Running head: PERCEPTION AND PRODUCTION OF MANDARIN TONES

Perception and Production of Lexical Tones by 3-Year-Old Mandarin-Speaking Children

Puisan Wong, Richard G. Schwartz, & James J. Jenkins

The Graduate Center, The City University of New York

Abstract

The present study investigated three-year-old children's perception and production of Mandarin lexical tones in monosyllabic words in isolation and in sentence final position. Thirteen three-year-old Mandarin-speaking children participated in the study. Tone perception was examined by a picture-pointing task and tone production was investigated by picture naming. To compare children's productions to the adult forms, four mothers of the 13 children were asked to say the same set of words to their children in a picture reading activity. The children's and mothers' productions were low-pass filtered at 500 Hz and 400 Hz respectively to eliminate segmental information. Ten Mandarin-speaking judges were recruited to identify the productions of tones from the filtered speech. The results revealed that three-year-old Mandarin-speaking children perceived the four lexical tones with high accuracy. Contrary to the findings in previous studies, three-year-old children's production of the four Mandarin tones was, however, not yet adult-like. The children had the greatest difficulties with the dipping tone.

Perception and Production of Lexical Tones by 3-Year-Old Mandarin-Speaking Children

Mandarin, the most widely spoken language in the world, is one of the tone languages that uses pitch differences to contrast word meanings (Yip, 2002). Mandarin has four lexical tones: Tone 1, Tone 2, Tone 3 and Tone 4. Each tone is realized largely by fundamental frequency changes (Li & Thompson, 1989). When spoken in isolation, Tone 1 has an essentially level fundamental frequency contour with a slight dip in the middle of the vowel and a slight rise toward the end of the syllable (Ho, 1976). Its tone contour starts and ends in a speaker's high pitch range (Shih, 1988). Tone 2 is a rising contour with a slight dip at the 15% point of the mean duration of the syllable nucleus (Ho, 1976). Its fundamental frequency contour starts at the speaker's mid pitch range and rises up to the high pitch range at the end (Shih, 1988). Tone 3 has a falling and rising contour with an inflection point coming after 41% of the mean duration of the whole syllable nucleus (Ho, 1976). It starts at the speaker's low mid pitch range, falls to the low pitch range and rises to the high mid pitch range. Tone 4 has a falling contour. It starts at the speaker's high pitch range and falls to the low pitch range at the end (Shih, 1988). Thus, the four tones are also referred to as the level tone (Tone 1), the rising tone (Tone 2), the dipping tone (Tone 3) and the falling tone (Tone 4). The same syllable spoken in different tones can have distinctive meanings. For example, when the syllable 'ma' is spoken with the level tone (i.e., 'ma1'), it means 'mother'; with the rising tone (i.e., 'ma2'), it means 'hemp'; with the dipping tone (i.e., 'ma3'), it means 'horse'; and with the falling tone (i.e., 'ma4'), it means 'scold'.

Several studies have examined young children's production of Mandarin tones. Some have been longitudinal studies with small numbers of children. The first such

study, published in 1951, documented the phonological development of a 28-month-old girl acquiring Mandarin as a first language in the U.S (Chao, 1973). The author, who was the grandfather of the child, recorded and analyzed the child's vocabulary and phonology during a one-month observation period. Very little information was provided about the child's productions of tones. Reportedly, the child produced tones in isolation correctly at the beginning of the study, and only had some difficulties with the tone sandhi rules, the change of tones when particular tones are adjacent to each other in connected speech.

Clumeck (1977) reported longitudinal data for three children. The study described the phonological development of M, a boy born to and raised in a Shanghainese- (another tone dialect in China) and Mandarin-speaking family residing in the U.S. Language samples were collected and analyzed from 14 to 32 months of age. The child used the rising pitch for all his words at the age of 1;10 and started to produce some level and falling tones at the age of 1;11. Another boy, P, was studied from 2;3 to 3;5 and a girl, J, from 1;0 to 2;10. Both children were born and raised in monolingual Mandarin-speaking families in the U.S. Tone productions in isolation and in utterance-final position were analyzed from imitated, elicited, and spontaneous speech. Both children mastered the level tone and the falling tone "almost" completely and had some difficulties with the production of the rising and dipping tones (Clumeck, 1980).

The most recent longitudinal study followed four children from 10 to 24 months in Beijing, China (Zhu, 2002). The children's productions of the four tones were reported to be stabilized between the age of 1;4-1;9.

Children's production of Mandarin tones has also been examined cross-sectionally in large groups of children. The first included 17 Mandarin-speaking children

(aged 1;6-3;0) from Taiwan. Ten of the seventeen children were also followed longitudinally for seven months. Tone productions were elicited using a picture naming task. The authors proposed four stages of tone development. At stage I (limited vocabulary, single-word utterances) the level and falling tones were predominant. At stage II (larger vocabulary, single-word utterances) all four tones were produced but the confusion between the rising tone and the dipping tone remained. At stage III (two and three-word utterances), they started to acquire the tone sandhi rules but the confusion between the dipping tone and the rising tone persisted. At stage IV (longer sentences) the rising and dipping tones were produced distinctly. However, the number of children in each age group, the ages of individual participants, or the age at which the children mastered the production of tones was not specified in the study.

The largest study to date included 129 Mandarin-speaking children (aged 1;6-4;6) who named and described pictures (Zhu & Dodd, 2000). The children's tone productions were transcribed by a Mandarin-speaking phonetician. The authors reported that the children in the youngest age group (aged 1;6-2;0) had mastered the production of tones in a variety of linguistic contexts.

The age at which children fully master the production of the four tones and the order of acquisition of the tones remain undetermined. Some studies have reported that tone production was mastered before the age of two (e.g., Zhu, 2002; Zhu & Dodd, 2000), whereas others suggested that tone production was not mastered by age three (e.g., Clumeck, 1977). Most studies found that level and falling tones were acquired before rising and dipping tones (e.g., Li & Thompson, 1977; Zhu, 2002) and that most of the tone errors involved a lack of distinction between rising and dipping tones (e.g.,

Clumeck, 1977; Li & Thompson, 1977). However, other studies reported that the rising tone was acquired first (e.g., Clumeck, 1977), and that most tone errors involved substitutions of the level tone for other tones (e.g., Zhu, 2002).

Several factors may have contributed to the discrepancies in the findings of these studies. The criteria for determining mastery of tone production were unspecified in most of these studies (Chao, 1973; Li & Thompson, 1977). Although some investigators provided information about error rates (Clumeck, 1977), only Zhu (2002) set criteria for the emergence and stabilization of tone production. A tone was considered to have emerged when a child produced the tone at least one time in his spontaneous or imitated speech. Stabilization was the age at which the child produced the tone with an accuracy rate of 66.7% and maintained a 66.7 % or higher accuracy rate in subsequent language samples. It is not surprising that the studies yielded conflicting results.

Procedural differences in the identification of children's tone productions may have also affected the findings. In most studies, children's accuracy in the production of tones was determined by one judge or transcriber (e.g., Chao, 1973; Clumeck 1977) and no study has reported inter- or intra-judge reliability on children's production of tones. Although inter-transcriber reliability was reported for other aspects of the children's phonological development such as vowels and consonants in Zhu and Dodd (2000) and Zhu (2002), no inter- or intra-transcriber reliability was reported for the transcription of tones. Without such measures, the accuracy of judges' transcriptions remains unknown. Also, most studies examined children's tone productions in spontaneous language samples (e.g., Chao, 1973; Clumeck, 1977; Zhu, 2002). Given young children's limited vocabularies and limited phonologies, the production contexts may be limited and poor

intelligibility may lead to an inaccurate gloss. Most studies determined the children's target tones with the support of the lexical, semantic, syntactic and contextual cues (e.g., Chao, 1973; Clumeck, 1977), which created expectations for the target tone and may have influenced transcription. The effect of these expectations may have been heightened in studies that used picture naming or picture description tasks (e.g., Li & Thompson, 1977; Zhu & Dodd, 2000). Although the children's targets were clear in these tasks, the judges' knowledge of the target words could have biased their tone transcription (Oller & Eilers, 1975). In some cases, parents interpreted the children's productions for the transcriber (e.g., Zhu, 2002); it is unclear how accurate the adults were in determining the children's target forms.

The very small sample size and the heterogeneity of the children may also have contributed to the divergent findings (e.g., Chao, 1973; Clumeck, 1977; Zhu, 2002). Only two published studies involved more than ten participants (i.e., Li & Thompson, 1977; Zhu & Dodd, 2000). The participant M (Clumeck, 1977) was exposed to two Chinese tone dialects at home and did not start using real words until after the age of 1;10. It seems likely that this child's development of tone differs from that of other monolingual children who were not late talkers. Developmental milestones and the language status of the children were not reported in most of the other studies. Therefore, it is unclear whether all the children had a comparable developmental and linguistic background.

Despite these divergent findings, most studies consistently found that the tone system is acquired relatively early and that mastery of tone production precedes mastery of segmental production (e.g., Clumeck, 1977; Li & Thompson, 1977). Zhu and Dodd (2000) found in their cross-sectional study that children between 1;6 and 2;0 made almost

no tone production errors, despite their finding that the error rate was 20% for vowels and 40% for consonants. In the longitudinal study, Zhu (2002) reported that when the four Mandarin tones were stabilized in the four children at 1;6 to 1;9, the children on average had only three stabilized consonants (Range = 1 to 5) in their repertoire.

Only one study also examined children's perception of Mandarin tones (Clumeck, 1977). Three objects were presented to the two children on each day of testing. The labels for two of the three presented-items formed a minimal pair different only in tone. The children were asked to identify the object when its label was presented. The two children were unable to discriminate between the rising and dipping tones at 3; 4 and 2; 9, respectively. However, the results may not be representative of a larger group of children due to the small sample size of the study. Also, as the author has pointed out, familiarity may also have influenced the children's performance given that the stimuli used in the study involved both nonsense words and real words. Thus, it is still unclear at what age Mandarin-speaking children master perception of the four tones.

The present study investigated both the perception and the production of Mandarin lexical tones in a sample of three-year-old children who were learning Mandarin as their first language in the U.S. Special efforts were made to ensure the homogeneity of the group, to reduce the degree of judge bias, and to compare children's productions with adults' productions. The objectives of the study were (1) to investigate the development of tone perception in 3-year-old children, (2) to examine the development of tone production in 3-year-old children, (3) to investigate the relation between the perception and production of lexical tones and (4) to go beyond the single transcriber who knows the identity of the semantic target.

Method

Tone Production

Participants

Children. Thirteen Mandarin-speaking children (6 girls and 7 boys) with a mean age of 3;0 (Range = 2;10 – 3;4) participated in the study. All of them were recruited in the New York and New Jersey areas. To determine the children's eligibility for the study, all children were administered an evaluation protocol which included a parent questionnaire, a Chinese-language test, an English-language test, a hearing screening and a language sample in Mandarin. According to the parent questionnaire, each child was brought up in a Mandarin-speaking home. All family members spoke only Mandarin to the child and the child had limited exposure to other languages, including English, or other tone dialects. Motor, social, emotional, cognitive, and language developmental milestones were all reported to be within normal limits.

In order to ensure that the children were monolingual Mandarin speakers with little proficiency in English, the Preschool Language Scale-3 (PLS-3) (Zimmerman, Steiner, & Pond, 1992) was administered to determine the children's English proficiency. All children scored more than one standard deviation below the mean in the total language score with a mean percentile rank of 2%, indicating very limited English proficiency.

The children's proficiency in Mandarin was examined with a language sample during play; story retelling; and the administration of a Chinese speech and language test,-- Language Disorder Scale of Preschoolers (LDSP, 學前兒童語言障礙評量表) (Lin & Lin, 1994), normed in Taiwan. The language samples revealed no language anomalies

in the children. Despite the fact that some test items in the LDSP were culturally biased (e.g., the pictures of some of the objects such as the piggy bank, the bar of soap, the construction site are different from those found in the American culture) all children scored higher than one standard deviation below the mean. The mean percentile rank was 49%.

All children passed the hearing screening at 1 KHz, 2 KHz, and 4 KHz at 20 dB HL under headphones using play audiometry (American Speech Language Hearing Association, 1997). One child (C12) passed the screening at 20 dB HL for all frequencies except for1K Hz in the left ear, which was passed at 25 dB.

Adults. Four of the 13 mothers were also recruited to participate in the study. All of them were native speakers of Mandarin who had resided in the U.S. for 2–11 years. Three mothers came from Mainland China and one was from Taiwan.

Stimuli

To minimize phonological effects on tone production, only monosyllabic words were tested in the study. The target words were drawn from a set of 72 monosyllabic words that are found in children's vocabulary and were represented in 78 pictures. Some words were depicted in more than one picture to check which representation would elicit the target form more successfully. The full set of pictures was presented to four Mandarin-speaking adults and 15 two- to three-year-old Mandarin-speaking children, including two children in Beijing. The pictures that elicited the most target forms from the 15 children were chosen for the final stimulus set.

The stimulus set consisted of 24 colored line drawings representing 24 monosyllabic words. Twelve of the words formed six minimal pairs across the six

different combinations of the four tones. The other 12 were singletons (words without minimal pairs in the set) balanced across the four tones (See Appendix A). The set of pictures was duplicated to form two sets of pictures. Each set was randomly ordered to form a different order of presentation. Each participant was randomly assigned to one of the two orders.

Procedures

Each child attended two one-hour sessions. Testing was conducted in a quiet room in the child's home, in a day care center or in a clinic. In the first session, the experimenter conducted the Chinese language test, tested the production and perception of tones and collected a language sample. In addition, the child's parents filled out the language questionnaire. The second session was used for hearing screening and the administration of the English test.

The tone production test was administered before the tone perception test. This was to prevent delayed imitations, and unequal or extra exposure to the target forms. The children were randomly assigned to one of two different orders of presentation. The 24 pictures were introduced one by one. Children's spontaneous productions were elicited by simple questions such as "這是什么 [what is this] ?" or "他在干嗎 [what is he doing] ? ". Numbers were elicited by asking the child to count the number of the same items in the picture. The child was asked to label each picture twice in a row. If the target form was not elicited, semantic cues were provided. If the child failed to produce the target form with cues and prompts, imitation of the target word was elicited. The children's productions were recorded on a DAT tape using a DAT recorder (Panasonic SV-255) via a condenser microphone.

The four mothers were asked to name each of the 24 pictures twice for their children at the end of the second session. Their productions were recorded on the same equipment.

Tones Production Judgments

Ten Mandarin-speaking adults were recruited as judges. Two sets of stimuli were prepared for the judges: one set of filtered stimuli for the experimental blocks and one set of natural (unfiltered) stimuli for the training blocks.

Stimuli

Filtered stimuli. The filtered stimuli in the experimental blocks for judgment were the target words produced by the 13 children and the four mothers described above. Only monosyllabic words spontaneously produced by the children either in isolation or in sentence-final position in the picture-naming task were included. Imitations, multisyllabic words, and noisy tokens were excluded. If a child produced the target form twice, the first production was used. The second production was included only when the first production was unsuitable due to noise or insufficient loudness. Thus, the set of stimuli for the judges' experimental blocks included 198 child-produced tokens of monosyllabic words. The target forms of 55 of them were in level tones, 54 were rising tones, 47 were dipping tones, and 42 were falling tones. Each of the 13 children contributed 12-19 tokens (Mean = 15.23).

The stimuli for the experimental blocks also included 92 tokens of adult-produced monosyllabic words. One word "mao4 [hat]" was excluded because none of the 13 children produced the word in the monosyllabic form. As a result, there were four subsets of 23 monosyllabic words, one from each of the four mothers. The adult-

produced tokens included 24 level tones, 24 rising tones, 24 dipping tones, and 20 falling tones.

The 198 child-produced tokens and the 92 adult-produced tokens were digitized as individual files using MultiSpeech 3700 (Kay Elemetrics). The word boundaries were determined primarily by the changes in the waveform amplitude with reference to the spectrographic display. Onset of the syllable was defined as the first visible increase of amplitude from zero in the waveform display corresponding to the beginning of the consonant in the spectrogram. The offset of the syllable was determined by the last excursion of the amplitude waveform from the baseline corresponding to the end of the formants in the spectrographic display. The maximum value of the fundamental frequency was measured for each of the 290 stimuli based on the pitch contours extracted by Praat Version.4.1.6 (Boersma & Weenink, 1992) with reference to the narrowband spectrogram display. Clicks and pops in four of the stimuli were reduced.

To eliminate the segmental information in the stimuli, all tokens were low-pass filtered using the Butterworth Low Pass filter in Cool Edit 2000 (Syntrillium Software). Adults' productions were low-pass filtered at 400 Hz because previous studies have shown that this filter cut-off was sufficient to eliminate most of the distinctive phonetic information while leaving the prosodic information intact (e.g., Cooper & Aslin, 1994; Friederici & Wessels, 1993; Jusczyk, Cutler, & Redanz, 1993). Because children tend to have higher fundamental frequencies, children's productions were low-pass filtered at a higher frequency, 500 Hz. All the stimuli were normalized for amplitude.

To check the quality of the filtered stimuli, all filtered tokens were presented to two native speakers of Mandarin. Neither Mandarin speaker was able to identify the

vowels or consonants in the stimuli with confidence. The tones of the five stimuli that were produced with the highest fundamental frequencies by the adults and the children were identified with 100% accuracy by the two Mandarin speakers, indicating that the filter cut-off was high enough to retain the tonal information in all tokens.

The 290 filtered stimuli were blocked by speaker. As a result, there were four blocks of stimuli produced by the four mothers and 13 blocks produced by the 13 children. Each block of adult-produced stimuli contained 23 tokens. The number of tokens in each child-produced block ranged from 12-19, depending on the number of monosyllabic words produced by the child during the picture-naming task. The 17 blocks of filtered stimuli were assigned to one of two sets. The assignment was pseudo random so that there was a balance between long and short blocks and the total number of trials was roughly equivalent. Experimental Set A (EA) contained two blocks of adult-produced stimuli and six blocks of child-produced stimuli. Experimental Set B (EB) contained the other two blocks of adult-produced stimuli and the other seven blocks of child-produced stimuli. There were a total of 139 stimuli in Experimental Set A and 151 stimuli in Experimental Set B.

Natural (unfiltered) stimuli. To familiarize the judges with the procedures and to ensure that all judges had the metalinguistic skill to label the lexical tones, 48 natural (unfiltered) stimuli were prepared for the training blocks. Half of the stimuli were monosyllabic morphemes in Mandarin and half of them were nonsense syllables with legitimate syllable structure and phoneme combinations. The nonsense syllables were included to determine whether the judges were able to identify the tones without relying on lexical or semantic information. The 48 monosyllabic syllables were spoken by a

female speaker and digitized. The stimuli were presented to a native speaker who identified the tones with 100% accuracy.

The 48 natural stimuli were randomly assigned to two training sets: Training Set A (TA) and Training Set B (TB). Each set contained 12 morphemes and 12 nonsense syllables balanced across the 4 tones. Thus, each set contained three morphemes and three nonsense syllables for each tone (See Appendix B).

Judges

Twelve adult speakers of Mandarin were recruited to make judgments on the tones produced by the 13 children and four mothers. All the judges were graduate students at the Graduate Center of the City University of New York. Two adults were excluded because one of them (J05) spoke five different tone dialects and the other (J10) did not meet the inclusion criterion of completing the training blocks with at least 80% accuracy. The remaining 10 adults (6F, 4M) with a mean age of 27.4 years (Range = 22-33 years old) passed a hearing screening at 500 Hz, 1K Hz, 2K Hz, and 4K Hz at 20 dB HL under headphones in a sound treated booth. A questionnaire was conducted to examine the language background of the judges. All judges reported learning Mandarin before the age of three years. Mandarin was reported to be the home and dominant language for all of them. Five judges came to the U.S. from Mainland China and five from Taiwan. The mean years of residence in the U.S. was 2.7 years (Range = 0.5 – 9 years). No history of hearing, speech or language difficulties was reported. None of the judges had taken any phonetics classes.

Procedures

The judges attended two 40-minute individual sessions within two weeks. In the first session, they were asked to fill out a questionnaire. Then, one of the two sets of training stimuli was selected randomly for presentation (e.g., TB). After the training block, they proceeded to complete judgments of one of the two experimental sets (e.g., EA). In the second session, they were presented the other set of the training stimuli (e.g., TA) followed by the other set of the experimental stimuli (e.g., EB). Hearing screening was conducted at the end of the second session.

During tone judgment, all the stimuli were presented at a comfortable hearing level via computer in a sound treated booth under headphones using customized identification testing software on a PC. The procedures were first verbally explained to the judges and again in writing on the computer screen. When the judge was ready, she used the mouse to click 'start' and the software randomly selected a block and presented the trials one at a time in random order. The judge could replay each stimulus an unlimited number of times. The judgment was indicated by a mouse click on the name of the tone category (i.e., Tone 1, Tone 2, Tone 3, or Tone 4) displayed on the screen. After the choice was made, the next trial was presented. After all the trials in a block were presented, there was a pause. The judge then clicked on a button to proceed to the next block when she was ready. The judges were allowed to take breaks at any time they desired. Responses of the judges were automatically recorded on a spreadsheet.

Only judges who achieved an accuracy level of at least 80% on the training blocks were included in the study. The 10 judges who were included in the study reached a mean accuracy of 97% (Range = 91.67% - 100%) in the training blocks across the two sessions.

Tone Perception

Participants

The participants for tone perception testing were the same 13 children who participated in tone production testing.

Stimuli

The stimuli were the same 24 monosyllabic words (6 minimal pairs, 12 singletons) used in production testing (See Appendix A). There were 36 trials in the perception testing. Twenty-four of the trials were designed to test children's perception of each of the 24 words without the minimal pair counterpart among the foils. The other 12 trials were designed to re-test the 6 minimal pairs with the minimal pair counterpart among the foils.

In each trial, the child was presented with four pictures on a sheet of paper. For the trials that contained a minimal pair, one picture in the foil had the same tone with the target and the other had the same tone as the other member of the minimal pair. Trials that did not contain a minimal pair had at least one foil that matched the target tone. The location of the target pictures was balanced across trials. Two sets of picture stimuli were organized into two random orders.

Procedures

Perception testing was conducted after production testing in the same quiet room. The children were randomly assigned to one of the two orders of presentation for perception testing. The examiner asked the child to point to the corresponding pictures by asking "哪個是 [which one is] … ?". After the guiding question had been asked in several trials, the experimenter sometimes simply presented the target word in isolation in

subsequent trials. Children's responses were recorded on a record sheet by the examiner. If the child did not respond after the question was presented two times, the examiner marked "no response" on the record sheet.

Results

Production of Tones

Interjudge Consistency

Tables 1 and 2 show the disagreement of the individual judge's categorization with the intended tones produced by the children and the adults. The results show that the judges were very consistent in their categorization of the adults' productions (See Table 1). The standard deviations for the level tone, the rising tone, the dipping tone and the falling tone were 4.39%, 5.36%, 4.39% and 3.5%, respectively. There was relatively greater variability in the judges' categorization of children's tone productions, particularly the children's falling tone (See Table 2). The standard deviation was 5.45% for the level tone, 8.12% for the rising tone, 7.24% for the dipping tone and 16.66% for the falling tone. Due to the fact that none of the judges consistently produced errors greater than two standard deviations above the means in their categorizations of each of the four tones produced by the children and adults and that when the four tones were collapsed, all the judges were highly consistent in their judgment of the children's and adults' productions (SD = 3.29% and 1.78%, respectively), subsequent analysis collapsed the results across all judges.

Production Accuracy

Table 3 shows the agreement of the judges' categorization with the intended tones produced by the children and the adults collapsed across all judges' responses. The

results revealed that the judges' perception of the adults' tone productions yielded high agreement with the speakers' intentions. The level tone, rising tone, and falling tone were categorized with over 95% accuracy. The dipping tone appeared to be relatively more difficult for the judges as the accuracy rate was 83.33%. Due to the low error rate in the judgment of the adults' productions, no statistical analysis was conducted.

The judges had more difficulty classifying the children's tone productions as the intended tones (See Table 3). The accuracy rate for categorizing the children's productions ranged from 46.42% to 76.92% for the four tones. Following the pattern found in the identification of adult productions, the dipping tone appeared to be most difficult for the judges. Judges' accuracy in categorization of children's production of level, rising, and falling tones ranged from 72.31% to 76.92%; however, the dipping tone was only classified as the intended tone 46.42% of the time (See Table 3). A one-way ANOVA for repeated measures yielded a significant overall effect of tones ($F(3,48) = 6.637$, $p = 0.001$). Post hoc pair-wise analysis (Tukey, HSD) revealed that the judges' identification of the children's dipping tones was significantly poorer than their identification of any of the three other tones (level tone vs. dipping tone: $p = 0.002$; rising tone vs. dipping tone: $p = 0.011$; falling tone vs. dipping tone: $p = 0.002$).

Chi-square tests were used to compare the judges' performance on the children's productions and their performance on the adults' productions. The results revealed that the judges' identification of the children's productions was significantly poorer than their identification of the adults' productions for each of the tones and when all the four tones were collapsed ($p \leq 0.0000$ in all cases).

Error Patterns

Table 4 presents the confusion matrixes for adult and child productions. Correct identifications by the judges are shown on the diagonal in boldface. There were similarities in the judges' error patterns for the adults' and children's productions. In both cases, when errors occurred, the judges were more likely to perceive the level tone and the dipping tone as a rising tone and the rising tone and falling tone as a dipping tone. However, the overall error rate was much higher for the children's productions than the adults' productions. Despite these similarities, the judges' errors in the identification of children's tone productions were more diverse than for the adults' productions. For example, the adults' level and rising tones were never confused with the falling tone, and the dipping tone was never identified as the level tone or the falling tone. However, all these confusions were found in the confusion matrix for children's productions.

Individual Differences in Tone Production

Tables 5 and 6 show the percent of identification errors for the tones produced by individual adult and child speakers, respectively. For adult speakers, the judges appeared to have greatest difficulty with Speaker A11 (14% errors overall), with error rates of 15% and 37% for the level tone and the dipping tone, respectively (See Table 5). For two other adult speakers (i.e., A10 and A12), identification of dipping tones was ≤ 90% accurate. Note however, that overall, all four adult speakers yielded lower error rates than any of the children (right hand column).

When comparing the error rates of the four tones produced by each child, ten of the 13 children had the highest error rate when the judges were categorizing their dipping tone (See Table 6). This implies that the accuracy rate for the production of the dipping

tone was the lowest for most children and that the production of the dipping tone was the least adult-like for most children.

The results indicated some individual differences in the mastery of the level tone, the rising tone, and the falling tone. It appears that there was little consistency among the children who were good at producing any of these three tones. For example, children who were the best at producing the level tone (i.e., C01, C03, C08, C14) are different from the children who were the best at producing the falling tone (i.e., C04, C05, C07, C13). Similarly, the child who was good at producing the rising tone (i.e., C06) was the worst at producing the level tone. Moreover, a different child appeared to have the greatest difficulty with each of the three tones. For example, C06 had the lowest accuracy rate in the production of the level tone; C04 had the lowest accuracy rate in the production of the rising tone, whereas C14 had the lowest accuracy rate in the production of the falling tone. Furthermore, children showed diverse patterns in their mastery of the tones. For example, C06 was poor in the production of the level tone but good at the production of the rising tone; C01, on the contrary, performed poorly on the production of the rising tone but well in the level tone. C14 performed badly on the falling tone and well in the level tone; C13, on the other hand, performed relatively poorly in the level tone but well in the falling tone.

When the four tones were collapsed, the error rate for productions of C02 (53.68%) was slightly greater than two standard deviations above the mean (50.38%). Background information and formal and informal testing did not reveal any differences between this child and the other children tested.

Perception of Tones

Table 7 presents the mean number correct and standard deviations for the perception of tones in the picture identification task; as well as the number correct for each child. As indicated, all children perceived the tones with high accuracy. The mean percent correct was 91% (Range = 78%-100%, SD = 10%) for the level tone, 95% for the rising tone (Range = 78%-100%, SD = 9%), 89% for the dipping tone (Range = 78%-100%, SD = 6%) and 88% for the falling tone (Range = 67% - 100%, SD = 14%). When the four tones were collapsed (right hand corner), the mean percent correct was 91% (range = 83%-97%, SD = 4%). Numerical differences across subjects were not reliable due to the small number of judgments.

Discussion

The results indicate that by the age of three, Mandarin-speaking children have mastered the perception of the four Mandarin lexical tones in monosyllabic words produced in isolation or in sentence-final position. All the children made very few errors in the picture identification task. It is unclear whether tone perception in other contexts (e.g., in continuous speech, sentence medial position, multisyllabic words) is fully mastered at age three, as the present study was not designed to answer such questions. Also, perception was assessed using "live-voice" presentation of items. Thus, the tones may have been "hyper-articulated" even more than in typical child-directed speech. Additional research is needed to determine if children can perceive tones as accurately in adult-directed speech as in child-directed speech.

The production results indicate that 3-year-old Mandarin-speaking children living in the U.S. have not yet fully mastered the production of tones in monosyllabic words.

This result conflicts with the findings of the recent cross-sectional study (Zhu & Dodd, 2000) and longitudinal study (Zhu, 2002) conducted in Beijing, China. In the large scale cross-sectional study, only 2 tone errors were found in the productions of 21 children in the youngest age group (i.e., 1;6-2;0) despite the fact that each of the children was asked to produce 44 words with one to three syllables and to provide descriptions to four pictures (Zhu & Dodd, 2000). The longitudinal study that followed four Mandarin-speaking children from 0;10.15- 2;0.15 years of age in Beijing also showed that children produced tones accurately in various contexts before the age of two years. The four children reached an accuracy of 66.7% or above in their production of the four tones and the use of tone sandhi rules in spontaneous speech.

The discrepancies between the findings of the present study and those of the previous studies might be attributed to a number of methodological differences. In the present study, filtered syllables were presented to a group of judges to identify the lexical tones. As a result, judges in the present study were deprived of the lexical, semantic, syntactic and contextual information about the children's productions. They were thus forced to base their judgments solely on the fundamental frequency contours. The judges in the previous studies listened to unfiltered speech and thus had semantic, syntactic and/or contextual information about the children's productions. Transcribers' expectations can influence transcription (Oller & Eilers, 1975). Thus, the transcribers' knowledge of the word targets in previous studies may have led to apparently earlier ages of acquisition. The exclusion of children's imitations in the present study may have also contributed to the difference in the findings.

The differences in the language environments may also account for the difference in the findings. The children in the present study were raised in Mandarin-speaking families residing in the U.S. These children were learning Mandarin as L1 in an L2 (English) environment. The exposure to two languages and limitation of L1 input in the ambient environment (e.g., television, radio) may cause the monolingual children in this study to develop differently from children who learn Mandarin in a predominantly Mandarin-speaking environment such as Beijing. Future research which directly compares groups of children using identical materials and procedures is needed to determine if this is the reason for the apparently slower development of tone production in Mandarin-learning children in the U.S.

Unlike previous studies in which the level tone and the falling tone were reportedly acquired before the rising and dipping tones, the results of the present study suggest that only the dipping tone was significantly more difficult for the children; the rising tone was produced with comparable accuracy to that of the level tone and the falling tone.

Two accounts can be provided for children's difficulties in producing the dipping tone. From the physiological point of view, the production of the dipping tone appears to be motorically more complex than that of the other three tones. Lexical tones are manifested mainly in the change of fundamental frequencies and the primary mechanism for the change of the fundamental frequencies is the regulation of the vocal fold tension by laryngeal muscles (Halle, Niimi, Imaizumi, & Hirose, 1990). Thus, the production of the dipping tone is relatively more difficult because it involves lowering and raising the fundamental frequencies at the right place and at the right time.

From the perspective of language input, the dipping tone may vary the most in the children's input because of phonological variation in continuous speech. The two most common sandhi rules in the production of Mandarin tones both involve a change of the dipping tone. In citation form or in sentence-final position, the dipping tone is a full dipping tone that has both a falling part and a rising component. However, when the dipping tone precedes another dipping tone, it changes to a rising tone (i.e., **dipping** + dipping → **rising** + dipping) (Sandhi rule 1); and when the dipping tone precedes any other tones, it only retains the falling part of its contour (i.e., **dipping** + level or rising or falling → **low falling** + level or rising or falling) (Sandhi rule 2). Given the variation of the dipping tone in different contexts, the input of the dipping tone is neither as consistent nor as frequent as the other three tones. Thus, it is likely that children have greater difficulty mastering the dipping tone, and, therefore, the dipping tone is mastered last.

The results of the present study also suggest that when children did not produce the tones accurately, their errors, as reflected in the judgments, were quantitatively and qualitatively different from the adult's error patterns. Compared to the adults, the children not only made more errors in their tone productions, but their error patterns were also more diverse. Error patterns that were not found in adult errors were found in children's errors. The findings of the present study support the findings in previous studies that children tended to confuse the rising and the dipping tones in their tone productions (e.g., Clumeck, 1980; Li & Thompson, 1977). However, other substitution and confusion patterns, such as confusions of dipping and falling tones, also occurred. Acoustic analysis of the children's and adults' productions will be required to examine the quantitative differences in the children's productions.

The findings that children perceived the tones with high accuracy but had not yet mastered the production of the tones suggest that tone perception precedes tone production and that accurate perception does not guarantee accurate production. One possible reason for the poor production in tone with apparently accurate perception is that the underlying perceptual representation of tones has not yet stabilized in Mandarin-speaking children living in the U.S. by the age of three years. Another reason could be the maturation of the coordinative structures necessary to produce the desired tone patterns consistently. Previous studies have shown that the mandible, tongue and vocal tract are not fully grown until 14-15 years of age (Kent & Vorperian, 1995) and that the descent of the larynx, the latter third of the tongue, the sensorimotor regulation of the mandibular system (Smith, Weber, Newton, & Denny, 1991) and the ratio of the length of the membranous portion of the vocal folds (Hirano, Kurita, & Nakachima, 1983) do not fully mature until adulthood. Thus, the phonological representation that underlies tone production may be more advanced than its realization. Additional research is needed to examine whether perception in other tasks (e.g., categorization) and in other contexts is fully accurate at young ages.

In summary, the present study found that three-year-old Mandarin-speaking children learning Mandarin in the U.S. perceived the four Mandarin tones accurately in monosyllabic words in isolation and in sentence final position. They had not yet fully mastered the production of the four tones in monosyllabic words. The dipping tone was the most difficult for the children to produce. Children's tone errors were both quantitatively and qualitatively different from those of adults.

We believe that the approach taken in this investigation reveals a more accurate picture of tone acquisition than previously used transcription approaches. Further research is needed to examine the acoustic details and the general course of production acquisition over time and across contexts as well as its relation to language acquisition.

References

American Speech-Language-Hearing Association. (1997). Guidelines for audiologic screening. Rockville, MD: ASHA

Boersma, P., & Weenink, D. (1992). Praat Version 4.1.6 [Computer software]. Amsterdam, The Netherlands: Institute of Phonetic Sciences. http://www.praat.org.

Chao, Y. R. (1973). The Cantian Idiolect: An analysis of the Chinese spoken by a Twenty-eight-month-old child. In C.A. Ferguson & D.I. Slobin (Eds.), *Studies of child language development* (pp. 13-33). New York: Holt, Rinehart & Winston.

Clumeck, H. (1980). The acquisition of tone. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child Phonology: Vol. 1. Production*. (pp. 257-275). New York: Academic Press.

Clumeck, H. V. (1977). Studies in the acquisition of Mandarin phonology. Doctoral dissertation, University of California, Berkerley.

Cool Edit 2000. [Computer software]. (1999). Syntrillium Software Incorporation.

Cooper, R. P., & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development*, 65(6), 1663-1677.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception and Psychophysics*, 54(3), 287-295.

Halle, P.A., Niimi, S., Imaizumi, S. & Hirose, H. (1990). Modern standard Chinese for tones: Electromyographic and acoustic patterns revisited. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo*, 24, 41-58.

Hirano, M., Kurita, S., & Nakachima, T. (1983). Growth, development and aging of the human vocal folds. In D. Bless and J. H. Abbs (Eds.), *Vocal Fold Physiology - Contemporary Research and Clinical Issues*, 22-43. San Diego: College-Hill.

Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica,* 33, 353-367.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64(3), 675-687.

Kay Elemetrics Multispeech Model 3700 [computer software]. Pine Brook, New Jersey: Kay Elemetrics Corp.

Kent, R. D., & Vorperian, H. K. (1995). Anatomic development of the craniofacial-oral-laryngeal systems: A review. *Journal of Medical Speech-Language Pathology*, 3, 145-190.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language* , 4, 185-199.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.

Lin, B., & Lin, N. (1994). *Language Disorder Scale of Preschoolers (學前兒童語言障礙評量表).* Department of Special Education, National Taiwan Normal University: Taipei, Taiwan.

Oller, D. K., & Eilers, R. E. (1975). Phonetic expectation and transcription validity. *Phonetica*, 31, 288-304.

Shih, C. (1988). Tone and intonation. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83-107.

Smith, A., Weber, C. M., Newton, J., & Denny, M. (1991). Developmental and age-related changes in reflexes of the human jaw-closing system. *Electroencephalography and Clinical Neurophysiology*, 81, 118-128.

Yip, M. (2002). *Tone.* Cambridge, United Kingdom: Cambridge University Press.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1992). *PLS-3: Preschool Language Scale-3*. San Antonio, TX : The Psychological Corporation

Zhu, H. (2002). *Phonological development in specific contexts: Studies of Chinese-speaking children*. Clevedon, England: Multilingual Matters.

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27, 3-24.

Author Note

Puisan Wong, Richard G. Schwartz, and James J. Jenkins, The Graduate Center, City University of New York.

This research was supported in part by a PSC-CUNY grant (Tone Perception and Production in 2-Year-Old Mandarin-Speaking Children) to the second author and by grants from the National Institute on Deafness and Other Communication Disorders (NIDCD) to the second author (5R01DC003885) and Winifred Strange (5R01DC00323-17).

We are especially grateful to Dr. Winifred Strange for providing the laboratory facilities in preparing and conducting the judgment portion of the project and her valuable and detailed comments and suggestions on the various versions of this manuscript. We express heartfelt appreciation to Dr. Gisela Jia for her generous help in running a pilot study on two children in Beijing, China and her insightful comments during the preparation of this manuscript. We particularly would like to thank Dr. Kanae Nishi for providing responsive technical support during the process of statistical analysis. We express our gratitude to Mr. Woonlam Hui for preparing the picture stimuli for the project, Mr. Bruno Tagliaferri who wrote the software program used for the judgment of tones, Mr. Gary Chant for the equipment and technical support, and Dr. Hwei Bing Lin and Dr. Min-Deh Wei for allowing us to use their clinic for data collection. We also would like to express our sincere thanks to Dr. Martin Gitterman, and Dr. Loraine Obler for their constructive comments on various versions of this manuscript. We are also indebted to the schools, physicians, and individuals who helped locate potential

participants. Finally, we are especially grateful to the participants and the parents of the children who participated in the project.

Correspondence concerning this article should be addressed to Puisan Wong, who is now at the Ph.D. Program in Speech and Hearing Sciences, the Graduate Center, The City University of New York, 365 Fifth Avenue, New York, NY 10016-4309. Email: pwong@gc.cuny.edu.

Table 1

Percent Errors for Each Judge in Identifying the Adults' Tone Productions

	Percent Error (%)				
Tone	Level	Rising	Dipping	Falling	4 Tones Collapsed
			Adult Productions		
Judges					
J01	0.00	0.00	**16.67**	0.00	4.35
J02	0.00	0.00	**16.67**	0.00	4.35
J03	8.33	**16.67**	12.50	0.00	9.78
J04	4.17	4.17	**16.67**	5.00	7.61
J06	4.17	4.17	**16.67**	10.00	8.70
J07	0.00	0.00	**25.00**	0.00	6.52
J08	**12.50**	0.00	8.33	0.00	5.43
J09	0.00	4.17	**20.83**	5.00	7.61
J11	4.17	8.33	**16.67**	0.00	7.61
J12	8.33	0.00	**16.67**	0.00	6.52
Mean	4.17	3.75	16.67	2.00	6.85
SD	4.39	5.36	4.39	3.50	1.78

Note: Errors are mismatches between the judgment and the target tone. For each judge, the tone that yielded the highest error rate is in bold.

Table 2

Percent Errors for Each Judge in Identifying the Children's Tone Productions

	Percent Error (%)				
Tone	Level	Rising	Dipping	Falling	4 Tones Collapsed
			Child Productions		
Judges					
J01	16.36	31.48	**46.81**	30.95	30.81
J02	21.82	25.93	**61.70**	42.86	36.87
J03	27.27	25.93	**57.45**	30.95	34.85
J04	23.64	37.04	**59.57**	4.76	31.82
J06	23.64	33.33	**53.19**	26.19	33.84
J07	18.18	18.52	**61.70**	4.76	25.76
J08	23.64	46.30	**51.06**	7.14	32.83
J09	14.55	20.37	**61.70**	38.10	32.32
J11	32.73	27.78	42.55	**45.24**	36.36
J12	18.18	31.48	**63.83**	4.76	29.80
Mean	22.00	29.81	55.96	23.57	32.53
SD	5.45	8.12	7.24	16.66	3.29

Note: Errors are mismatches between the judgment and the target tone. For each judge, the tone that yielded the highest error rate is in bold.

Table 3

Judges' Accuracy in Categorizing Adults' and Children's Tone Productions

Tone	Level	Rising	Dipping	Falling
Accuracy (% Correct)				
Adult Productions	95.83%	96.25%	83.33%	98.00%
Child Productions	76.56%	72.31%	46.42%	76.92%

Note: Accuracy of judgments is the percent of match between the judged tone and the target.

Table 4

Judges' Responses to Adults' and Children's Tone Productions

Judges' Responses	Level	Rising	Dipping	Falling
		Adult Productions		
Target Tones				
Level	**95.83 %**	3.75 %	0.42 %	0.00 %
Rising	0.83 %	**96.25 %**	2.92 %	0.00 %
Dipping	0.00 %	16.67 %	**83.33 %**	0.00 %
Falling	0.00 %	0.00 %	2.00 %	**98.00 %**
		Child Productions		
Target Tones				
Level	**78.00 %**	10.91 %	1.64 %	9.45 %
Rising	4.81 %	**70.19 %**	19.44 %	5.56 %
Dipping	2.34 %	39.79 %	**44.04 %**	13.83 %
Falling	3.57 %	1.67 %	18.33 %	**76.43 %**

Note. Correct responses are in bold.

Table 5

Judges' Error Rate in the Identification of Tones Produced by Each Adult Speaker

	Percent Error (%)				
Tone	Level	Rising	Dipping	Falling	4 Tones Collapsed
Adult Speaker					
A06	0.00	**8.33**	0.00	6.00	3.48
A10	0.00	1.67	**10.00**	0.00	3.04
A11	15.00	3.33	**36.67**	0.00	14.35
A12	1.67	1.67	**20.00**	2.00	6.52
Mean	4.17	3.75	16.67	2.00	6.85
SD	6.29	2.73	13.54	2.45	5.23

Note. Errors are mismatches between the judgment and the target tone. For each speaker, the tone that resulted in the highest error rate is in bold.

Table 6

Judges' Error Rate in the Identification of Tones Produced by Each Child Speaker

	Percent Error (%)				
Tone	Level	Rising	Dipping	Falling	4 Tones Collapsed
Child Speaker					
C01	2.00	66.00	**67.50**	22.50	38.89
C02	34.00	45.00	**94.00**	36.67	53.68
C03	2.50	26.67	**45.00**	23.33	20.83
C04	10.00	**50.00**	33.33	0.00	26.47
C05	33.33	30.00	**52.50**	0.00	30.77
C06	**60.00**	0.00	10.00	50.00	33.33
C07	15.00	14.00	**47.50**	6.67	21.25
C08	2.00	14.00	**50.00**	30.00	22.22
C10	30.00	37.50	**65.00**	15.00	40.77
C11	30.00	12.50	**86.67**	13.33	33.33
C12	40.00	16.00	**70.00**	30.00	36.47
C13	43.33	25.00	**50.00**	2.50	31.43
C14	2.50	23.33	25.00	**70.00**	27.86
Mean	23.44	27.69	53.58	23.08	32.10
SD	19.04	18.03	23.43	20.67	9.14

Note. Errors are mismatches between the judgment and the target tone. For each speaker, the tone that resulted in the highest error rate is in bold.

Table 7

Number Correct (out of 9 possible for each tone) in the Perception of Tones by Each Child

	Number Correct				
Tone	Level	Rising	Dipping	Falling	4 Tones Collapsed
Child					
C01	8	7	7	8	30
C02	9	9	8	9	35
C03	7	9	8	8	32
C04	9	9	8	6	32
C05	9	7	8	8	32
C06	9	9	8	9	35
C07	7	9	8	7	31
C08	9	8	8	9	34
C10	9	9	7	9	34
C11	7	9	9	9	34
C12	7	9	8	9	33
C13	9	9	8	6	32
C14	8	8	9	6	31
Mean #	8.23	8.54	8.00	7.92	8.23
Mean %	91.45	94.87	88.89	88.03	90.81
SD	10.30	8.63	6.42	13.95	4.45

Note: For each child, the tone(s) that yielded the lowest accuracy rate is (are) in bold.

Appendix A

Target Words for Perception and Production Testing

	Non Minimal Pairs				Minimal Pairs			
	Word	Tone	Pinyin	Translation	Word	Tone	Pinyin	Translation
1	哭	1	ku 1	cry	汤	1	tang 1	soup
2	花	1	hua 2	flower	糖	2	tang 2	candy
3	灯	1	deng1	lamp	三	1	san 1	three
4	门	2	men 2	door	伞	3	san 3	umbrella
5	牛	2	niu 2	cow	书	1	shu 1	book
6	球	2	qiu 2	ball	树	4	shu 4	tree
7	马	3	ma 3	horse	鱼	2	yu 2	fish
8	手	3	shou 3	hand	雨	3	yu 3	rain
9	笔	3	bi 3	pen	毛	2	mao 2	hair
10	二	4	er 4	two	帽	4	mao 4	hat
11	四	4	si 4	four	脚	3	jiao 3	foot
12	抱	4	bao 4	hold	叫	4	jiao 4	shout

Note. Tone 1 = Level Tone, Tone 2 = Rising Tone, Tone 3 = Dipping Tone, Tone 4 = Falling Tone. Pinyin is the official romanization system of Chinese.

Appendix B

Stimuli in the Training Blocks for the Judges.

Tone	Level	Rising	Dipping	Falling
		Training Set A (TA)		
Nonsense Syllables				
1	nai 1	dang 2	lai 3	kuan 4
2	nou 1	kun 2	cuo 3	kei 4
3	kuai 1	gua 2	kuo 3	kuan 4
Morphemes				
1	xin 1 心	yao 2 遙	xu 3 許	yang 4 樣
2	mao 1 貓	tong 2 同	kuan 3 款	ci 4 次
3	jian 1 間	chen 2 陳	li 3 理	wei 4 味
		Training Set B (TB)		
Nonsense Syllables				
1	dei 1	kong 2	ce 3	niu 4
2	lian 1	ding 2	mie 3	jiong 4
3	gun 1	kan 2	que3	keng 4
Morphemes				
1	yue 1 約	mo 2 蘑	chang 3 場	hui 4 會
2	ti 1 剔	xie 2 鞋	guo 3 果	mai 4 麥
3	ke 1 科	liu 2 流	huang 3 恍	ku 4 庫

Note. The transcription is in Pinyin.