinese-speaking children and explored the differences in the multiword data set between the patterns described in terms of PCC, acceptability/intelligibility rating and error types. Although it was demonstrated that normative patterns can be useful for illustrating phonological development in hearing and DHH children, the scarcity of balanced and comparable data is still one of the most severe problems in speech assessment research. It was likely that uncontrolled factors (e.g., age of hearing loss diagnosed, hearing loss degree, hearing device, intervention program, age, gender, family background, and so forth) resulted in a wide diversity in PCC and intelligibility level in our multiword data set. Nevertheless, the Taiwan Mandarin data presented in this study provided supplementary materials with comprehensive data annotation for the cross-linguistic review of consonant acquisition research (Li & To, 2017; McLeod & Crowe, 2018). On top of that, methodological considerations should be established to facilitate thorough comparisons of both consonant and vowel acquisition by different speaker groups across different languages (Edwards & Beckman, 2008; Hochmann et al., 2011). Shriberg and colleagues suggested that clinically relevant metrics should include the proportions of correctly produced consonants or vowels, as well as common and uncommon distortions, omissions, and substitutions (Shriberg et al., 1997). They proposed various modified versions of PCC and percentage of vowels correct (PVC), including adjusted and revised versions such as PCC-A, PCC-R, PVC-A, and PVC-R, along with an amalgamation version, PPC-R. To derive the indicators, high-quality data, especially clinically authentic data, are needed, for which close and intensive collaborations between linguists,

Figure 10. (A, B) Percentage of consonants correct (PCC) and IntScore/AccWord by hearing experience for CI/HA children. CI = cochlear implant; HA = hearing aid; NH = normal hearing.

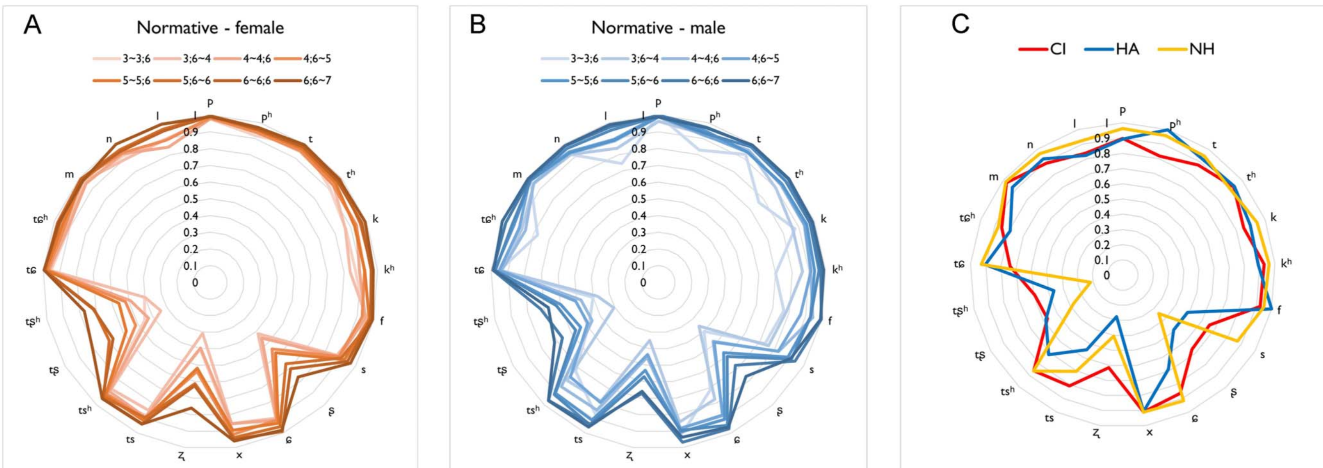


speech therapists, practitioners, or maybe engineers as well, are required.

Annotation tasks and granularity depend on speech assessment purposes. In this study, acceptability-based (AccWord) and pronunciation fluency–based (IntScore) judgments were conducted. Although intelligibility is a key component shared by these two rating methods, they are not equivalent to each other. In the normative data set, acceptability was rated based on whether the words were understandable at multiple linguistic levels. However, as they were single-word recordings, the conducted acceptability rating did not account for issues crucial to intelligibility judgments in continuous speech (e.g., fluency, intonation and comprehensibility). In the multiword data set, intelligibility was rated by overall pronunciation clarity and fluency, but without word-level evaluations. Mandarin Chinese is a tone language, so tone reduction in continuous speech is likely to make the association between word meaning and phonetic representation a complicated research issue.

Rating categorical degrees of intelligibility is a specific task that differs in the linguistic domain (word, sentence, discourse), the linguistic aspects (fluency, intonation, accuracy of pronunciation), and the type of perceptual judgment (whether or to what degree the word or sentence is understandable and acceptable). Constructing large-scale data sets of child speech with comparable intelligibility or comprehensibility ratings requires well-thought-out operational criteria. In addition to the linguistic and perceptual domains and aspects, the interactive and functional properties of human speech and communication should be considered too. The ICS is a parent-reported measure of children's intelligibility with different speaking partners (McLeod et al.,

Figure 11. (A, B, and C) Illustrative summaries of consonant production in normative and multiword data sets. CI = cochlear implant; HA = hearing aid; NH = normal hearing.



2015). Together with conventional expert-conducted judgments, the assessment by children's parents is valuable because it corporates the parent–child dynamic that serves a specific communicative purpose in speech production (Levelt, 1989; Tomasello, 2000). Instead of adopting existing quantitative measures, customized metrics that include suprasegmental properties that are crucial for efficient social communication should be designed for speech assessment, especially for children whose speech is unintelligible (Bauman-Waengler, 2019; McLeod & Baker, 2017; Pommée et al., 2022; Shriberg et al., 2020).

Phonemic transcription focuses on phonologically contrastive features rather than phonetic details of speech sounds. Shriberg et al. (1997) suggested concrete criteria for PCC design and use in clinical applications, including the following: (a) Data should be designed, collected, and analyzed; (b) speakers should be selected and assessed; and (c) assessment measures should be standardized and reliably obtained. The current study addressed some of these concerns. In the normative data set, age and gender were controlled, leaving out dialectal, regional, and socioeconomic factors. The results may therefore be biased. An operational definition of "normative data set" and practical guidelines regarding "normative patterns" are needed from a methodological perspective. When there is consensus on variables that should be controlled and features that should be characterized for a normative data set, patterns of phonological development in DHH children may be more explicitly presented and generally accepted in academic and clinical practice. Then we could go one step further and construct "normative data sets" for different subject groups of DHH children.

Applicability of Assistive Technology

Automatic assistive tools can reduce the workload of speech-language pathologists for transcribing speech data and obtaining a systematic report of speech production (Chang & Yeh, 2019). Normative data facilitate the development of model-assisted tools designed to detect and assess early-stage speech problems in young children, for example, intelligibility scoring, phoneme recognizer, and tone category prediction (Y.-S. Lin et al., 2024; Y.-S. Lin & Tseng, 2021; Tseng et al., 2023). Beyond serving as training data for computational model construction, the signal-aligned phoneme boundary information in the two data sets

Table 10. Norm-referenced, comparative analysis of syllable-initial consonant errors.

Summarized error patterns in normative data set	CI	HA	NH
1. Place of articulation errors > manner of articulation errors	✓	×	✓
2. Stopping > frication	✓	✓	✓
3. Stopping > affrication	×	×	×
4. De-aspiration > aspiration	×	×	✓
5. Low accuracy rate of retroflex sounds	✓	✓	✓
6. Mixed use of /n, l, ʐ/	✓	✓	✓

Note. CI = cochlear implant; HA = hearing aid; NH = normal hearing.

presented in this study is also useful for phoneme-level research on language-related abilities and acoustic-prosodic analysis in young children (Hazan & Barrett, 2000; Polyanskaya & Ordin, 2015; Redford & Howson, 2022; Treiman & Cassar, 1997; Yang et al., 2015). Referring to normative patterns of phonological development, educational materials can be specifically developed for automatic assessment and training programs in speech therapy (Petinou & Theodorou, 2019).

Previous research has reported that children tend to perform better in single-word speech production for emerging speech sounds, whereas established speech sounds are more accurate in conversational speech (Morrison & Shriberg, 1992). Developmental patterns of PCC and AccWord/IntScore were correlated among subjects in the normative data set (single-word) and the hearing data set (multiword). While individual consonants may differ slightly, overall patterns were comparable. The findings that PCC was correlated with AccWord/IntScore in both data sets suggest that for initial screenings of consonant production accuracy, single-word recordings may yield similar assessment results to multiword recordings, but at a much lower cost in terms of data collection and processing. While segment accuracy rates and phonological processes (e.g., epenthesis, omission, and substitution) are efficient indicators for the clinical screening of possible speech production problems, phonological properties (e.g., lexical tone and syllable structure) should also be considered. For instance, the low accuracy rate of /o/ in the normative data set may be attributed to the word "dinosaur" /kʰoŋ loŋ/, which has a nasal coda and a Tone 3–Tone 2 combination (Tseng, 2024b). For future studies, prosodic characteristics of naturalistic speech production (e.g., rhythm and intonation) should also be considered to account for a more comprehensive description of children's speech performance. Aside from a complex account of linguistic parameters, for automated child speech assessment, a high precision rate of model prediction is required in clinical applications (e.g., phoneme recognition model and tone classifier). Therefore, access to high-quality, annotated training data like those presented in the current study is crucial for accurate and reliable evaluations.

Conclusions

This study advanced the understanding of consonant production in different groups of Taiwan Mandarin-speaking children using a large annotated normative data set and a multiword data set of hearing and hard of hearing children. While phonemic transcription focuses on phonologically contrastive features, intelligibility encompasses a broad range of communicative effectiveness. The normative data set features developmental growth curves and error analysis based on phonological characteristics. Comparative studies with a multiword data set further support potential applications for utilizing these normative patterns as reference standards. Despite the limitations of uncontrolled child-level variables in the data sets used in this study, the correlation between consonant accuracy rate and acceptability/intelligibility rating across data sets suggests that single-word speech assessments may yield reliable insights as effective as continuous speech. Moreover, large-scale, annotated data sets of child speech from a wide variety of speaker groups and speaking styles may be greatly beneficial to the development of model-assisted tools for early detection of speech issues, especially in DHH children.

Data Availability Statement

The data sets generated and/or analyzed during the current study are not publicly available due to ethical restrictions but are available from the corresponding author on reasonable request. Descriptions about the project can be found at https://tmc.ling.sinica.edu.tw/corpus_list_en/.

Acknowledgments

This study was financially supported by the Institute of Linguistics, Academia Sinica, the National Science and Technology Council, under Grants NSC98-2410-H-001-066, NSC99-2410-H-001-097, MOST109-2410-H-001-087-MY2, MOST111-2410-H-001-015-MY3, NSTC114-2410-H-001-067-MY3, and Children's Hearing Foundation (awarded to Principal Investigator: Shu-Chuan Tseng).

References

Amayreh, M. M., & Dyson, A. T. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research, 41*(3), 642–653. https://doi.org/10.1044/jslhr.4103.642

Bauman-Waengler, J. (2019). *Articulation & phonology in speech sound disorders: A clinical focus* (6th ed.). Pearson Education. https://www.pearson.com/en-us/subject-catalog/p/Bauman-Waengler-Articulation-and-Phonology-in-Speech-Sound-Disorders-A-Clinical-Focus-Plus-Pearson-e-Text-2-0-Access-Card-Package-6th-Edition/P200000000757

Bernthal, J. E., Bankson, N. W., & Flipsen, P. J. (2016). *Articulation and phonological disorders—Speech sound disorders in children* (8th ed.). Pearson Education.

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (Version 6.2.14). Retrieved May 24, 2022, from http://www.praat.org/

Chang, Y.-N., & Yeh, L.-L. (2019). Assessment practices followed by speech-language pathologists for clients with suspected speech sound disorder in Taiwan: A survey study. *Rehabilitation Practice and Science, 47*(1), 31–47. https://doi.org/10.6315/TJPMR.201906_47(1).0004

Cho, S.-C. (2008). *The phonological development of 3 to 6 year-old preschool children in Taiwan* [Unpublished master's thesis, National Taipei University of Nursing and Health Sciences].

Crowe, K., & McLeod, S. (2020). Children's English consonant acquisition in the United States: A review. *American Journal of Speech-Language Pathology, 29*(4), 2155–2169. https://doi.org/10.1044/2020_AJSLP-19-00168

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English–speaking children. *Clinical Linguistics & Phonetics, 17*(8), 617–643. https://doi.org/10.1080/0269920031000111348

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). OUP Oxford. https://doi.org/10.1093/oso/9780199215782.001.0001

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics, 22*(12), 937–956. https://doi.org/10.1080/02699200802330223

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics, 28*(4), 377–396. https://doi.org/10.1006/jpho.2000.0121

Hochmann, J. R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: Different roles in early language acquisition. *Developmental Science, 14*(6), 1445–1458. https://doi.org/10.1111/j.1467-7687.2011.01089.x

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language, 27*(1), 3–42. https://doi.org/10.1017/S030500099900402X

Hustad, K. C., Mahr, T. J., Natzke, P., & Rathouz, P. J. (2021). Speech development between 30 and 119 months in typical children I: Intelligibility growth curves for single-word and multi-word productions. *Journal of Speech, Language, and Hearing Research, 64*(10), 3707–3719. https://doi.org/10.1044/2021_JSLHR-21-00142

Iyer, S. N., Jung, J., & Ertmer, D. J. (2017). Consonant acquisition in young cochlear implant recipients and their typically developing peers. *American Journal of Speech-Language Pathology, 26*(2), 413–427. https://doi.org/10.1044/2016_AJSLP-16-0073

Jain, R., Barcovschi, A., Yiwere, M., Corcoran, P., & Cucu, H. (2023). Adaptation of whisper models to child speech recognition. *Proceedings of Interspeech 2023,* 5242–5246. https://doi.org/10.21437/Interspeech.2023-935

Jeng, J.-Y. (2011). The phonological processes of syllable-initial consonants spoken by the preschool children of Mandarin Chinese. *Journal of Special Education, 34,* 135–170. https://doi.org/10.6768/JSE.201112.0137

Kim, M., Kim, S. J., & Stoel-Gammon, C. (2017). Phonological acquisition of Korean consonants in conversational speech produced by young Korean children. *Journal of Child Language, 44*(4), 1010–1023. https://doi.org/10.1017/S0305000916000258

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics, 33*(2), 363–374. https://doi.org/10.2307/2529786

Lee, S. A. S., Nguyen Phuoc, T. M., Truong Thi, N. Q., Dang Thi, H. T., Ho Thi, L. M., & Ha, N. C. (2024). Phonological acquisition in Vietnamese-speaking children with central dialect. *Speech, Language and Hearing, 27*(2), 87–103. https://doi.org/10.1080/2050571X.2023.2289252

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*(2), 185–199. https://doi.org/10.1017/S0305000900001598

Li, X. (2018). *The phonological development of Putonghua-speaking children with and without speech sound disorders* [Doctoral dissertation, The University of Hong Kong].

Li, X. X., & To, C. K. S. (2017). A review of phonological development of Mandarin-speaking children. *American Journal of Speech-Language Pathology, 26*(4), 1262–1278. https://doi.org/10.1044/2017_AJSLP-16-0061

Lin, G. B.-G., & Lin, M.-H. (1994). A study on the development of language abilities in preschool children in Taiwan. *ROC Bulletin of Special Education, 10,* 259–281. https://bse.spe.ntnu.edu.tw/BSE/upload/journal/prog/f374b026_20220329.pdf [PDF]

Lin, Y.-H. (2007). *The sounds of Chinese with audio CD* (Vol. 1). Cambridge University Press.

Lin, Y.-S., & Tseng, S.-C. (2021). Classifying speech intelligibility levels of hearing and hearing-impaired children in two continuous speech styles. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 7763–7767. https://doi.org/10.1109/ICASSP39728.2021.9414986

Lin, Y.-S., Tseng, S.-C., & Jang, J.-S. R. (2024). Leveraging phonemic transcription and whisper toward clinically significant indices for automatic child speech assessment. *Proceedings of Interspeech 2024,* 2440–2444. https://doi.org/10.21437/Interspeech.2024-1887

Liu, Y.-F., Tseng, S.-C., & Jang, J.-S. R. (2016). Deriving disyllabic word variants from a Chinese conversational speech corpus. *The Journal of the Acoustical Society of America, 140*(1), 308–321. https://doi.org/10.1121/1.4954745

Lockart, R., & McLeod, S. (2013). Factors that enhance English-speaking speech-language pathologists' transcription of Cantonese-speaking children's consonants. *American Journal of Speech-Language Pathology, 22*(3), 523–539. https://doi.org/10.1044/1058-0360(2012/12-0009)

May-Mederake, B. (2012). Early intervention and assessment of speech and language development in young children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology, 76*(7), 939–946. https://doi.org/10.1016/j.ijporl.2012.02.051

McLeod, S., & Baker, E. (2017). *Children's speech: An evidence-based approach to assessment and intervention*. Pearson.

McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology, 27*(4), 1546–1571. https://doi.org/10.1044/2018_AJSLP-17-0100

McLeod, S., Crowe, K., & Shahaeian, A. (2015). Intelligibility in Context Scale: Normative and validation data for English-speaking preschoolers. *Language, Speech, and Hearing Services in Schools, 46*(3), 266–276. https://doi.org/10.1044/2015_LSHSS-14-0120

McLeod, S., Harrison, L. J., & McCormack, J. (2012). The Intelligibility in Context Scale: Validity and reliability of a subjective rating measure. *Journal of Speech, Language, and Hearing Research, 55*(2), 648–656. https://doi.org/10.1044/1092-4388(2011/10-0130)

Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech, Language, and Hearing Research, 35*(2), 259–273. https://doi.org/10.1044/jshr.3502.259

Petinou, K., & Theodorou, E. (2019). Promoting speech intelligibility through phonologically dense targets. *Clinical Linguistics & Phonetics, 33*(10–11), 978–990. https://doi.org/10.1080/02699206.2019.1597168

Polyanskaya, L., & Ordin, M. (2015). Acquisition of speech rhythm in first language. *The Journal of the Acoustical Society of America, 138*(3), EL199–EL204. https://doi.org/10.1121/1.4929616

Pommée, T., Balaguer, M., Mauclair, J., Pinquier, J., & Woisard, V. (2022). Intelligibility and comprehensibility: A Delphi consensus study. *International Journal of Language & Communication Disorders, 57*(1), 21–41. https://doi.org/10.1111/1460-6984.12672

Poole, I. (1934). Genetic development of articulation of consonant sounds in speech. *The Elementary English Review, 11*(6), 159–161. https://www.jstor.org/stable/41381777

Priester, G. H., Post, W. J., & Goorhuis-Brouwer, S. M. (2011). Phonetic and phonemic acquisition: Normative data in English and Dutch speech sound development. *International Journal of Pediatric Otorhinolaryngology, 75*(4), 592–596. https://doi.org/10.1016/j.ijporl.2011.01.027

Räsänen, O. (2012). Computational modeling of phonetic and lexical learning in early language acquisition: Existing models and future directions. *Speech Communication, 54*(9), 975–997. https://doi.org/10.1016/j.specom.2012.05.001

Redford, M. A., & Howson, P. J. (2022). Acoustic correlates and listener ratings of function word reduction in child versus adult speech. *The Journal of the Acoustical Society of America, 152*(3), 1463–1475. https://doi.org/10.1121/10.0013835

Romani, C., Galuzzi, C., Guariglia, C., & Goslin, J. (2017). Comparing phoneme frequency, age of acquisition, and loss in aphasia: Implications for phonological universals. *Cognitive Neuropsychology, 34*(7–8), 449–471. https://doi.org/10.1080/02643294.2017.1369942

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37*(1), 55–63. https://doi.org/10.1044/jshd.3701.55

Shriberg, L., Austin, D., Lewis, B. A., Mcsweeny, J. L., & Wilson, D. L. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*(4), 708–722. https://doi.org/10.1044/jslhr.4004.708

Shriberg, L. D., Kent, R. D., Mcallister, T., & Preston, J. L. (2020). *Clinical phonetics* (5th ed.). Pearson Education. https://www.pearson.com/en-us/subject-catalog/p/clinical-phonetics/P200000001833/9780136945529

Shriberg, L., & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders, 47*(3), 256–270. https://doi.org/10.1044/jshd.4703.256

So, L. K. H., & Dodd, B. J. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language, 22*(3), 473–495. https://doi.org/10.1017/S0305000900009922

Templin, M. C. (1957). *Certain language skills in children: Their development and interrelationships*. University of Minnesota. https://doi.org/10.5749/j.ctttv2st

To, C. K., Cheung, P. S., & McLeod, S. (2013). A population study of children's acquisition of Hong Kong Cantonese consonants, vowels, and tones. *Journal of Speech, Language, and Hearing Research, 56*(1), 103–122. https://doi.org/10.1044/1092-4388(2012/11-0080)

Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics, 10*(4), 401–413. https://doi.org/10.1075/prag.10.4.01tom

Treiman, R., & Cassar, M. (1997). Can children and adults focus on sound as opposed to spelling in a phoneme counting task? *Developmental Psychology, 33*(5), 771–780. https://doi.org/10.1037//0012-1649.33.5.771

Tseng, S.-C. (2019). ILAS Chinese spoken language resources. In *The third international symposium on linguistic patterns in spontaneous speech* (pp. 13–20). Institute of Linguistics, Academia Sinica, Taipei.

Tseng, S.-C. (2024a). Corpus-based research on speech acquisition and automatic assessment of Taiwan Mandarin-speaking children aged 3 to 6. In S.-C. Tseng & E. Zeitoun (Eds.), *Linguistic diversity, but unity in research: Celebrating the twentieth anniversary of the Institute of Linguistics, Academia Sinica* (pp. 475–502). Institute of Linguistics, Academia Sinica. https://www.ling.sinica.edu.tw/item/zh-tw?act=publish_book&code=view&bookID=143

Tseng, S.-C. (2024b). Tone acquisition in Chinese-speaking children: Developmental data of tone acceptability and contour pattern. *Proceedings of Speech Prosody 2024,* 51–55. https://doi.org/10.21437/SpeechProsody.2024-11

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, 25*(9), 784–803. https://doi.org/10.3109/02699206.2011.565906

Tseng, S.-C., & Liu, Y.-F. (2021). Segment and tone production in continuous speech of hearing and hearing-impaired children. *Proceedings of Interspeech 2021*, 606–610. https://doi.org/10.21437/Interspeech.2021-757

Tseng, S.-C., Liu, Y.-F., & Lu, X.-L. (2023). Model-assisted lexical tone evaluation of three-year-old Chinese-speaking children by also considering segment production. *Proceedings of Interspeech 2023,* 3909–3913. https://doi.org/10.21437/Interspeech.2023-161

Usha, G. P., & Alex, J. S. R. (2023). Speech assessment tool methods for speech impaired children: A systematic literature review on the state-of-the-art in speech impairment analysis. *Multimedia Tools and Applications, 82,* 35021–35058. https://doi.org/10.1007/s11042-023-14913-0

Wellman, B. L., Case, I. M., Mengert, I. G., & Bradbury, D. E. (1931). *Speech sounds by young children*. The University of Iowa.

Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research, 48*(5), 1065–1079. https://doi.org/10.1044/1092-4388(2005/074)

Wu, H., Lu, F., Yu, B., & Liu, Q. (2020). Phonological acquisition and development in Putonghua-speaking children with autism spectrum disorders. *Clinical Linguistics & Phonetics, 34*(9), 844–860. https://doi.org/10.1080/02699206.2019.1702720

Yang, J., Brown, E., Fox, R. A., & Xu, L. (2015). Acoustic properties of vowel production in prelingually deafened Mandarin-speaking children with cochlear implants. *The Journal of the Acoustical Society of America, 138*(5), 2791–2799. https://doi.org/10.1121/1.4932165

Zhang, X., Qin, F., Chen, Z., Gao, L., Qiu, G., & Lu, S. (2020). Fast screening for children's developmental language disorders via comprehensive speech ability evaluation—Using a novel deep learning framework. *Annals of Translational Medicine, 8*(11), Article 707. https://doi.org/10.21037/atm-19-3097

Appendix A

Sinica Child Balanced Wordlist: Stimuli Used for the Normative Data Set

Word *Pinyin* IPA	**Word *Pinyin* IPA**
hen *mǔjī* /mu tɕi/	ride horse *qímǎ* /tɕʰi ma/
bee *mìfēng* /mi fəŋ/	walk *zǒulù* /tsou lu/
dinosaur *kǒnglóng* /kʰoŋ loŋ/	turn off light *guāndēng* /kwan təŋ/
eagle *lǎoyīng* /lao iŋ/	sweep *sǎodì* /sao ti/
turtle *wūguī* /u kwei/	sleep *shuìjiào* /ʂwei tɕjao/
rabbit *tùzi* /tʰu tsɨ/	go shopping *mǎicài* /mai tsʰai/
tiger *lǎohǔ* /lao hu/	climb mountain *páshān* /pʰa ʂan/
crab *pángxiè* /pʰaŋ ɕje/	put on clothes *chuānyīfú* /tʂʰwan i fu/
spider *zhīzhū* /tʂɯ tʂu/	TV *diànshì* /tjen ʂɯ/
swan *tiāné* /tʰjen ə/	tire *lúntāi* /lun tʰai/
hedgehog *cìwèi* /tsʰɨ wei/	window *chuānghù* /tʂʰwaŋ xu/
building blocks *jīmù* /tɕi mu/	donut *tiántiánquān* /tʰjen tʰjen tɕʰyen/
straw *xīguǎn* /ɕi kwan/	ugly duckling *chǒuxiǎoyā* /tʂʰou ɕjao ja/
hot dog *règǒu* /ʐə kou/	clock *shízhōng* /ʂɯ tʂoŋ/
cake *dàngāo* /tan kao/	teacup *chábēi* /tʂʰa pei/
mango *mángguǒ* /maŋ kwo/	leather shoes *píxié* /pʰi ɕje/
juice *guǒzhī* /kwo tʂɯ/	toy *wánjù* /wan tɕy/
milk *niúnǎi* /njou nai/	button *niǔkòu* /njou kʰou/
strawberry *cǎoméi* /tsʰao mei/	dish *pánzi* /pʰan tsɨ/
grapes *pútáo* /pʰu tʰao/	crayon *cǎisèbǐ* /tsʰai sə pi/
steak *niúpái* /njou pʰai/	thermometer *wēndùjì* /wən tu tɕi/
apple *píngguǒ* /pʰiŋ kwo/	football *zúqiú* /tsu tɕʰjou/
sushi *shòusī* /ʂou sɨ/	jigsaw *pīntú* /pʰin tʰu/
car *qìchē* /tɕʰi tʂʰə/	Monopoly *dàfùwēng* /ta fu wəŋ/
chopsticks *kuàizi* /kʰwai tsɨ/	steamed bread *mántóu* /man tʰou/
airplane *fēijī* /fei tɕi/	blow bubbles *chuīpàopào* /tʂʰwei pʰao pʰao/
train *huǒchē* /hwo tʂʰə/	school *xuéxiào* /ɕye ɕjao/
ears *ěrduo* /ɚ two/	kitchen *chúfáng* /tʂʰu faŋ/
teeth *yáchǐ* /ja tʂʰɯ/	living room *kètīng* /kʰə tʰiŋ/
mouth *zuǐbā* /tswei pa/	garden *huāyuán* /xwa yen/
speak *shuōhuà* /ʂwo hwa/	fountain *pēnshuǐchí* /pʰən ʂwei tʂʰɯ/
write *xiězì* /ɕje tsɨ/	cloud *báiyún* /pai yn/
have a meal *chīfàn* /tʂʰɯ fan/	moon *yuèliàng* /ye ljaŋ/
get wet in rain *línyǔ* /lin y/	cliff *duànyá* /twan jai/
swim *yóuyǒng* /jou joŋ/	birthday *shēngrì* /ʂəŋ ʐɯ/

Appendix B

Eighteen Sentences: Stimuli Used for the Multiword Data Set

Phrases/sentences	**Pinyin**	**IPA**
I can watch TV.	*wǒ kě yǐ kàn diàn shì*	/wo kʰə i kʰan tjen ʂɯ/
We can watch TV.	*wǒ men kě yǐ kàn diàn shì*	/wo mən kʰə i kʰan tjen ʂɯ/
Please pass me the book. (singular)	*qǐng nǐ bǎ shū gěi wǒ*	/tɕʰiŋ ni pa ʂu kei wo/
Please pass me the book. (plural)	*qǐng nǐ men bǎ shū gěi wǒ*	/tɕʰiŋ ni mən pa ʂu kei wo/
He likes cakes.	*tā xǐ huān dàn gāo*	/tʰa ɕi xwan tan kao/
They like dancing.	*tā men xǐ huān tiào wǔ*	/tʰa mən ɕi xwan tʰjao u/
Three adults.	*sān gè dà rén*	/san kə ta ʐən/
A child.	*yí gè xiǎo hái*	/i kə ɕjao xai/
A chicken.	*yì zhī xiǎo jī*	/i tʂɯ ɕjao tɕi/
A box of coloring pencils.	*yì hé cǎi sè bǐ*	/i xə tsʰai sə pi/
(Someone) won't come.	*bú huì lái*	/pu xwei lai/
I can write.	*wǒ huì xiě zì*	/wo xwei ɕje tsɨ/
I can eat by myself.	*wǒ huì zì jǐ chī fàn*	/wo xwei tsɨ tɕi tʂʰɯ fan/
Black dogs are beautiful.	*hēi sè de gǒu hěn piào liàng*	/xei sə tə kou xən pʰjao ljaŋ/
Do you like dinosaurs or not?	*xǐ bù xǐ huān kǒng lóng*	/ɕi pu ɕi xwan kʰoŋ loŋ/
He likes yellow.	*tā xǐ huān huáng sè*	/tʰa ɕi xwan xwaŋ sə/
He likes yellow cakes.	*tā xǐ huān huáng sè de dàn gāo*	/tʰa ɕi xwan xwaŋ sə tə tan kao/
Don't stay in rain.	*bù kě yǐ lín yǔ*	/pu kʰə i lin y/

Note. Adapted with permission from Tseng et al. (2011).