Journal of Child Language Acquisition and Development – JCLAD
Vol: 5 Issue: 3 165-182, 2017, September
ISSN: 2148-1997

Tone Acquisition at the One-Word Stage in Taiwan Mandarin: A corpus study

Received : 13.03.2017
Accepted : 08.09.2017
Published : 30.09.2017

I-Ping Wan[1]
National Chengchi University
Han-Chieh Yang[2]
Lo-Sheng Hospital

Abstract

Evidence for early developmental patterns of tone acquisition was found by investigating a longitudinal study of 8 children (0;6-1;8) in Taiwan Mandarin. The present study found that in terms of the order of the emergence of tones, there seems to be a universal tendency whereby high-level tones are likely to be acquired earlier than high-falling tones, followed by low-falling tones and rising tones. As for the distribution patterns of tones in monosyllabic words, high-falling tones are the most common, high-level tones are the next common, and low-falling tones are the least common. In disyllabic words, a language-specific phenomenon in Taiwan made the endearment tone, T3+T2, the most common pattern, with the high-level sequence, T1+T1, taking second place, followed by the morphologically-conditioned pattern, T2+T3. Binary foot structure has been supported; however, no rhythmic effect has been found between tones and disyllables. Although children acquired four tones early, they were not able to stabilize the system at the one-word stage. Articulatory theory, physiological effort, phonetic cues, Markedness constraints, and influence from the caretakers' speech and linguistic environment provide some explanations for the order of emergence of tone acquisition and the distribution patterns of tones.

Keywords tone acquisition, one-word stage, age and order of emergence of tones, age of stablizaton of tones, distribution patterns of tones, endearment tone, Taiwan Mandarin

1. Introduction

A number of studies have suggested that infants are sensitive to prosodic cues during their early developmental stage, and children start to acquire suprasegmental features such as stress, intonation, pitch and tone very early in the developmental process (e.g., Clumeck, 1980; Demuth, 1996; Kaplan & Kaplan, 1971; Mehler et al., 1988). Most researchers have pointed out that suprasegmental features, such as intonation and stress, are acquired earlier and better than segments (e.g., Crystal, 1986; Demuth, 1996; Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988). Tone is considered one of the most salient features in tone languages, and a number of studies have discussed tone acquisition in Cantonese, Mandarin,

[1] Bio: Dr. I-Ping Wan is a Professor in the Graduate Institute of Linguistics and Center for Mind, Brain and Learning at National Chengchi University in Taiwan. Her research interests span both phonology and psycholinguistics, and she has collected a number of spoken corpora in Mandarin including speech errors, aphasic speech, child language acquisition, and phonologically-disordered children's speech. Contact: ipwan@nccu.edu.tw

[2] Bio: Han-Chieh Yang received an M.A. in linguistics at National Chengchi University, and another M.A. in speech and language pathology at National Taipei University of Nursing and Health Sciences. She is currently affiliated with Lo-Sheng Hospital, Ministry of Health and Welfare, in Taiwan.

Taiwanese, and Thai (e.g., Mandarin spoken in the United States: Chao, 1951, 1968; Clumeck, 1977, 1980; Taiwan Mandarin: Chen & Kent, 2009; J. Hsu, 2003; Li & Thompson, 1977; Wong, 2008, 2012, 2013; Wong, Schwartz, & Jenkins, 2005; Beijing Mandarin: Zhu, 2002; Zhu & Dodd, 2000; Cantonese: So & Dodd, 1995; To, Cheung,& McLeod, 2013; Tse, 1991; Taiwanese: H. Hsu, 1989; Tsay, 2001; Thai: Tuaycharoen, 1977).

Taiwan Mandarin has four distinctive lexical tones and a neutral tone. The conventionally accepted inventory of tones involves T1, high-level tones [55], T2, rising tones [35], T3, low-falling tones [21], and T4, high-falling tones [51]. Note that these tone numbers are not intended to indicate underlying sequences, but simply to show the pitches involved in the contour tones. In addition, Taiwan Mandarin is analyzed as having a range of possible surface monosyllables: V, CV, GV, VG, VN, CVG, CVN, CGV, GVG, GVN, CGVG, CGVN. The maximal syllable is CGVX, with C a [+consonantal] segment, G a glide, V the nucleus vowel, and X either a nasal or a glide.

1.1. Previous studies

In Mandarin, the majority of previous research dealt with production, and few focused on perception (For an integrated view on the studies related to tone acquisition, see Singh and Fu (2016) for more evidence from perceptual and production perspectives). In perception, evidence from a habituation-based paradigm suggested that tone categories in Cantonese and Mandarin infants seem to emerge at the age of 4 months (Yeung, Chen, & Werker, 2013). In the production studies, at least one of the following three issues has usually been discussed: (1) the age and order of emergence; (2) the age and order of stabilization; (3) the distribution patterns of tones.

In the past, the majority of the relevant studies focused on the chronological order of tone development (i.e., the emergence of tones) and error patterns or error rates of tones. Three ways of data collections have been adopted, and the studies prior to the year 2000 were largely based on longitudinal data, which involve descriptive diary reports or case studies (e.g., Chao, 1951, 1968; Clumeck, 1977, 1980; J. Hsu, 2003; Li & Thompson, 1977). The second approach was to present experimental methods (Zhu, 2002; Zhu & Dodd, 2000) and the third one included acoustic analysis for children's developmental course of tone acquisition (Chen & Kent, 2009; Wong, 2008, 2012, 2013; Wong et al., 2005).

Most of the prior studies focusing on the age and order of emergence found that high-level tones are acquired earlier than high-falling tones and rising tones, which are acquired earlier than low-falling tones whereas Clumeck's study (1977) suggested a conflicting order in which rising tones occur earlier than the rest of three tones. Secondly, children have mastered their tonal systems with few errors by the age of 2;0, possibly at the one-word stage (e.g., Li & Thompson, 1977). However, a late tonal development was proposed by Clumeck (1977, 1980), suggesting that the children would not complete the acquisition of the tonal system at the two-word utterance stage or around the age of 2;4 to 3;0. Finally, very few studies discussed the issue involving the distribution patterns of tone acquisition, and they seemed to lump the tone inventory together with monosyllables, disyllables and multisyllables. Li and Thompson (1977) and J. Hsu (2003) found that

children are more likely to produce high-falling tones, followed by high-level tones, and rising tones and low-falling tones are the least frequent.

The new approach including elicited production experiments also confirmed some of the aforementioned studies. Zhu (2002) and Zhu and Dodd (2000) presented quantitative and qualitative measures by using 66.7% criterion in a cross-sectional experiment for providing a certain degree of phonological accuracy and consistency in defining when the tonal system is considered stabilized. They found that rising tones emerged about one month after high-level and high-falling tones had first emerged. Low-falling tones were the last to emerge at the age of 1;4, and children emerged and mastered the four tones at age of 2.

However, using acoustic evidence for children's developmental course of tone acquisition yielded difference analyses. Chen and Kent (2009) investigated the early prosodic development of children at underdeveloped ages (0;7-1;6). They documented the children acquiring Taiwan Mandarin tones along with caretakers' speech by systematically measuring F_0 variation at the transition from babbling to producing the first fifty words. They agreed with most of the prior studies and found that falling tones are more prominently produced than high-level and rising tones, high-level tones occur more frequently than mid/low tones, and the distribution patterns of F_0 contours in babbling and early-word stages are pretty much alike; the ones in infants' reflect their caretakers' data as well. On the contrary, Wong (2008, 2012, 2013) and Wong, Schwartz and Jenkins (2005) presented a phonetic evidence with a more rigorous experimental control in a low-pass filtered speech to eliminate lexical information and reserve tone information from Taiwan Mandarin speaking children. They suggested that children produce the most adult-like tones at the isolated monosyllabic words in the following order: high-level tones > high-falling tones > rising tones > low-falling tones. However, in their studies, children did not stabilize the four tones in monosyllabic words even when they reached the age of 3, and children did not stabilize the four-tone system in disyllabic words till they reached the age of 5 or 6.

Cross-linguistic studies from Thai, Cantonese, and Taiwanese generally agreed that level tones are in general acquired earlier than contour tones; however, it is uncertain whether high-level tones would have a better chance of being acquired earlier than mid- or low-level tones (Thai: Tuaycharoen, 1977; Cantonese: So & Dodd, 1995; Tse, 1991; Taiwanese: H. Hsu, 1989; Tsay, 2001). Tuaycharoen (1977) in Thai showed what is to date the earliest age of mastery of a tonal language system to be 1;4. So and Dodd (1995) and Tse (1991) both agreed that the Cantonese children generally control and master the production of lexical tones by the age of 2 at the two-word stage. Tsay (2001) in Taiwanese presented percentages of error rates as criteria for defining how tones are maintained and she also adopted the phonetic measurements based on the corrected tokens. The lower error rates were also in support of an early completion of tone acquisition for Taiwanese children, and the phonetic result showed that Taiwanese children acquire an entire phonetic gesture including the pitch dimension and duration of tonal contrasts among different tones found in adult speech.

In conclusion, except for Clumeck (1977), who suggested the early emergence of rising tones, the rest of prior studies in general agreed that high-level tones are acquired earlier than high-falling tones which emerge earlier than rising tones, and low-falling tones emerge in the last order. A few studies discussed the age and order of stabilization, and the results showed a discrepancy between the early and late developmental course of tone acquisition. Very few studies worked on the distribution patterns of tones, and by combining tones in monosyllables with tones in disyllables or multisyllables, children were more likely to produce high-falling tones, followed by high-level tones, and rising tones and low-falling tones were the least frequent.

1.2. Theoretical analyses of tones

A number of relevant studies provided articulatory explanations, markedness, phonetic cues, or theoretical phonological framework to account for universal patterns or language-specific phenomenon on tones. Evidence from the articulatory effort theory led Ohala and Ewan (1972) and Ohala (1978) to report that rising tones are cross-linguistically longer than falling tones, and need more energy with muscular control, so rising tones are supposedly more difficult than falling tones to produce. Vihman (1996) proposed that there is a degree of markedness to pitch, and suggested that a falling pitch movement is a natural gesture of speech production and required less physiological effort than rising pitch movement.

Many scholars in production and perceptual studies have discussed Mandarin tones in relation to the phonetic cues, including duration, intensity and F_0. In duration, Fu and Zeng (2000), Tseng (1990), and Xu, Tsai and Pfingst (2002) all found that low-falling tones have longer duration than rising tones, followed by high-level tones and/or high-falling tones. In intensity, Ong and Yang (1997), and Xu et al. (2002) agreed that high-falling tones are the strongest in intensity, and high-level tones are the next strongest, and the least strong are the low-falling tones. Tsao (2008) investigated tone discrimination in Taiwan Mandarin infants between 10 and 12 months of age by testing their abilities to distinguish tone contrasts. The result showed that high-level tones and low-falling tones are a highly distinct tone pair, followed by a moderately distinct tone pair of rising and high-falling tones, and rising and low-falling tones are minimally distinct from one another.

Regarding theoretical implications, Yip (2002) presented a theory of the markedness of tone in which contour tones are more marked than level tones, rising tones are more marked than falling tones, and high-level tones are more marked than low-level tones. Such a theory would predict that high-level tones are acquired earlier than other contour tones, falling tones are acquired earlier than rising tones, low-level tones are acquired earlier than high-level tones.

1.3. Spoken corpus

As with the occurrence of frequency, high-falling tones had greater frequency than rising tones in the world's languages (Li & Thomspon, 1978). Token counts from a speech corpus (N=604,916), *NCCU Corpus of Spoken Chinese*,

showed a rank order in which high-falling tones had the highest frequency of occurrence (39%), followed by low-falling tones (23%), then high-level tones (21%), and finally rising tones, being the least common (17%)[3]. If each tone is expected to occur with approximately equal frequency, the frequency of the actual tokens is greater than the 25% expected by chance for high-falling and high-level tones (i.e., there are four tones in Taiwan Mandarin).

1.4. Binary foot structure

Citation tones in Mandarin are nearly always discussed in isolated words, and only when tone sandhi processes are involved will more than two syllables in sequences be discussed. Lin (2007) listed the four basic tones in monosyllables and 15 tonal combinations in disyllables, showing all the possible distribution patterns of tones in Mandarin. Note that the pattern T3+T3 is not allowed in the surface form[4].

Recently, the preference for disyllables in language acquisition has shed light on binary foot structure in cross-linguistic studies (e.g., Demuth, 2006; Demuth & Johnson, 2003; Demuth & Tremblay, 2008; Miyokada, 2012; Ota, 1999), suggesting a universal tendency. Many phonologists treated binary foot structure as a higher constraint than faithfulness or markedness in optimality theory. As for prosodic phonology, there was a rhythmic effect in the relative metrical prominence between tones and disyllables. H. Hsu (2006) and Lu (2011) found the preference of low-high or weak-strong (i.e., iambic) contrast in pitch from dialectal studies and experiments, respectively, in Mandarin.

1.5. Research issues

The most controversial issue in the aforementioned studies seemed to be children's stabilization of tones at the developmental stage of two years of age and beyond, and few studies were related to the developmental course at prelinguistic stages. In addition, many of the studies were collected from diary reports and they might include meaningful and non-meaningful utterances in the children's spontaneous speech and imitative responses as Li and Thompson (1977) included non-meaningful tone tokens, and Zhu (2002) involved imitative forms in the studies. Moreover, very few studies have discussed the distribution patterns of tones by sorting out disyllables from monosyllables, and the studies in Wong (2008, 2012, 2013) and Wong et al. (2005) by far were probably the first series of work presenting evidence on tones from monosyllabic and disyllabic words. Finally, dialectal differences might offer different results; studies such as Zhu (2002) or Zhu and Dodd (2000) had the speech samples in Beijing Mandarin, and for the contour pattern on the third tone, there was a pitch falling downward to the

[3] The corpus was collected and provided by Prof. Kawai Chui at National Chengchi University. The data contained four tones plus neutral tone in spontaneous speech. Since the present study focused only on the four tones, the token and type for the neutral tone were excluded for the discussion.

[4] Such a pattern is not a legal phonological output in phonetic representation due to the tone sandhi rule where the low-falling tone in the first syllable would change to a rising tone when the two syllables in sequence both have underlying low-falling tones.

low pitch and rising upward at the end point in Beijing Mandarin, but there was a rather simple low pitch falling downward in Taiwan Mandarin. Another dialectal difference was the special phenomenon addressed by Duanmu (2007), the endearment tone T3+T2 specially used in the reduplicative words (low-falling tones plus rising tones). Such a pattern has been found to be very popular in Taiwan Mandarin. However, no report on tone acquisition has ever been done in relation to such a special phenomenon.

Due to some inconclusive and inconsistent findings in cross-linguistic studies, the present study aims to investigate the developmental patterns of tone acquisition at underdeveloped stages in Taiwan Mandarin by investigating digitally recorded spontaneous vocalizations of 8 children, aged 0;6 to 0;7 in the beginning and 1;6-1;8 by the end of one-word stage. The data to be included involved meaningful monosyllabic and disyllabic words, and were collected and dated before the children started to enter the two-word stage[5]. Multisyllabic words were insufficient for computing results, so they were not compared quantitatively for the present study. Topics involved in the present study include the following: (1) the age and order of emergence; (2) the age and order of stabilization; (3) the distribution patterns of tones in monosyllables and disyllables detected by age-tracking children's speech samples at one-word stage in Taiwan Mandarin-speaking children.

2. Methodology

2.1. Participants

Eight children (3 boys, 5 girls) were assessed longitudinally in their earlier development in this study. Seven Taiwan Mandarin children began to enter the two-word stage at 1;6 and one child began to produce two words at 1;8. To gain information concerning the acquisition of tone, eight paid children (3 boys, 5 girls), who acquired Mandarin only as their first language, were assessed longitudinally between the ages of 0;7 and 1;8[6]. At the beginning of the data collection, the children's ages were between 0;6 and 0;8 (mean age= 0;7.17, SD= 0.4 months). At the end of the one-word stage, the children's ages were between 1;4 and 1;8 (mean age= 1;5.53, SD= 0.09 months). The eight participants were all healthy and had not been diagnosed with any hearing or intellectual impairment, and all uttered their speech in Mandarin only with their main caretakers.

2.2. Data collection and processing

The entire data collection session followed the criteria proposed by Vihman (1996), Vihman and McCune (1994), and Zhu (2002). Before the data collection session began, the mothers were trained to get detailed guidelines and procedures as to what they were expected to assist the children to produce more speech output, and they did not help collect the data. Data collection took place every other week and was recessed for three weeks

[5] One of the children did not enter the two-word stage till 1;8, whereas the rest of the children in this study completed the one-word system before 1;6.

[6] The other 10 children's data were not included in the present study since they acquired Mandarin (70%) and other Chinese languages (e.g., Taiwanese or Hakka) (30%) as the first languages.

during Lunar New Year.

All children were video-recorded and digital-recorded at their homes during bi-weekly 60-minute sessions. Audio recordings were made with the AKG cardinal microphone and SONY digital voice recorder (ICD-UX513F) along with SONY Handycam digital video camera recorder (SONY DCR-SR40). The video files helped the authors and the research team members easily decode the spontaneous utterances by the children's gestures and eye movements, and the sound files provided high quality audio signals.

The usual methodology for collecting tone data is to rely on the native-speaker linguist's intuitions as to the categorization of the tone in the spoken utterance with lexical information. This methodology is subject to some problems of listener/transcriber bias, as noted in Wong's (2013) study. However, one could still argue that the perception that tones carried with lexical information might be more valid psycholinguistic measure than the acoustic properties of tones. As stated in Demuth (2006), the challenge with the field of phonological acquisition has been the lack of longitudinal phonetically transcribed data from multiple children between the ages of 1 and 2. Besides, it is difficult to perform elicited production experiments with children at underdeveloped stages below the age of 2. In the first author's other corpus based on a picture-naming task in a cross-sectional study from 225 Taiwan Mandarin children (0;7-6;00), the author has adopted a program for automatically aligning the transcript of acoustic signal to proper sound files from the forced alignment system, *HTKAlignementScript_ILAS*.[7] Such a program helped distinguish the meaningful words from non-meaningful ones, and the ratio in the elicited production experiment at underdeveloped stages (Mononsyllables: Meaningful words vs. Non-meaningful words = 1.2 vs. 1; Disyllables: Meaningful words vs. Non-meaningful words: 2.1 vs. 1) was similar to the one found in the spontaneous speech. Out of 729 tokens in monosyllabic words, and only 379 words that were selected are meaningful, and out of 957 tokens in disyllabic words, 660 words show the meaningful form, suggesting a rather strict screening process for this study; assuming that the data to be discussed below are sufficiently reliable to support the analysis.

2.3. *Data analysis*

Meaningful words were classified under the following three conditions: 1) when children's vocalizations would be identified as words that are matched more than two segments of the adult form; 2) when children's tone matched the adult target; 3) when children produced the words more than twice with similar phonological shapes across different uses. For example, in [twej51] 'correct', [tɕʰjow35] 'ball', and [fan51] 'rice', many children would produce them as in [tej51], [tʰjow35], and [fa51]. Since more than two segments

[7] Professor Chiu-yu Tseng and her research team in the Phonetics Lab at Academia Sinica have been developing the Mandarin forced alignment system for many decades. The program is a Hidden Markov Model Speech Recognition Toolkit along with Praat tools (Boersma, 2001). This forced alignment package can be used for elicited production, but does not provide good alignment results for spontaneous speech.

matched the adult form, examples like these that appear more than twice in different contexts were considered meaningful words. However, if the tones were not recognizable, they were classified as non-meaningful words. In the study, 379 meaningful monosyllabic words were selected from 729 tokens (Meaningful words vs. Non-meaningful words: 1.1 vs. 1), and 660 meaningful disyllabic words were selected from 957 tokens (Meaningful words vs. Non-meaningful words: 2.2 vs. 1)[8].

The recording was started before their pre-meaningful speech occurred, and a tone was considered to have emerged when a child could produce it more than twice in a meaningful word in spontaneous speech. The utterances of children along with the utterances of their main caretakers have all been transcribed in broad phonemic transcription. In this study, the data collection was drawn from the point when these children began to produce the first meaningful word. For IPA transcription, 10% of audio recordings of the database were phonetically transcribed by the second author; 90% by six trained research assistants, and 30% (311 syllables) of samples, which were selected from the database at random, were verified by the first author to have a 90% interrater reliability (Cronbach's alpha=.899). Intrarater reliability was calculated as the percent agreement between the coder's original transcription and his/her repeated transcription for 935 syllables (Cronbach's alpha=.948). When there appeared to be a questionable utterance such as a tone that was not clearly audible, or when there was a conflict between the two transcribers, the data were not included in the study. Transcription conventions were based on Chao's tone markers. An imitative response to a verbal stimulus has not been considered as a word, and such data were not considered in this study.

Following the same criterion for an accuracy rating formulated in the studies of Zhu (2002) and Zhu and Dodd (2000), a tone in the study was considered to have reached the level of stability at the point when the accuracy rating in the child's sample had reached 66.7% of phonological accuracy and consistency.

- Accuracy rating = the number of times a tone is produced correctly / the number of opportunities for the tone in the sample x 100%.

3. Results and Discussion

The following table summarizes the children's information, the age range during which the children contributed the data for the current study, the emergence of the children's first word and tone, and the total number of tokens for the one-word stage involving monosyllables and disyllables in the eight children.

[8] The current study is selected from a corpus which is by far the largest database involving Taiwan Mandarin-speaking children's spontaneous speech, approximately estimated over 300,000 tokens (N=18, 0;6-5;10). This corpus has been collected by the first author and the research team from 18 Taiwan-speaking children between 2011 and the present.

Journal of Child Language Acquisition and Development – JCLAD
Vol: 5 Issue: 3 165-182, 2017, September
ISSN: 2148-1997

Table 1
Participants

Child	Gender	Age range	Emerge T	Emerge W		N		%
KL	M	0;7-1;6	1;1	Monosyllable	1;2	Monosyllable	53	31%
				Disyllable	1;1	Disyllable	117	69%
DD	M	0;7-1;6	0;10	Monosyllable	0;10	Monosyllable	77	47%
				Disyllable	0;11	Disyllable	86	53%
SH	M	0;7-1;6	0;9	Monosyllable	0;11	Monosyllable	38	30%
				Disyllable	0;9	Disyllable	88	70%
JJ	F	0;8-1;6	1;0	Monosyllable	1;0	Monosyllable	80	47%
				Disyllable	1;1	Disyllable	91	53%
WW	F	0;8-1;6	0;10	Monosyllable	0;10	Monosyllable	25	25%
				Disyllable	1;2	Disyllable	74	75%
YC	F	0;7-1;6	1;0	Monosyllable	1;0	Monosyllable	46	35%
				Disyllable	1;1	Disyllable	85	65%
YJ	F	0;6-1;8	0;9	Monosyllable	0;9	Monosyllable	17	40%
				Disyllable	0;9	Disyllable	25	60%
LC	F	0;7-1;6	1;0	Monosyllable	1;1	Monosyllable	43	31%
				Disyllable	1;0	Disyllable	94	69%
				Total		Monosyllable	379	36%
						Disyllable	660	64%

Table 1 shows that eight children (3 boys, 5 girls) were assessed longitudinally in their earlier development in this study. Seven Taiwan Mandarin children began to enter the two-word stage at 1;6 and one child began to produce two words at 1;8. Each child had a different time point for the emergence of their first meaningful word, and Emerge T and Emerge W in the column referred to the age recorded the first time a child was able to produce a tone or a word, respectively. This table showed the age at which monosyllables and disyllables emerged in children's first words: four children produced monosyllables earlier than disyllables, three children showed the opposite order, and 1 child produced the two syllable types almost simultaneously. Generally speaking, disyllables (N=660, 64%) far outnumbered monosyllables (N=379, 36%) across children and within children. This result suggests that disyllables are more likely to emerge at children's first-word stage.

The fact that these children preferred to produce disyllables might have largely been influenced by the caretakers' speech, which had more disyllables (N=6,172, 78%) than monosyllables (N=1,729, 22%) as well as by linguistic environment, in which the speech corpus (N=604,916), *NCCU Corpus of Spoken Chinese,* showed Taiwan Mandarin speakers had a preference to produce more disyllables (N=368,997, 61%) than tri-syllables

(N=96,789, 16%), monosyllables (N=90,738, 15%), and multisyllables (N=48,392, 8%). The children's data might also confirm the universal preference in cross-linguistic studies for binary foot structure, which can be viewed as a higher constraint than faithfulness or markedness (e.g., Demuth, 2006; Demuth & Johnson, 2003; Demuth & Tremblay, 2008; Miyokada, 2012; Ota, 1999;).

3.1. Age and order of the emergence of tones

With regard to the age and order of the emergence of tones, Table 2 presents the age when each child uttered his or her first meaningful word with recognizable lexical tone.

Table 2
Age of the emergence of tones

	KL	**DD**	**SH**	**JJ**	**WW**	**YC**	**YJ**	**LC**
High-level tone	1;1*	0;10	0;9*	1;0*	0;10*	1;0*	1;0*	1;0*
Rising tone	1;3	0;10	0;11	1;0	0;11	1;0	1;2	1;0
Low-falling tone	1;2*	0;10	0;11	1;0	1;2*	1;1*	0;9	1;0
High-falling tone	1;1*	0;10*	0;11	1;0*	0;11*	1;2*	0;9*	1;1*

*Note: * = The child uttered a recognizable meaningful word for the first time.*

This table summarizes the age at which the four tones emerged in Taiwan Mandarin. High-level or high-falling tones were the earliest and existed in the children's data collected at the time when they were about to produce first recognizable meaningful words. This table shows that high-level tones were earliest in the sample of the children's speech; 4 out of 8 children acquired high-level tones, and 2 out of 8 children acquired high-falling tones. As for rising tones, all children had great confusion and variation in their production systems so some of the data were not recognizable. The following table presented the developmental course of each tone.

Table 3
Order of the emergence of tones

Participant \ Month	KL	DD	SH	JJ	WW	YC	YJ	LC
0;6	N/A	N/A	N/A	N/A	N/A	N/A	None	N/A
0;7	None	None	None	N/A	N/A	None	None	None
0;8	None	None	None	None	None	None	None	None
0;9	None	None	1	None	None	None	4	None
0;10	None	4	1, 2	None	1	None	4	None
0;11	None	1	1, 2, 3	None	4	None	3, 4	None
1;0	None	1, 4	1, 2, 3	1, 4	1, 4	1	1, 3, 4	1
1;1	1, 4	3	Complete	2, 3	1, 2, 4	1, 3	1, 3, 4	1, 4
1;2	1, 3, 4	Complete		Complete	1, 2, 4	1, 3, 4	1, 3, 4	1, 4
1;3	Complete				Complete	Complete	Complete	Complete

Note: The numbers 1-4 refer to high-level, rising, low-falling and high-falling tones, respectively. "N/A" means the data had not yet been collected at that age. "None" means that the child has not yet acquired any of the four tones in Mandarin while "Complete" means that the child had acquired, but not necessarily stabilized, all four tones of Mandarin at that age.

All these children in general showed a similar pattern of order of tone acquisition. Four children first acquired high-level tones, two children first acquired high-falling tones, and the other two children acquired both tones simultaneously. At the end of tonal completion, four children had acquired rising tones, one child had acquired low-falling tones, one child had acquired high-falling tone, and two children had acquired rising and low-falling tones at the same time. Although there were some individual differences, not all of the children first acquired high-level tones before high-falling tones; no data have been reported so far regarding the fact that children acquired rising or low-falling tones earlier than high-level or high-falling tones. All children had completed their four tones by 1;3, but they had not entirely stabilized the tonal production system. Based on the data, the rank order of the emergence of tones is suggested as follows: high-level tones > high-falling tones > low-falling tones > rising tones. Such an order of tone acquisition showed a slightly different order from the one found in the majority of scholars who agreed that children emerge low-falling tones in the last order.

3.2. Theoretical explanations

Evidence from articulatory effort theory (Ohala, 1978; Ohala & Ewan, 1972; Vihman, 1996), phonetic cues, involving duration and intensity (Fu & Zeng, 2000; Tseng, 1990; Xu et al., 2002), or the theory of markedness of tone (Yip, 2002) can all provide some explanations for such a rank order found in Taiwan Mandarin children. First, high-falling tones required less physiological effort so children need less energy and muscular control to produce such a natural gesture of falling pitch movement. Rising tones were the last tone to emerge in children's speech production since the level of difficulty would be greater for children as they would need more articulatory and physiological effort to produce the rising pitch. Secondly, phonetic cues might suggest that children easily acquired shorter duration and stronger intensity than vice versa, and falling tones have shorter duration and stronger intensity than rising tones. The physiological factors and articulatory complexity in relation to the order of tone acquisition have also been supported by Wong (2013). Finally, Yip's (2002) theory predicted the right rank order in which high-level tones are acquired earlier than other contour tones, and high- and low-falling tones are acquired earlier than rising tones. In her theory, rising tones would be the last one to be acquired. Therefore, all these relevant studies provide explanations for why rising tones are in the last order to emerge in the four-tone system, and this study showed a conflicting result to Clumeck (1977) who proposed the early emergence of rising tones over the rest of three tones.

In comparison to cross-linguistic studies, level tones seem to be acquired earlier than contour tones in Taiwan Mandarin as well as in Cantonese, Thai and Taiwanese, and as for contour tones, the study confirmed the cross-

linguistic findings in which high-falling tones are acquired earlier than low-falling tones.

3.3. Distribution patterns of tones

The final issue deals with the distribution patterns of tones, which are classified into monosyllables and disyllables. This study followed Lin's (2007) work which shows four basic tones plus 15 possible tone patterns in disyllables.

Table 4 *The distribution patterns of tone acquisition*

Tone	**0;9**	**0;10**	**0;11**	**1;0**	**1;1**	**1;2**	**1;3**	**1;4**	**1;5**	**1;6**	**1;7**	**1;8**	**Total**
Monosyllables													
T_1 (High-level)	1	1	1	1	3	4	5	6	1	7	N/A	N/A	30
T_2 (Rising)	0	0	0	0	0	1	2	12	7	3	1	2	28
T_3 (Low-falling)	0	0	0	0	4	2	2	6	1	1	N/A	N/A	16
T_4 (High-falling)	6	2	2	2	15	11	4	4	0	12	N/A	N/A	58
Subtotal	6	2	5	3	24	30	18	20	3	21	N/A	N/A	132
Disyllables													
T_1+T_1	0	0	0	1	5	12	6	6	4	0	N/A	N/A	34
T_1+T_2	0	0	0	0	0	1	1	2	0	0	N/A	N/A	4
T_1+T_3	0	0	0	0	0	0	0	0	0	0	N/A	N/A	0
T_1+T_4	0	3	2	0	0	1	1	0	0	0	N/A	N/A	7
T_2+T_1	0	0	0	0	0	0	0	0	0	0	N/A	N/A	0
T_2+T_2	0	0	0	0	0	3	0	0	0	0	N/A	N/A	3
T_2+T_3	0	0	0	0	3	3	9	2	7	3	N/A	N/A	27
T_2+T_4	0	0	0	0	0	0	0	3	0	0	N/A	N/A	3
T_3+T_1	0	0	0	0	2	2	5	0	0	0	N/A	N/A	9
T_3+T_2	0	0	0	0	7	13	13	7	4	3	N/A	N/A	47
T_3+T_4	0	0	0	0	1	0	0	0	0	0	N/A	N/A	1
T_4+T_1	1	0	0	0	0	0	0	1	0	0	N/A	N/A	2
T_4+T_2	0	0	0	0	1	0	0	0	0	0	N/A	N/A	1
T_4+T_3	0	0	0	0	0	0	0	0	0	0	N/A	N/A	0
T_4+T_4	1	0	1	0	0	5	2	5	2	0	N/A	N/A	16
Subtotal	2	3	6	2	15	37	28	25	7	3	N/A	N/A	154

Note: The tokens presented in this sample were the stabilized tones that had reached 66.7% of phonological accuracy and consistency; No data were found at 0;6-0;8 . Only one child stabilized rising tones at the first-word stage at 1;7-1;8. The rest of the children had entered the two-word stage after 1;6, so the columns were marked "N/A".

Based on the 66.7% criteria of phonological accuracy and consistency and compared to the tokens (Monosyllables=379 vs. Disyllables=660) presented at the stage of emergence of tones in Table 1, the table shows that 35% of monosyllables and 23% of disyllables were the corrected or stabilized tones, suggesting that the children do not stabilize the four-tone system yet, and they seem to have more difficulties in stabilizing the tones in disyllabic words at the one-word stage. In addition, this table shows that the tone patterns in the children's speech samples were not distributed evenly.

Regarding tones in monosyllables, this shows that at the one-word stage, high-falling tones occurred more often than high-level tones, followed by rising tones, and low-falling tones were the least common. Note that the distribution patterns in the children were deeply influenced by the caretakers' speech for monosyllabic words (N=1,729), in which an informal count of tone frequency showed that, for caretakers, high-falling tones were, in fact, more frequent than the other three tones (32%, N=553); high-level tones, 21% (N=363); rising tones, 19.7% (N=341); low-falling tones, 19.4% (N=335), neutral tones, 7.9% (N=137). In addition, the token counts from the speech corpus (N=604,916), *NCCU Corpus of Spoken Chinese*, showed that high-falling tones are the most frequent, suggesting that the linguistic environment seem to be related to children's preferred patterns as well.

Regarding tones in disyllables, there were in total 15 possible tone combinations in sequences of disyllables to be expected by chance, but so far the tonal patterns T1+T3, T2+T1, and T4+T3 have not been found. The endearment-tone pattern, T3+T2, was the most remarkably common one in the reduplicative forms, followed by the high-level sequence T1+T1, and T2+T3 was the third most common type. Preference for the 12 available tonal patterns is not equally distributed in the tokens. This suggests that Taiwan Mandarin children have a preference for such tonal patterns involving T3+T2, T1+T1, and T2+T3 when they produce disyllabic words at the one-word stage.

This finding was not surprising since the endearment tone T3+T2 was a unique language-specific phenomenon in Taiwan Mandarin, and such a pattern has been reported to be very frequent, particularly in Taiwan (Duanmu, 2007).

Yang (2013) examined the tone combination in sequences of disyllables in motherese, and found that the caretakers tended to produce the endearment tone, T3+T2, in their child directed speech most often. Since high-level tones were easily acquired and emerged, it should not be surprising to find that the high-level sequence, T1+T1, is the second most preferred pattern. The third most common pattern could be viewed as a morphologically-conditioned tone pattern, T2+T3 (N=27). Twenty-six cases in the current study were derived from the tone sandhi rule, and the underlying form was

the low-falling sequence, T3+T3, and in only one case was the true rising-low-falling combination of T2+T3 in the underlying representation.

Evidence from the speech corpus, *NCCU Corpus of Spoken Chinese*, showed the preferred rank order in adults' spontaneous conversation in the following order: T2+T3 > T4+T2 > T2+T1 > T4+T4 > T1+T4 > T4+T2 > T3+T4 > T1+T1. Children's disyllabic word production seemed not to have a close influence with the linguistic environment.

As for prosodic phonology, iambic foot structure was proposed in support of low-high contrast in pitch from H. Hsu's (2006) and Lu's (2011) studies. Although the present data supported binary foot structure, the children did not show a preference for low-high (iambic) pitch contrast in the disyllabic words. Tsao (2008) presented a perceptual study where rising and low-falling tones are the minimally distinct pair for the infants aged of 10 and 12 months to distinguish tone contrasts; however, such a minimally distinct pair turned out to be the preferred combination in disyllabic words for the present study.

The present study showed that children will not be able to stabilize and master the four-tone system at the one-word stage since the tones in monosyllables contained only 35% of accuracy rate, and the ones in disyllables included even lower, 23%, of accuracy rate. In the earlier study from Li and Thomspon (1977) in which children mastered their tonal systems with few errors by the age of 2;0, possibly at the one-word stage, and in the phonetic study of Chen and Kent (2009) where the distribution pattern of F_0 contours in babbling and early-word stages are alike, the lower accuracy rate in the present study instead provided a conflicting result. Since this study mainly focused on the tone acquisition at the one-word stage, and it might not be able to fully compete with the claims of Wong (2008, 2012, 2013) and Wong et al. (2005) for a support of the lengthy developmental course of tone acquisition.

Mandarin has only four tones, and one would possibly imagine that more complex tonal system such as Cantonese, Thai and Taiwanese would require longer period of timeline to stabilize the entire course; however, evidence from the majority of cross-linguistic acquisition studies on tone production, for which data have been reported so far, showed that the order of emergence-stabilization interval seems to be very short. Another perceptual study suggested that phonetic categories for tone evolve earlier than those for vowels and consonants even at the age of 4 months from Cantonese and Mandarin infants (Yeung et al., 2013). Although perceptual and production studies have sometimes showed conflicting results, the question remains unsolved to explain why Mandarin children are able to distinguish the four-tone system at the age of 4 months, acquire tonal production early, but stabilize the system after the age of 3; there is a long delay between the age of 4 months and the age of 3 or beyond. Moreover, conflicting statements have always raised another question as to whether the data sources are drawn from production/articulation or from acoustics/perception since they might yield different results (Myers, 2016a, 2016b).

4. Conclusions

Evidence from the current study has shown that, in terms of the child's age and the order of tone acquisition, there seems to be a universal tendency whereby high-level tones are likely to be acquired earlier than high-falling tones, followed by low-falling tones and rising tones. In terms of the stabilization of tone acquisition, this study does not support the early developmental course of tone acquisition, and children do not stabilize the four-tone system before 1;8 at the one-word stage. As for the distribution of tone patterns, high-falling tones are the most common in monosyllabic words, and the endearment tone, T3+T2, is the most preferred combination in disyllabic words at the one-word stage. Many sources from the points of view of articulatory theory, physiological effort, phonetic cues, and markedness constraints have provided some explanations to account for both the universal tendencies and the language-specific patterns found in the present study. Few studies have ever discussed the relative frequency of occurrence of tones in terms of perceptual explanations, so there is a need to find out whether children have a perceptual preference or bias towards any particular tonal patterns. The linguistic environment (i.e., caretaker's speech or speech corpora) might also provide a greater influence in modifying children's acquisition system by using innate articulatory and auditory templates (Locke, 1980, 1983; Kent, 1992).

Acknowledgments

We are very grateful for the comments and suggestions made by the two anonymous reviewers. Although the paper is due to the collaboration of the two authors, the respective contribution to the paper is 90% for the first author, and 10% for the second author. The first author is solely responsible for the content of the paper. Thanks also go to Prof. Chiu-Yu Tseng and her research team for providing free Mandarin alignment software in her phonetics lab. We are also grateful for obtaining free access to the NCCU corpus of spoken Chinese provided by Prof. Kawai Chui. For helpful discussions on this work and related issues, we thank Prof. Janice Fon and Prof. Feng-Ming Tsao. This research was supported by National Science Council Grants (NSC 100-2410-H-004-187, NSC 101-2410-H-004-182, NSC 102-2410-H-004-107) to the first author in Taiwan.

References

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Chao, Y. R. (1951). The Cantian idiolect: An analysis of the Chinese spoken by a twenty-eight-months-old child. In W. J. Fischel (Ed.), *Semantic and oriental studies*. Berkeley and Los Angeles: University of California Press.

Chao, Y. R. (1968). *A grammar of spoken Chinese.* Berkeley and Los Angeles: University of California Press.

Chen, L. M., & Kent, R. D. (2009). Development of prosodic patterns in Mandarin-learning infants. *Journal of Child Language, 36*(01), 73-84.

Clumeck, H. (1977). *Studies in the acquisition of Mandarin Phonology.* Ph.D. Dissertation. University of California, Berkeley.

Clumeck, H. (1980). The acquisition of tone. In G. D. Allen and S. Hawkins (Eds.), *Child Phonology. Vol: Production.* 257-275.

Crystal, D. (1986). Prosodic development. In Fletcher, P. and Garman, M. (eds.), *Language Acquisition* (2nd edition). Cambridge, Cambridge University Press.

Demuth, K. (1996). The prosodic structure of early words. In J. Morgan & K. Demuth (Eds.), Signal to Syntax: Bootstrapping from Speech to Grammarin Early Acquisition. Mahwah, N.J.: Lawrence Erlbaum Associates. 171-184.

Demuth, K. (2006). Cross-linguistic perspectives on the development of prosodic words. *Language and speech*, *49*, 129.

Demuth, K., & Johnson, M. (2003). Truncation to subminimal words in early French. *Canadian Journal of Linguistics*, *48*, 211-241.

Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of child language*, *35*(01), 99-127.

Duanmu, S. (2007). *The Phonology of Standard Chinese*. 2nd Edition. Oxford: Oxford University Press.

Fu, Q. & Zeng, F. (2000). Identification of temporal envelope cues in Chinese tone recognition, *Asia Pacific Journal of Speech, Language, and Hearing*, *5*, 45–57.

Hsu, J. H. (1987). *A study of the various stages of development and acquisition of Mandarin Chinese by children in Chinese milieu.* National Science Council project report.

Hsu, H. C. (2006). Revisiting Tone and Prominence in Chinese. *Language and Linguistics*, *7(1)*, 109-137.

Hsu, J. H. (2003). *A study of the stages of development and acquisition of Mandarin Chinese by Children in Taiwan.* Crane Publishing: Taipei.

Kaplan, E. & Kaplan, G. (1971). The Prelinguistic Child. In J. Eliot (Ed.), *Human Development and Cognitive Process.* New York: Holt, Rinehart, Winston. 359-381.

Kent, R. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, and C. Stoel-Gammon (Eds.) *Phonological Development: Models Research, Implications. Timonium*, MD: York Press.

Li, C. N. & Thompson, S. A. (1977). Acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, *4*(2), 185-199.

Lin, Y. H. (2007). *The sounds of Chinese*. Cambridge: Cambridge University Press.

Locke, J. (1980). The prediction of child speech errors: implications for a theory of acquisition. In G. H. Yeni-Komsnian, J. F. Kavanagh, & C. A. Ferguson (Eds.) Child Phonology, I: Production. New York: Academic Press.

Locke, J. (1983). *Phonological Acquisition and Change.* New York: Academic Press.

Lu, P. H. (2011). *A Linguistic Exploration of Tone in Chinese Personal Names*. Master's thesis, Tsing Hua University: Hsinchu, Taiwan.

Miyakoda, H. (2012). Markedness within phonological theory: Focus on acquisition and disorder data. *Acoustical Science and Technology*, *33*(3),142-146.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143-178.

Myers, J., (2016a). Psycholinguistics, overview. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, & J. Myers (Eds.), *Encyclopedia of Chinese language and linguistics.* Leiden, Netherlands: Brill.

Myers, J., (2016b). Psychological reality of linguistic structure. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, & J. Myers (Eds.), *Encyclopedia of Chinese language and linguistics.* Leiden, Netherlands: Brill.

Ohala, J. J. (1978). The production of tone. In V. Fromkin (ed.), *Tone: a linguistic survey.* New York: Academic Press.

Ohala, J. J. & Ewan, W. (1972). Speed of pitch change. *Journal of the Acoustical Society of America*, *53*, 345.

Ong, S. & Yang, C. (1997). On the energy and duration of the four tones of Mandarin speech). *Journal of Technology*, *12*(1), 125-129.

Ota, M. (1999). *Phonological theory and the acquisition of prosodic structure: Evidence from child Japanese.* Ph.D. dissertation. Georgetown University.

Singh, L., & Fu, C. S. (2016). A new view of language development: the acquisition of lexical tone. *Child development*, *87*(3), 834-854.

So, K. H. & Dodd, B. J. (1995) Acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*, *22*(3), 473-495.

Tsao, F.-M. (2008). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chinese Journal of Psychology*, *50*, 111–124.

Tsay, J. (2001) Phonetic parameters of tone acquisition in Taiwanese. In M. Nakayama (Ed.), *Issues in East Asian Language Acquisition*. Tokyo: Kuroshio Publishers, 205-226.

To, C. K., Cheung, P. S., & McLeod, S. (2013) A population study of children's acquisition of Hong Kong Cantonese consonants, vowels, and tones. *Journal of Speech, Language, and Hearing Research*, *56*(1), 103-122.

Tse, A. C.-Y., (1991) The acquisition process of Cantonese phonology: A case study. *HKU Theses Online (HKUTO)*.

Tseng, C. (1990) *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Taipei: Institute of History and Philology.

Tuaycharoen, P. (1977) *The Phonetic and Phonological Development of a Thai Baby: From Early Communicative Interaction to Speech*. Ph.D. dissertation. School of Oriental and African Studies, University of London.

Vihman, M. M. (1996) *Phonological Development*. Oxford: Blackwell.

Vihman, M. M., & McCune, L. (1994) When is a word a word? *Journal of child language*, *21*(03), 517-542.

Wong, P. (2008). Development of lexical tone production in disyllabic words by 2- to 6-year-old Mandarin-speaking children. Doctoral Dissertation. The Graduate Center of the City University of New York.

Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics*, *40*(1), 141-151.

Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America*, *133*(1), 434-443.

Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1065-1079.

Xu L., Tsai, Y. & Pfingst, B. (2002). Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," *Journal of the Acoustical Society of America*, *112*(1), 247-258.

Yang, H. J. (2013). *Tone acquisition in Taiwan Mandarin.* MA thesis. National Chengchi University, Taipei.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, *68*, 123–139.

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

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

Zhu, H. & Dodd. B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language*, *27*3, 3-42.