Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Tone acquisition in Chinese-speaking children: Developmental data of tone acceptability and contour pattern

Shu-Chuan Tseng

Institute of Linguistics, Academia Sinica
tsengsc@gate.sinica.edu.tw

Abstract

This paper presents developmental data from 798 Chinese-speaking children aged three to six. An overall assessment of tone acquisition is reported with descriptive results from tone acceptability judgement, pitch contour type labeling and F0 slope patterns. The order of tone acquisition is Tone 4, Tone 1, Sandhi Tone 3, Tone 3, and Tone 2. High-register tones are generally acquired before low-register tones. Pitch contour types of perceptually acceptable tones primarily conform to phonologically predicted tone contours including Sandhi Tone 3. Unacceptable tones are actually more diverse. Based on the F0 property, the proportions of immediately erroneous slope patterns in Tone 2 are higher than in Sandhi Tone 3. As a whole, our developmental data consistently suggests that Tone 2, instead of Sandhi Tone 3, is the most challenging tone to acquire for Chinese-speaking children.

Index Terms: phonological development, acceptability, pitch contour

1. Introduction

Phonological development data focus on phoneme acquisition with percentages of correct vowels and consonants. These percentages can be used as quantitative indices for assessing children's speech acquisition status [1-4]. Nonetheless, tonal aspects play a vital role in speech communication in a language such as Mandarin Chinese. Previous literature in speech acquisition and pathology reported that lexical tones are acquired at two to three years of age, normally earlier than segments [2, 5, 6]. Despite tones' essential word-prosodic features, there are no unanimously accepted tone quality metrics. As a matter of fact, proper tone production implies listeners can infer meaning without difficulty. From the perspective of spoken word processing, how pitch contour is perceived by auditory judgement, how to interpret and assign tone category by mapping to the phonological system of tones, and how word meaning is decoded by referring to segmental and suprasegmental information are three different, but correlated processing modules of tones [4]. That is also one of the reasons why judging monosyllabic tones by whether or not they match the canonical form can only reveal part of their role in speech production.

This paper reports descriptive data on tone acquisition in 3 to 6-year-old Chinese-speaking children in Taiwan. With our developmental data, we assess tone quality by acceptability, pitch contour perception and speech signal representation. As tone is a word prosody feature, tone acceptability judgement of syllables is conducted by specifically referring to the meaning of the affiliated word. For future automatic assessment applications, this paper also examines to what extent objective acoustic cues can be employed to represent the phonetic properties of child speech [7, 8].

1.1. Tone inventory in Mandarin Chinese

Mandarin Chinese has four canonical tone categories, represented by high-level (Tone 1), low-rising (Tone 2), low-falling-rising (Tone 3), and high-falling (Tone 4) as well as an unstressed, neutral tone [9, 10]. They distinguish word meaning beyond its segmental components. Phonetically speaking, in within- or across-word contexts, tone coarticulation and transition are accompanied by different types and degrees of adaptations in tonal contours [11]. Phonologically, Tone 3 has three predicted variants, having a low-falling-rising contour (canonical Tone 3), a low-rising contour (Sandhi Tone 3), and a low-falling contour (half Tone 3), where only the first low-falling part is preserved and the second rising part is omitted. In this paper, we use Hanyu Pinyin to transcribe Chinese characters. Tones are noted with diacritics on top of the main vowels, i.e., *yāo* (to invite), *yáo* (to swing), *yǎo* (to bite), *yào* (to want).

1.2. Acquiring lexical tones

Tone is acquired early in childhood around three. Children aged three to six have 99% of their tones correct [2]. Tone acquisition is completed before speech sounds for Beijing Mandarin, Taiwan Mandarin, and Cantonese [3, 5, 6, 12]. Low-register tones in Mandarin Chinese are less adult-like than high-register tones for children [13]. The order of tone acquisition in hearing-impaired children is similar to that of developing children, only with lower accuracy rates [14]. F0 and spectral features of Chinese-speaking children's speech do not appear as distinctive in tone production as adults' speech until eleven [15]. Specifically, Tone 1 and Tone 4 share more similar qualities in onset register level and contour trend with their adult counterparts than Tone 2 and Tone 3 [16]. Tone discrimination and spoken word recognition were examined to explore how tone contrast is acquired by Cantonese-speaking children [17]. With the results of pitch contour judgement of disyllabic words in Chinese-speaking hearing-impaired and hearing children, the effects of syllable position on tone accuracy were reported in first and second language acquisition [18, 19]. Usage-based functions of comprehension and speech communication have increasingly attracted attention to linguistic research and speech pathology applications [7, 8, 20, 21], suggesting that word-level features should also be considered for tone acquisition research.

2. Data and methodology

Descriptive, but comprehensive results of two experiments conducted on 798 preschoolers with normal hearing are presented. In the first experiment, spoken words are evaluated to see if the way tones are produced facilitates or prohibits word

 10.21437/SpeechProsody.2024-11

comprehension. In the second experiment, pitch contour types are labeled by pitch trends perceived by listeners. The token order was randomized in both experiments. Neither the annotator participated in the other experiment. In addition, the main F0 slope types are examined to explore possible correlations with pitch type perception and signal representation. This is done to lay a preliminary foundation for future automatic system development.

2.1. Sinica Phonological Development Corpus

The *Sinica Phonological Development Corpus* contains speech recordings of 798 preschool children, as shown in Table 1 [22]. None of them had known or diagnosed diseases related to language, hearing, or cognitive development. They all passed a pure-tone audiometric screening at 1, 2, and 4 kHz at 20 dB. Seventy multisyllabic words in the *Sinica Child Balanced Wordlist* were used for data elicitation, as shown in Table 2 [23]. Onset consonants occur at least once in word-initial and -final positions. Disyllabic tones are balanced except for the neutral tone. For data collection, CapiAssess was installed on a MacBook Air Pro Retina 13.3 laptop with a Sony ECM MS907 microphone. 55,860 words were digitized at a sampling rate of 16 kHz and processed by the publicly available *ILAS Phone Aligner* to obtain syllable boundaries for F0 extraction in PRAAT [24].

Table 1: *Subjects*

AGE	3~3;6	3;6~4	4~4;6	4;6~5	5~5;6	5;6~6	6~6;6	6;6~7	Total
Male	10	40	64	55	65	64	79	22	399
Female	21	52	58	64	53	62	60	29	399

Table 2: *Sinica Child Balanced Wordlist*

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

2.2. Tone acceptability judgement

For the tone acceptability experiment, two annotators were instructed to judge whether the tone in each word is properly produced by referring to word meaning. A tone is labeled “acceptable” if the word associated with the produced tone is understandable without difficulty. A tone is labeled "unacceptable" if the tone quality hinders comprehension. Two annotators listened to sound files of 55,860 recorded words independently and gave their acceptability judgment to all 118,104 syllables. They were instructed to pay attention primarily to word meaning and not to specific pitch contours. The agreement rate is 0.983. Cohen’s Kappa is low, 0.42 due to our skewed data with acceptable tones as the dominant majority. Inconsistent cases for which no consensus can be achieved after discussion sessions, were assigned “unacceptable” as the final label.

2.3. Pitch contour type labeling

Different from the tone acceptability experiment, the pitch contour type labeling experiment was designed to examine to what extent a pitch contour perceived by native speakers corresponds to the phonological tone categories. Two annotators listened to sound files of syllables segmented from affiliated words. The annotators were instructed to select one from flat, rising, falling-rising, falling, short, and unclassifiable to describe the pitch contour type they heard. Register level was not considered in this experiment. All 118,104 segmented syllables were labeled with an agreement rate of 0.84. Cohen’s Kappa is 0.76. For the analysis of proportions presented below, we only use results of 99,391 syllable tokens annotated with consistent pitch contour types.

2.4. F0-based contour

Preliminary acoustic analysis of tone contour was conducted to seek possible objective metrics for automatic assessment of tones to supplement impressionistic judgements. Adopting the tone visualization approach proposed in [16], F0 features representing phonological contrast of tones such as onset and offset values, slopes of fitted F0 lines (at most two lines) and time stamps of turning points in the case of two fitted lines are used. F0 ranges defined as F0 values allocated between 0.01% and 0.99% of each speaker are normalized for cross-speaker comparison. All speech data of 55,860 words were processed with the same procedure using manually verified syllable boundaries. Figure 1 shows F0 contours of four lexical tones in Mandarin Chinese spoken by a native adult female speaker.

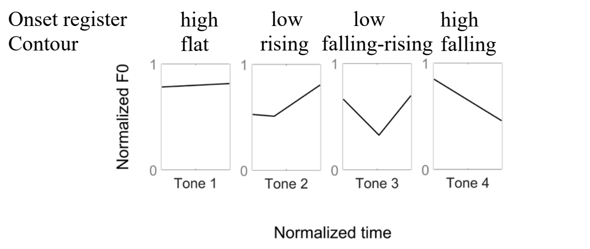

Figure 1: *F0-visualization of tones, adapted from [16]*

3. Tone acquisition in disyllabic words

As the *Sinica Child Balanced Wordlist* has a balanced design for the four lexical tones in disyllabic words, this paper focuses on the results of 49,476 disyllabic words. Developmental data in terms of proportion analysis are reported in the order of acceptability judgement, pitch contour type labeling, and F0 slope patterns.

3.1. Tone acceptability by meaning association

For multisyllabic words, processing of word meaning and tone quality acts at different linguistic levels. In our experiment, listeners judge tone acceptability explicitly based on whether the word that contains the to-be-judged tone can be understood and not based on the perceived pitch contour of the tone. Table 3 shows the developmental results of tone acceptability. Because Sandhi Tone 3 was found to be particularly problematic or in some studies even the most problematic for children, we added an extra "Sandhi Tone 3" category. The "Tone 3" category contains all Tone 3 syllables, including Sandhi Tone 3; the “Sandhi Tone 3” category consists of only Sandhi Tone 3 syllables.

Table 3: *Acceptability judgement results*

(%)		All	Tone 1	Tone 2	Tone 3	Sandhi Tone 3	Tone 4
Female		97.96	98.94	96.29	97.78	98.75	98.91
	3~3;6	96.85	98.81	94.15	96.43	100	98.25
	3;6~4	96.63	98.38	92.75	96.79	98.08	98.91
A	4~4;6	96.51	98.01	94.73	95.83	94.83	97.36
G	4;6~5	97.87	98.97	96.29	97.53	98.44	98.7
E	5~5;6	98.43	99.12	96.77	98.51	100	99.5
	5;6~6	98.97	99.4	98.29	98.92	100	99.41
	6~6;6	98.97	99.27	98.14	98.96	100	99.61
	6;6~7	99.11	99.89	98.33	98.71	100	99.43
Male		97.57	98.74	95.75	97.14	98.5	98.86
	3~3;6	95.56	98.13	90.29	97.08	100	98
	3;6~4	96.31	98.36	94.5	94.69	95	97.92
A	4~4;6	96.52	98.24	93.93	95.7	95.31	98.44
G	4;6~5	97.24	98.52	94.34	97.12	100	99.21
E	5~5;6	97.44	98.56	95.16	97.12	100	99.33
	5;6~6	98.15	99.12	97.28	97.66	98.44	98.59
	6~6;6	98.69	99.13	97.97	98.52	100	99.2
	6;6~7	99.3	99.72	98.57	99.43	100	99.55

Consistent with previous research, the acceptability rates of each tone, including Sandhi Tone 3, reach the 90% acquisition threshold with an increase across age groups. Although tones are acquired early, we still can observe the development of tone accuracy until the age of seven. And there is a clear difference between 95% and 99% tone accuracy in children aged three and six. High-register tones, Tone 1 and Tone 4, have higher acceptability rates than low-register Tone 2 and Tone 3. It should be noted that neither Sandhi Tone 3 nor Tone 3 in general have the lowest acceptability rates, but Tone 2.

3.2. Tone 2 is challenging

We then examined Tone 2 results in terms of gender and syllable position. Table 4 shows similar development patterns of Tone 2 acquisition in both genders. Tone 2 in the second syllable position apparently causes more difficulty than in the first position. For prospective works, not only on children’s but also on adults’ speech, analysis should be conducted on syllable- and word-level factors that may play a role in tone production. To capture nuanced differences in tone accuracy, sophisticated indicator selection is required. According to our results, we would suggest using accuracy rates of non-word-initial Tone 2 as an efficient index.

Table 4: *Acceptability rates of Tone 2 in disyllabic words*

(%)		Female		Male	
Position in word		Initial	Final	Initial	Final
All Tone 2 tokens		97.07	94.97	96.61	94.29
	3~3;6	96.1	90.84	90.45	90
	3;6~4	93.88	90.83	96.02	91.92
A	4~4;6	95.45	93.5	94.74	92.55
G	4;6~5	97.16	94.83	95.29	92.73
E	5~5;6	97.68	95.21	96.5	92.9
	5;6~6	98.68	97.64	97.87	96.27
	6~6;6	98.79	97.05	98.5	97.08
	6;6~7	98.43	98.14	98.97	97.9

3.3. Tone category vs. pitch contour type

Table 5 shows that the majority of Tone 1 tokens are perceived as having a flat contour, Tone 2 tokens a rising contour, Tone 3 tokens a falling contour, Sandhi Tone 3 tokens a rising contour, and Tone 4 tokens a falling contour. In our disyllabic data, very few tokens of non-Sandhi Tone 3 were perceived to have a falling-rising contour type. Overall, the contour types heard by native adult listeners conform to phonologically predicted contour patterns. Increasing proportions of phonologically predicted contours are also observed in older age groups.

Table 5: *Pitch contour type perception (row total = 100%)*

AGE	(%)	Flat	Rising	Falling-rising	Falling	Unclassifiable
3~3;6	Tone 1	53.53	18.65		3.23	24.6
	Tone 2	11.89	59.91		9.68	18.53
	Tone 3	2.69	6.45		77.02	13.84
	Sandhi Tone 3		74.19		9.68	16.13
	Tone 4	1.51	0.43		89.68	8.39
3;6~4	Tone 1	50.34	17.32		6.15	26.19
	Tone 2	10.96	63.57		9.07	16.40
	Tone 3	2.72	5.21		82.38	9.69
	Sandhi Tone 3	2.17	90.22		2.17	5.43
	Tone 4	0.69	0.51		89.42	9.38
4~4;6	Tone 1	51.92	17.98		5.4	24.69
	Tone 2	10.98	65.74		6.89	16.39
	Tone 3	2.73	5.60		79.71	11.95
	Sandhi Tone 3	5.74	82.79		3.28	8.20
	Tone 4	0.96	0.49		90.11	8.44
4;6~5	Tone 1	54.83	19.64		3.65	21.88
	Tone 2	9.46	68.64		6.77	15.13
	Tone 3	2.1	5.88	0.04	82.32	9.66
	Sandhi Tone 3	4.2	94.12		0.84	0.84
	Tone 4	0.45	0.31		92.18	7.06
5~5;6	Tone 1	56.17	17.27		4.05	22.51
	Tone 2	8.6	68.55		7.89	14.96
	Tone 3	2.4	6.25	0.07	82.17	9.11
	Sandhi Tone 3	1.69	88.98		2.54	6.78
	Tone 4	0.56	0.37		92.23	6.84
5;6~6	Tone 1	55.46	19.3		2.9	22.35
	Tone 2	5.71	78.57		4.74	10.98
	Tone 3	1.88	6.12		82.61	9.39
	Sandhi Tone 3	2.38	92.06		0.00	5.56
	Tone 4	0.37	0.11		93.44	6.08
6~6;6	Tone 1	57.13	18.55		2.14	22.19
	Tone 2	6.78	77.64		4.4	11.18
	Tone 3	1.77	5.7		83.54	8.99
	Sandhi Tone 3	1.44	92.81		1.44	4.32
	Tone 4	0.24	0.17		93.65	5.95
6;6~7	Tone 1	59.5	15.93		2.39	22.18
	Tone 2	5.71	79.22		5.1	9.97
	Tone 3	2.12	5.80		84.23	7.76
	Sandhi Tone 3	1.96	94.12			3.92
	Tone 4	0.33	0.26		93.07	6.34

3.4. Tone perception: Meaning association vs. contour type

When a Chinese word is spoken, tones affect meaning decoding. Nevertheless, it is unclear how tone shape is concretely related to this process. Figure 2 shows that for perceptually judged acceptable tones, their pitch contour perception corresponds to what is expected from canonical tones for all tone categories except Tone 1. Comparing the results of both experiments, we notice that only around half of Tone 1 tokens judged properly produced for comprehension were perceived to have a flat contour. The other half either have rising contours or are unclassifiable. This is likely due to tone coarticulation that causes a descending or ascending pitch transition phase across the syllable boundary. It is not trivial to assign pitch contour type to multisyllabic words based on impressionistic judgement alone. This would also imply considerable confusion in the speech signal representation used to automatically process tone quality.

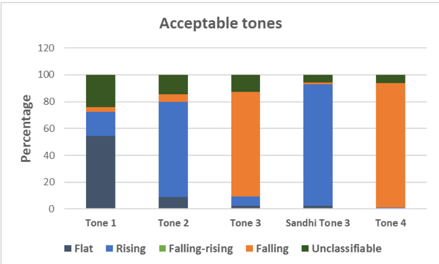

Figure 2: *Pitch contour types in ACCEPTABLE tones*

The pitch contour labeling results for unacceptable tones are shown in Figure 3. "Unacceptable tones" are meant to have "incorrect" pitch contours that prohibit word comprehension. Figure 3 indeed demonstrates larger proportions of pitch contour types different from canonical tone contours than acceptable tones in Figure 2. In particular, Tone 1, Tone 2, and Sandhi Tone 3, have diverse pitch contour types. However, not all unacceptable tones are "incorrect". A majority of Tone 3 and Tone 4 were still labeled with falling contours that matched the predicted tone contour patterns.

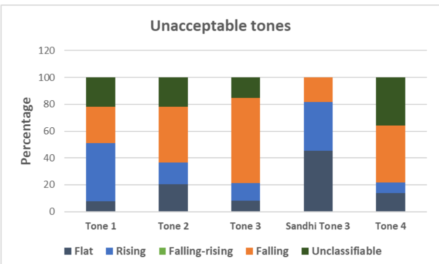

Figure 3: *Pitch contour types in UNACCEPTABLE tones*

3.5. F0 slope types

Differences in acoustic signal representation such as F0 slope types should to some degree reflect tone acquisition stages. Negative/positive slopes (-/+) imply falling/rising trends in the linearized F0 contours. Table 6 presents the four main slope types (-, +, +-, -+), making up 97% of the overall data, equivalent to 95,505 disyllabic words. Distribution across age groups is symmetrical. As stated above, Tone 1's flat contour could be affected by tone coarticulation and thus cause confusion in the acoustic signal. But for Tone 2 and Sandhi Tone 3, a single fitted line with a falling slope would indicate an inadequate tone pattern. This would be considered erroneous tone production. For both Tone 2 and Sandhi Tone 3, the percentages of negative-slope tokens decrease with age, with Tone 2 larger than Sandhi Tone 3. Tone 2 acquisition seems slower than Sandhi Tone 3. Besides, the percentages of erroneous Tone 4 tokens with a positive slope are very low, which is also consistent with our acceptability and pitch contour type results.

Table 6: *Main slope pattern of linearized F0 contour (Total of each age group under each tone category = 100%)*

(%)		AGE							
	Slope	3~3;6	3;6~4	4~4;6	4;6~5	5~5;6	5;6~6	6~6;6	6;6~7
Tone 1	-	40.06	38.59	38.65	35.04	34.89	29.55	31.44	31.67
	+	22.92	22.76	23.71	25.75	25.02	28.2	26.88	25.74
	+ -	5.27	7.89	7.74	7.41	8.13	9.36	7.88	8.58
	- +	31.74	30.76	29.9	31.79	31.95	32.89	33.8	34.01
Tone 2	-	**24.58**	**23.49**	**22.73**	**20.53**	**20.94**	**15.92**	**16.3**	**17.28**
	+	30.06	28.87	29.38	30.46	29.74	33	31.37	31.4
	+ -	4.64	4	4.13	3.79	4.04	3.05	2.98	2.59
	- +	40.72	43.64	43.76	45.22	45.28	48.02	49.35	48.73
Tone 3	-	73.25	77.65	78.31	78.19	78.14	78.85	79.34	79.79
	+	2.72	2.56	2.26	2.66	2.26	2.42	2.13	1.99
	+ -	8.87	7.96	7.5	7.59	7.94	7.42	7.6	7.51
	- +	15.16	11.82	11.94	11.56	11.66	11.31	10.93	10.71
Sandhi Tone 3	-	**25.81**	**11.96**	**10.74**	**5.04**	**6.78**	**8**	**5.76**	**3.92**
	+	29.03	42.39	50.41	47.06	42.37	41.6	46.04	45.1
	+ -	6.45	4.35	2.48	2.52	3.39	3.2	0.72	0
	- +	38.71	41.3	36.36	45.38	47.46	47.2	47.48	50.98
Tone 4	-	77.68	74.97	74.55	74.4	73.79	74.01	72.79	74.88
	+	1.73	1.82	1.56	1.46	1.28	1.23	1.4	1.65
	+ -	14.63	17.3	17.84	19.11	19.54	20.52	20.9	19.78
	- +	5.96	5.90	6.04	5.04	5.38	4.24	4.92	3.69

4. Conclusion

This study presents three types of analyses of developmental data from a corpus of 798 preschool Chinese-speaking children that supplement previous tone acquisition studies. The order of tone acquisition based on acceptability judgement is Tone 1, Tone 4, Sandhi Tone 3, Tone 3, and Tone 2. This is slightly different from previous studies. Tone 2 and Sandhi Tone 3 both have rising contours. Tone 2 seems to be the most difficult tone, while Sandhi Tone 3 did not cause apparent difficulty despite Tone 3 having three variants. Besides, we found a wide variety of pitch contour types and F0 slope patterns in Tone 1. This may be due to varied phonetic properties in different phonological contexts. However, clear development trends towards phonologically predicated tone contours can be identified from perspectives of word meaning, pitch type, and F0 property. This paper presents multi-faceted tone acquisition data that can serve as the foundation for prospective research on the interrelationship between overall speech ability and tone production as well as for potential system development for clinical applications.

5. Acknowledgements

We are grateful for the support of Academia Sinica (LING-112-002-01) and the National Council for Science and Technology (MOST-108-2410-H-001-041, 109-2410-H-001-087, and 111-2410-H-001-015) as well as teachers, parents and children who participated in the recording project from 2017 to 2019.

6. References

[1] S. A. S. Lee, T. M. Nguyen Phuoc, N. Q. Truong Thi, H. T. Dang Thi, L. M. Ho Thi, and N. C. Ha, "Phonological acquisition in Vietnamese-speaking children with central dialect," *Speech, Language and Hearing,* pp. 1-17, 2023.

[2] G. B.-G. Lin and M.-h. Lin, "A Study on the Development of Language Abilities in Preschool Children in Taiwan, R.O.C," *Bulletin of Special Education,* vol. 10, pp. 259-281, 1994.

[3] Z. Hua and B. Dodd, "The phonological acquisition of Putonghua (modern standard Chinese)," *Journal of child language,* vol. 27, no. 1, pp. 3-42, 2000.

[4] S. McLeod and S. Verdon, "A review of 30 speech assessments in 19 languages other than English," *American Journal of Speech-Language Pathology,* vol. 23, no. 4, pp. 708-723, 2014.

[5] C. N. Li and S. A. Thompson, "The acquisition of tone in Mandarin-speaking children," *Journal of Child Language,* vol. 4, no. 2, pp. 185-199, 1977.

[6] C. K. To, P. S. Cheung, and S. McLeod, "A population study of children's acquisition of Hong Kong Cantonese consonants, vowels, and tones," *Journal of Speech, Language, and Hearing Research,* vol. 56, pp. 103-122, 2013.

[7] G. P. Usha and J. S. R. Alex, "Speech assessment tool methods for speech impaired children: a systematic literature review on the state-of-the-art in Speech impairment analysis," *Multimed Tools Appl,* pp. 1-38, Mar 6 2023, doi: 10.1007/s11042-023-14913-0.

[8] A. Chen, R. Zhao, G. Huang, A. Li, and H. Cheung, "Successful lexical tone production of Mandarin Chinese autistic children with intellectual impairment," *Int J Lang Commun Disord,* vol. 58, pp. 1912-1926, May 4 2023, doi: 10.1111/1460-6984.12881.

[9] S. Duanmu, *The phonology of standard Chinese*. OUP Oxford, 2007.

[10] Y.-H. Lin, *The Sounds of Chinese with Audio CD*. Cambridge University Press, 2007.

[11] Y. Xu, "Contextual tonal variations in Mandarin," *Journal of phonetics,* vol. 25, no. 1, pp. 61-83, 1997.

[12] P. Wong, R. G. Schwartz, and J. J. Jenkins, "Perception and production of lexical tones by 3-year-old, Mandarin-speaking children," *Journal of Speech, Language, and Hearing Research,* vol. 48, pp. 1065-1079, 2005.

[13] P. Wong, "Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions," *Journal of Phonetics,* vol. 40, no. 1, pp. 141-151, 2012.

[14] S.-C. Peng, J. B. Tomblin, H. Cheung, Y.-S. Lin, and L.-S. Wang, "Perception and production of Mandarin tones in prelingually deaf children with cochlear implants," *Ear and hearing,* vol. 25, no. 3, pp. 251-264, 2004.

[15] N. Rhee, A. Chen, and J. Kuang, "Going beyond F0: The acquisition of Mandarin tones," *J Child Lang,* vol. 48, no. 2, pp. 387-398, Mar 2021, doi: 10.1017/S0305000920000239.

[16] S.-C. Tseng and Y.-F. Liu, "Segment and Tone Production in Continuous Speech of Hearing and Hearing-Impaired Children," presented at the Interspeech 2021, 2021.

[17] L. K. So and B. J. Dodd, "The acquisition of phonology by Cantonese-speaking children," *Journal of child language,* vol. 22, no. 3, pp. 473-495, 1995.

[18] Y.-C. Hao, "Contextual effect in second language perception and production of Mandarin tones," *Speech Communication,* vol. 97, pp. 32-42, 2018, doi: 10.1016/j.specom.2017.12.015.

[19] H.-F. Chang, H.-Y. Gu, and J.-H. Wu, "A pilot study on human listener evaluation and computerized tone-contour analysis of Mandarin disyllable utterances by hearing-impaired students.," *Bulletin of Special Education,* vol. 26, pp. 221-245, 2004.

[20] M. Tomasello, "The social-pragmatic theory of word learning," *Pragmatics,* vol. 10, no. 4, pp. 401-413, 2000.

[21] S. McLeod and E. Baker, *Children's speech: An evidence-based approach to assessment and intervention*. Pearson, 2017.

[22] S.-C. Tseng, "Corpus-based research on speech acquisition and automatic assessment of Taiwan Mandarin-speaking children aged 3 to 6," in *Diversity in Linguistics, but Unity in Research: Celebrating the Twientieth Anniversary of the Institute of Linguistics, Academia Sinica*, S.-C. Tseng and E. Zeitoun Eds.: Academia Sinica, 2024, pp. 475-502.

[23] S.-C. Tseng, "ILAS Chinese spoken language resources," presented at the the third International Symposium on Linguistic Patterns in Spontaneous Speech, Taipei, 2019.

[24] P. Boersma and D. Weenink. "Praat: doing phonetics by computer. Version 6.2.14. Retrieved 24 May 2022 from http://www.praat.org/."