NORMATIVE DATABASE OF WORD PRODUCTION OF PUTONGHUA-SPEAKING CHILDREN - BEIJING ARTICULATION NORMS PROJECT:CASS_CHILD_WORD

Jun Gao, Aijun Li, Ziyu Xiong, Jiaxuan Shen
Institute of Linguistics
Chinese Academy of Social Sciences
Beijing, China

Ying Pan
Health Care Department
Beijing Maternal and Child Health Care Hospital
Beijing, China

Abstract

The database is based on normative data of the phonological development of Putonghua-speaking children. Single word production of around 4000 normally-developing Putonghua-speaking children aged between 1;6 and 6;0 years in Beijing city proper was collected. The speech samples were obtained through picture-naming task. What was of primary concerns were the order of the acquisition of speech sounds and the typical error patterns in normally-developing children within different age bands. The normative data can offer reference for theoretical exploration, clinical assessment, intervention and rehabilitation, and help improve children speech recognition.

Keywords—normative data, phonological development, Putonghua-speaking children, word production, CASS_CHILD_WORD

I. INTRODUCTION

In terms of the production aspect of phonological development, compared with adults' speech production, normally-developing children's speech production will experience the development from being deviant form adults' targets to being adult-like. Some speech sounds are acquired earlier by children than others. And there is a great deal of individual differences among children in speech sound errors they make. What's more, children with speech delays or disorders also make errors when they speak. The difference between speech-delayed and -disordered children and normally-developing children lies only in the number of the speech sounds they make errors with not in what errors they make and what phonological processes they use. "It therefore seems likely that children with delayed development constitute the lower range of normal development" (Zhu, 2002). In this sense, large sampling with narrow age bands is crucial to establish the norm of phonological development.

In order to provide effective screening tool for clinical assessment, single-word productions of around 4000 children, aged 1.5 year to 6 year, in the city proper of Beijing, were collected cross-sectionally to obtain normative data. The aim for the collection of normative data is to give reliable and comprehensive description of the phonological profile of normally-developing children. The labeled data can be used for the exploration of phonological development theories, cross-linguistic comparison, clinical practice for screening out children with speech delays and disorders and improving child speech recognition.

II. NORMATIVE DATABASE: CASS_CHILD_WORD

A. Word collection and selection

The words for picture-naming task were collected from a variety of digital sources, including the transcribed Chinese texts from the database of CHILDES (Child Language Data Exchange System, http://childes.psy.cmu.edu/) (including folders of *Beijing2* (with the permission of Dr. Twila Tardif), *Chang*, *Context*, *Zhou1*, *Zhou2* under the path of 'EastAsian/Chinese/'), classic Chinese and foreign fairy tales, popular Chinese children fables and stories, CASS (Chinese Academy of Social Sciences) transcribed texts of the recordings of children's speech in Zuojiazhuang Kindergarten, transcribed texts of CASS_CHILD corpus of the longitudinal recordings of one child between 1 and 4 years and three one-hour-long pieces of the longitudinal recording of two other children and transcribed texts of the recordings of the Question-Answer dialogues between 16 children and their parents (Gao, Li & Xiong, 2012), words from Putonghua-CDI (PCDI) questionnaires (411 words for 8-16m infants; 799 words for 16-30m children) (Tardif, Fletcher, Zhang, Liang, 2008), many kinds of books particular for 1-3 year olds (story books and books to teach children words), word lists in the textbooks for students of primary schools.

A custom software derived a word list from these materials by segmenting the texts into words and tagged the words with parts of speech through ICTCLAS (2011) (http://sewm.pku.edu.cn/QA/reference/ICTCLAS/FreeICTCLAS/) and automatically transformed the word characters into Chinese Pinyin (Chinese romanisation system). At the same time word frequency was counted for each word to be a kind of reference for word selection in the later phase. The results of the auto-segmentation and -transformation were checked manually. The auto-transformation of characters into Pinyin was checked for words with more than one pronunciation and words with neutral tones and words with retroflexers. The phonological properties of each word were analyzed as well, such as the number of syllables, syllables of the word with lexical tones, syllables of the word without lexical tones, syllable initials, syllable finals, lexical tones, types of syllable initials (e.g. stops, fricatives, affricates, nasals, laterals and approximants), syllable finals (type I) (e.g. Open, Stretched, Round, Protruded), syllable finals (type II) (e.g. monophthongs, diphthongs, triphthongs, vowels with nasal endings). For each syllable of the word, its position in the word was specified, i.e. initial, medial or final. These phonological details of the words were prepared for making test papers for children's production and also for the analysis of the children's production of these words.

Out of the raw texts collected, around 15,000 words were extracted. From these words, around 1000 words were chosen based on word frequency, familiarity to the children and imageability of the words. Owing to the fact that nouns can be embodied more easily with pictures, most words chosen were nouns. According to categories of nouns in PCDI (Tardif et al., 2008), nouns related to food, animals, vegetables, fruits, daily commodities, furniture, toys, colors, body parts, transportation tools and roles in the stories were chosen. When we were choosing words, we avoided choosing words which were culturally- and times- restricted or words that were popular only for the time being, like the roles in the popular cartoons, or proper names, like KFC and McDonald's, or famous brand names for cars like BMW.

Regarding the 1000 words, a child is not able to produce all of them at a time. And for the practice of collecting a large number of samples, time of recording of one sample should be limited to a reasonable extent. Practically, a test paper of around or below 100 words is an ideal size both for a child and for large sampling. Taking into the cognitive ability of children into account, out of the 1000 words, two probe word lists were created, one of 300 words for children below 4 years and the other of 500 words for children above 4 years. The two probe word lists were made with the consideration to cover in Putonghua phonological system syllable initials, syllable finals, lexical tones, lexical tone combinations of two-syllable words, combinations of syllable initials and type I finals, combinations of syllable initials and type II finals, all kinds of neutral tones, all kinds of retroflexers, Tone3-Tone3 tone sandhi, minimal pairs of words, and pairs of words with the same syllable finals.

In Putonghua phonological system, there are 22 initials (including zero syllable initial), 37 finals, 4 lexical tones, 16 two-syllable lexical tone combinations (including the original tone combination of Tone3 and Tone3). Syllable finals can be categorized under either type I (Open, Stretched, Round, Protruded) or type II (e.g. monophthongs, diphthongs, triphthongs, vowels with nasal endings). Different type II syllable finals can be grouped under the same type I syllable finals. For instance, all the type II finals starting with [i] (e.g. [i], [ia], [iou], [in], etc.) are grouped under Stretched; all the type II finals starting with [u] (e.g. [u], [ua], [uo], [uan], etc.) under Round; all the type II finals starting with [y] (e.g. [y], [yn], etc.) under Protruded; and the rest of the type II syllable finals are grouped under Open. Both syllable-initial-type-I-final combinations and syllable-initial-type-II-final combinations have their own phonotactic rules. Syllable-initial-type-II-final combinations form syllable types. In Putonghua phonological system, there are around 400 syllables (syllable-initial-type-II-final combinations) without considering lexical tones and 55 syllable-initial-type-I-final combinations. For some syllable-initial-type-I-final combinations, even for older children, only one or two words appropriate can be found to represent them. Therefore, the 300-word list covered less syllable-initial-type-I-final combinations than the 500-word list which covered all the 55 combinations. The two wordlists had more than half words in common. To cover syllable-initial-type-I-final combinations, in the 500-word list, less familiar words like *Buddha*, *well* (noun) and *axe* were included. Due to the limited number of words, the two word lists couldn't cover the assessment of the syllable initials in all word positions (such as word-medial and word-final). Most words in the two word lists were nouns and a few were verbs like *to fish*, *to cry*, *to brush teeth*; adjectives like *big*, *fat*, *short*, *many*, *thick*, *light*, *soft* and adverbs like *fast*.

B. Making test papers

Out of the two word lists, papers were made for children of different ages. According to past research (Li, Zhu, Dodd et al., 2000; Zhu, 2002; Si, 2006; Yuan, 2009), most errors children make are on syllable initials. So the first requirement each test paper must meet was to cover all the syllable initials in Putonghua phonological system. The way to meet the requirement was to have each paper cover all kinds of the syllable-initial-type-I-final combinations existent in the word lists. The coverage of combinations entailed the coverage of all the syllable initials and syllable finals (type II) as well and at the same time limited the size of each paper. A custom paper-making software counted how many kinds of combinations there were in the word list and identified for each combination what were the words. And the counting was based on all the syllables in the words. The software also analyzed in the word list the number of neutral tones, retroflexers, Tone3-Tone3 tone sandhi, and minimal pairs and pairs of words with the same finals. For type II syllable finals, missing ones were spotted after the first run of the software and filled in automatically. For each automatically-made paper, there went with specifications of the distribution of the phonological properties of the paper, such as syllable initials, finals, lexical tones, two-syllable-lexical-tone combinations, words with neutral tones, retroflexers and tone sandhi, minimal and pairs of words with the same finals. Based on the information, more words were added manually to make the paper cover all the phonological aspects characteristic of Putonghua phonological system and with which we wanted to test children.

Three kinds of papers were made out of the two word lists for three age groups of children. One kind of papers was for children aged 18 to 30 months, another kind of papers was for children aged 31 to 47 months, and the third kind for children at and above 48 months. The first two kinds of papers were made out of the 300-word list. The third kind was made out of the 500-word list. For each kind, there were 10 papers, each having reasonable number of words and similar phonological distribution. The number of words tested with 18-30 months was around sixty, that with 31-47 months around eighty-five and that with 48 months and above around one hundred and fifteen.

C. Papers for 18-30 months

To cover the syllable initial-type-I-final combinations and at the same time to limit the number of words in this kind of papers for the youngest children, all the syllables in a word were taken into consideration. Ideally, to satisfy the 48 combinations featured by the 300-word list, less than 48 words were needed, for some words had more than one syllable. Monosyllabic words, disyllabic and multisyllablic words were pooled to be the candidates for the 48 syllable-initial-type-I-final combinations. The paper-making software first counted the combinations in the word list based on all the syllables of the words and grouped words under each combination kind. If there were 48 kinds of combination in the word list, there should have been 48 kinds to be filled up in each paper. The software started with randomly choosing a word among the word list and analyzing the combination property of each syllable of the word. Each time a word was chosen out of the word list, one kind or more than one kind of combination was filled and the word chosen was removed from the candidate list for next papers. In selecting words for a paper, the priority was given to the word candidates in which the combination properties of all the syllables of the word had not been touched upon or to the word candidates with least overlapping in terms of combination properties with the words which were already in the paper.

When making a paper, if for some syllable-initial-type-I-final combination there were no words left in the candidate list due to the choices of the previous papers, all the words removed related to this combination when making the previous papers would be drawn back for making the current paper. From the words returned, the software would randomly choose one to fill the initial-final combination for the current paper. In consequence, different papers would share some same words.

After all the combinations were fulfilled, the software analyzed the types of syllable finals (type II) to check whether the words chosen in the paper covered all the type II syllable finals. If some were missing, the software would add to the paper automatically the words which contained the syllable finals the paper needed.

After the auto paper-making, manual adjustments were made to ensure the words in each paper cover all the basic phonological aspects in Putonghua phonological system.

D. Papers for 31-47 months

For the older group of children who were able to be tested with more words, papers were made by choosing words separately first from monosyllabic words and then from multisyllabic words (including both disyllabic and multisyllabic words). That is, the paper-making software counted the initial-final (type I) combinations both in the monosyllabic words and in multisyllabic words in the word list. The software chose words from the two parts of the word list separately and words in the papers had to cover all the combinations in each part. Therefore, the papers for 31-47 months were made up of two parts, one part from monosyllabic words and the other part from multisyllabic words. The monosyllabic-word part in the paper had to cover all kinds of the combinations in the monosyllabic words in the word list. For the multisyllabic-word part in the paper, all the syllables of the words together had to cover all the combinations. Specifically, a disyllabic word might cover two kinds of combinations, for each syllable covered one combination. A trisyllabic word might cover three kinds of combinations. And in choosing words from the multisyllable-word part, the word candidate with the syllable type (the initial-type-II-final combination) which was different as much as possible from the syllable type of the monosyllabic word of the same initial-type-I-final combination had the advantage to be chosen. This requirement aimed to make the paper cover more different syllable structures, i.e. the initial-final (type II) combinations, in case that phonological development in children is not in an across-board way but in an item-based way. The other steps of making this kind of papers were the same as those for the youngest children.

E. Papers for 48 months and above

For the oldest group who know much more and be able to be tested with more words, to make a paper include more words, the papers also consisted of two parts, one part made up of monosyllabic words and the other part of multisyllabic words. But this time, with regard to the multisyllabic word part, the paper-making software only focused on the first syllable of the multisyllabic words. The software only considered the combination properties of the first syllable of the word and ignored the combination properties of the rest in the same word. As a result, more words were needed to cover all kinds of combinations than those for the papers for 31-47m. Except this, the logic of making this kind of papers was the same as that for 31-47 months olds.

F. Participants

We cooperated with Beijing Municipal Bureau of Health, Beijing Maternal and Child Health Hospital and its subordinated district Maternal and Child Health Care Hospitals (Chaoyang, Haidian, Xicheng Nanqu, Xicheng Beiqu). They were responsible for recruiting and organizing children who met the requirements of the project to record.

This project was cross-sectional sampling. Speech samples of around 4000 children were collected, of which 320 samples were for pilot study, and 3640 were for formal recording. Children were enlisted from four districts in the city proper, Chaoyang, Haidian, Xicheng Nanqu, Xicheng Beiqu. In each district around 1000 children were sampled. The requirements for children were being healthy, with normal hearing ability and normal intelligence, and risen in Putonghua-speaking family. The way to screen Putonghua-speaking children was to see whether the child was born in Beijing with Beijing registered permanent residence. However, the way of screening children inevitably included the children only one of whose parents was a Beijinger and who might be exposed to dialects in the family. In fact, due to the difficulty in recruiting children, the participants also included the children whose parents were not Beijingers and who were exposed to accented Putonghua. The age range was 1.5 to 6 years. The age bands were as follows: eight bands below 3 years: 18-20m, 21-22m, 23-24m, 25-26m, 27-28m, 29-30m, 31-32m, 33-35m; eight bands above 3 years: 36-38m, 39-41m, 3.5y, 4y, 4.5y, 5y, 5.5y, 6y.

Errors occur most likely between 2 and 3 years (Zhu, 2000; Si, 2006), so we had narrower age band below the age of 3 years. For pilot study, each age band sampled around 20 children, half boys and half girls. For formal sampling, each age band recorded around 220 children, half boys and half girls.

Parents signed the consent form and filled up a questionnaire (modified from a questionnaire of Anamnese BabyLab Utrecht of Utrecht University) about the basic information of the child and the family. The basic information of a child included name, sex, age, hearing ability, the health status of the mother when she was pregnant and the health status of the child when he or she was just born, the birth province of the mother and the father and the dialects of the main caretakers of the child, the motor development of the child, the month age the child began to speak, the now speaking ability. Other information included: whether the child and the family members including extended family members had illnesses like alexia, stuttering, hyperactivity, hearing disability, epilepsy, autism, or depression; how much time the caretakers communicate and interact with the child; whether the child is exposed to songs, children songs, poems, nursery songs, doggerels; whether the child watches TV and whether the child can sing or dance to the song.

For children younger than 3 years, we mainly recorded at child health care departments of community health service centers. For children older than 3 years who go to kindergarten, we tested in the kindergartens.

G. Recording

Each child was tested individually. The administering took around 20 minutes.

Laptops (Dell or ThinkPad) were used to record. Some laptops recorded with a special soundcard (Lexicon IO22). When a soundcard was used, the speech was recorded with two microphones (AKG). One was headset (C520, AKG) for left channel and the other was desk-stand (C1000, AKG) for right channel. The laptops which recorded without the soundcard recorded with the microphone the laptops themselves had.

A platform was developed for recording. The platform called both Cool Edit pro 2.0 (44 KHz, 16 bit) and the recording software of the windows system to record, and presented pictures at the same time according to the preset test papers. The window of Cool Edit pro 2.0 was shown on the laptop screen which was for the examiners to monitor the amplitude of the sound. If the child spoke too low according to the amplitude of the waveform, the examiner would ask the child to speak louder. The recording software of the windows system was hidden and recording background as a backup. The windows system recording software saved recording picture by picture while Cool Edit pro 2.0 saved recording after the whole paper was finished. The window for Cool Edit pro 2.0 was adjusted to one fifth of the screen of the laptop and the window for presenting pictures of words took up four fifths of the screen. At the bottom of the picture-showing window the standardized prompts about the target word and the target word were displayed. At the top right corner of the picture-showing window, there was a button for playing the prerecorded production of the words by a female Putonghua native speaker. At the right bottom corner of the picture-showing window, there were two dots, one green and the other yellow. Green meant the correct pronunciation and yellow meant the incorrect pronunciation. The judgment was made on site for each picture by the examiner during recording. The child's production was elicited mainly by the picture or by the prompt questions posed by the accompanying caretakers or by the examiner or sometimes by imitating what the caretakers or the examiner produced or the prerecorded sounds. For recording taking place at the community health service centers, usually there were accompanying caretakers while in the kindergartens, the child came to record alone. In Putonghua, for the same word meaning, there can be more than one expression. Therefore, if the child didn't produce the word exactly as what we required, the examiner would ask the child to produce the way exactly as what we wanted. Recording usually took place in the office or classroom, so there was noise.

Before recording, the administrator in charge of the project set up the resources for recording, including pictures, prompt questions, test papers, and prompt sounds. Most pictures used to elicit the children's responses were of high definition and downloaded from www.nipic.com. For each paper, the pictures of more familiar words and the pictures which could be named easily were presented first and more difficult pictures whose responses needed prompts were presented in the later part. Within each part, the pictures were presented randomly.

On recording, the basic information of the child based on the questionnaire the caretakers filled in was input and the paper proper to the child's age was chosen (for each kind of papers, 10 papers were chosen in turn for different children within the same age range). During recording, two parts of information of recording would be input. They were basic information of the child and the family, and basic information of recording, like recording venue, recording date, name of the examiner, recording environment (in the office or in the classroom), the paper tested. After recording, basic information about the child's performance, the child's personality, the extent of cooperation, whether the child was accented would be input. And then, the examiner would fill up a paper recording form. All the information was stored in the database based on each child. The consent form and the questionnaire and the recording form written by the examiner were scanned and stored in the database as well.

During recording, the starting time and ending time for each picture were marked on Cool Edit pro 2.0 with its cue-marking function. And the times were stored in a log text file in the database with the information of the test words and the order of the test words. In testing, the pictures could go back for a second production, if necessary. Each child had a recording number comprised of recording venue plus the number of laptop plus recording date plus the serial number of recording. The recording sound file and the log text file for test words was named in the same way and stored in the database based on each child.

H. Text transcribing and phonetic labeling

A transcribing and labeling platform was developed to administrate transcribing and labeling procedures and data, give basic checks on transcribing and provide convenience to the labelers.

There were two-step labelings and two-step checks. The first step of labeling included text transcribing and phonetic labeling. Through the platform, the labelers could download the designated sounds by restricting the age or the recording venue or the recording number or the child's name. The labelers downloaded the sound from the server and the transcribing and labeling was uploaded to the server in real time. The submitted transcribed and labeled sound was stored in the database with the information of the labeler and the date of submission.

The labelers transcribed only what the child spoke (speeches of the accompanying parents were not transcribed) and identified the errors the child made if any. All the speeches made by the child including test words and other utterances were required to be transcribed and labeled. For each child under labeling, the platform offered for the labelers' reference the test information (basic information of the child and the family, basic information of recording, children's performance) and the test log for the child where words tested and the order of testing and starting time and ending time for the word were displayed. This reference of the test log could ensure to some extent that the labelers wouldn’t do the wrong transcription when the child's word production was confusing due to erroneous pronunciation. The function of the platform was to make transcribing as automatically as possible and to ease the typing workload of the labelers. When the platform opened a sound file through Praat, the labeler selected on Praat the part of waveform for a particular word and clicked the labeling button, the labeling window popped up. If the word to be transcribed was the word in the test paper, the labeler could choose the text for the word from several possible words offered by the platform which were calculated based on the selection time of the word and the time recorded by Cool Edit pro 2.0 during recording. When one word was chosen, its Pinyin and syllable initial and syllable final and lexical tone were displayed automatically in a table. So there was no need for the labeler to type. What the labeler needed to do was to label the error regarding syllable initial, final and lexical tone. At the same time, parts of speech would be offered automatically. If the pronunciation was correct, the labelers didn't need to label anything. For the words or utterances produced by the child which were not in the test paper, the labeler transcribed them manually and then transformed the characters into Pinyin by clicking a transforming button in the labeling window and divided the syllable into initial, final and lexical tone in the table form by clicking another button in the labeling window. For these non-testing words, the platform made a log of the past input words. Once one of these words was produced for a second time, the labeler could choose it in a pull-down menu under *frequent words*. Thus, the labeler didn't need to type the word once again. Besides text transcribing and error labeling, the labelers were required to label parts of speech of the word, whether the word was produced with neutral tone or with retroflexer, whether the child knows the word or not and whether the sound of the selected word was good without noise or poor with noise. To the labeling, the labelers could add comments.

The labeling was a kind of broad phonetic transcription with Chinese Pinyin. The errors made by the children could mostly be transcribed into another sound existent in Putonghua phonological system (Li, Zhu, Dodd et al., 2000; Zhu, 2002; Si, 2006; Yuan, 2009). Therefore, in the first-step labeling, for those errors which could be identified as another sound in Putonghua phonological system, we used the symbol in Chinese Pinyin system to label them whereas for those errors which could only be identified vaguely as being deviant from the adult's target, we used an asterisk * to indicate that the sound was not standard. For each syllable, which part (syllable initials, finals and lexical tones) was erroneous was identified and into what sound the child mispronounced the target was labeled. For example, if the labeler judges that the child pronounces *gou3* (Chinese Pinyin, *dog*) as *dou3* (Chinese Pinyin), he/she will label the target syllable initial *g* (Chinese Pinyin) with *d* (Chinese Pinyin). Labelers from Phonetics Lab of CASS and iFLYTEK Co., Ltd were engaged in transcribing and labeling. Before labeling, all the labelers were trained by the first author. The reliability of the identification of the erroneous sounds among labelers was higher than that of the specification or labeling of the error.

After the labeler finished transcribing and labeling, he/she could click a checking button to check whether all the words in the test paper of the child were transcribed or not. After checking, they could upload to the server the transcription and labeling.

After the first step of labeling, experienced labelers would do the first step of checking. In the first step of checking, all the information transcribed and labeled in the first step labeling was written into different tiers in the textgrid. So, the checkers checked the labeling in a classic Praat wav-textgrid way.

For the second step labeling, the labelers downloaded the sound file and the corresponding textgrid where auto-segmentation was made at the time of downloading. The main purpose for this labeling was to identify the precise starting time and ending time of each segment based on the results of auto-segmentation. The second step labeling would be subjected to the second step checking.

In the first step checking and the second step checking, the results of the first step labeling and the first step checking and the results of the second step labeling and the second-step checking could be exported for comparison and for reliability checking. The results of each step of labeling and checking were stored separately in the database. The checking results would not overwrite the results of the labeling.

I. Data analysis and error patterns

The analyses to be conducted include what phonemes are likely to be pronounced incorrectly by what age of children; what phonological processes are involved; what errors are the typical ones for a phoneme; whether the existence of some errors entails the existence of other errors, what is the percentage of the errors in the syllable initials, finals, lexical tones and the whole phoneme inventory for children at a certain age; for a phoneme, for a certain age group, what is the time of emerging and being acquired, what is the percentage of children in the age band who make pronunciation errors.

Besides the aspects mentioned above and the error patterns found in past researches, another important phenomenon needs to be paid attention to is that some errors of syllable finals are triggered by the errors of syllable initials. That is, due to the phonotactic rules of the syllable-initial-type-I-final combinations where some syllable finals can only go with a certain type of syllable initials, if the syllable initial is mispronounced into another one, the original final cannot follow the new initial according to the phonotactic rule, so the original final is changed accordingly into another final. In the past research, syllable initial and final errors are considered separately which will not make a full understanding of phonological development. The error pattern where the deviation of the syllable finals is originated from the errors of the syllable initials may reveal that children are sensitive at a very early stage to phonotactics in Putonghua phonological system, especially the phonotactics of syllable-initial-type-I-final combinations which are unique in Chinese.

The qualitative and quantitive error pattern data are to be presented for reliable clinical assessment in the 2- or 3-month age bands for children under the age of 3 years.

III. Ongoing and next-step researches

To effectively differentiate normally-developing children and children with speech delays and disorders, another important predicator needed is how long the errors endure or in what time the errors will be suppressed in development. This needs longitudinal data. Besides cross-sectional recording, we are conducting the longitudinal recording based on the test papers designed for the cross-sectional sampling. Another importance of having longitudinal data is that even for the same phoneme, the child will make different phonological processes on it at different ages. For example, for a child, YY (a longitudinal case study in CASS), she can't produce either [tɕ] or [tɕ'] correctly at around 55 months for her difficulties in producing affricates. She substituted these two sounds with [ɕ]. But at the earlier stage around 30 months, she could produce [tɕ] correctly though she could not produce the aspirated counterpart [tɕ'], substituted by [tɕ]. [ts] [ts'], and [tʂ] [tʂ'] had similar developmental pattern. When she was around 57 months, she developed to be able to produce [tɕ], [ts] and [tʂ], but could not produce [tɕ'], [ts'] and [tʂ'] which were substituted by [ɕ], [s], [ʂ] respectively. From these developments, it seems that the child is sensitive to features under segments. To understand fully the underlying mechanisms of phonological acquisition, it is best to have perception data to see whether the production errors are derived from perception impairment or due to articulation limitation. For the child, YY, when she was around 57 months, she could not produce [tɕ'], [ts'] and [tʂ'] and substituted them with [ɕ], [s], [ʂ] respectively. When she was tested with the discrimination and identification of the minimal pairs involving these phonemes, she could identify and discriminate accurately and quickly. Even though this means she has no problem with perception, maybe during development, in some period, her discrimination of these phonemes is confusing. In the next step, we will design a word text database with which older children with less phoneme errors can be tested with the production of more words related to the erroneous phonemes which can cover syllable structures as many as possible and with the perception with minimal pairs involving the phonemes. Furthermore, samples of children with speech delays and disorders will be collected.

Based on the normative data and sounds, machine assessment and transcription can be possible. According to the normative data, we can know what sounds one phoneme is usually mispronounced as. With the limited number of candidates for the judgment of a phoneme for the machine, its identification and transcription can be more accurate. Thus clinicians who are not phonetic professionals can do the assessment.

Acknowledgment

Thanks are given to iFLYTEK Co., Ltd for recording and labeling, and Beijing Municipal Bureau of Health, Beijing Maternal and Child Health Hospital, and its subordinated district Maternal and Child Health Care Hospitals (Chaoyang, Haidian, Xicheng Nanqu, Xicheng Beiqu) and all the community health service centers and kindergartens which helped recruit and organize children and all the children and parents participating in the project.

III. References

[1] H. Zhu, Phonological development in specific contexts: Studies of Chinese-speaking children: Multilingual Matters Ltd., 2002.

[2] J. Gao, A. Li, and Z. Xiong, "MANDARIN MULTIMEDIA CHILD SPEECH CORPUS: CASS_CHILD," in the Oriental-COCOSDA 2012.

[3] T. Tardif, P. Fletcher, Z. Zhang and W. Liang, Hanyu Goutong Fazhan Liangbiao Shiyong Shouce: Beijing University Press, 2008.

[4] W. Li, H. Zhu, B. Dodd, et al. "Phoneme acquisition by Chinese-speaking children," Journal of Psychology, 2000.

[5] Y. Si, "Case study of phoneme acquisition of a Mandarin-speaking child, " Contemporary Linguistics, 2006.

[6] Y. Yuan, "Syllable initial errors by Putonghua-speaking children" M.A. thesis. Beijing: Beijing Normal University, 2009.

[This paper was published in O-COCOSDA, 2013]