Interlanguage tone patterns in Thai pre-school children: A preliminary corpus analysis

Abstract:

The aim of this paper is to present a preliminary analysis by providing a detailed corpus study of interlanguage tone patterns made by 11 pre-school children in Thailand who learn Mandarin as their global language at Thai-Chinese International School in Bangkok, Thailand. The novelty of this work is to collect and analyze some data from a highly reliable corpus, and provide Praat acoustic parameters for the tone distribution by looking at various tone combinations measured by plotting the F0 variations in Mandarin. These units involved in the errors are classified by segmenting word shapes into monosyllabic, disyllabic or multi-syllabic. Evidence from the data shows that Thai children have no problem producing Mandarin tones, and are more likely to utter with disyllabic lexicons. Different from the traditional analysis on the error patterns, this study has suggested that interlanguage can be viewed as the transitional process between mother tongue and target language, and those tone error patterns are not fully related to their mother tongue, possibly suggesting that Thai children might have constructed an independent linguistic system gradually towards Mandarin. In the near future, a phonetically perceptual task in a low-pass filter for reserving tone information will need to be conducted in the hope of eliminating the transcribers' bias to rely on the lexical information.

Keywords: interlanguage tone patterns, Thai pre-school children, spoken corpus, Mandarin Chinese

1. Introduction

Earlier studies have suggested that infants are sensitive to prosodic cues during their language developmental stage, and children start to acquire suprasegmental features such as stress, intonation, pitch and tone very early in the developmental process (e.g., Kaplan and Kaplan 1971; Clumeck 1980; Mehler et al. 1988; Demuth 1996). Evidence from Crystal (1986), Mehler et al. (1988) and Demuth (1996) suggested that suprasegmental features, such as intonation and stress, are acquired earlier and better than segments. Evidence from 8 children (0;6-1;8) in Taiwan Mandarin leads Wan and Yang (2017) to suggest the similar pattern and agreed that tone is considered one of the most salient features in tone languages.

Due to different contour and level tones in the tone languages, studies reported tone acquisition varied in Cantonese, Mandarin, Taiwanese, and Thai (e.g., Mandarin spoken in the United States: Chao 1951, 1968; Clumeck 1977, 1980; Taiwan Mandarin: Li and Thompson 1977; Hsu 2003; Wong, Schwartz and Jenkins 2005; Wong 2008, 2012, 2013; Chen and Kent 2009; Wan and Yang 2017; Beijing Mandarin: Zhu and Dodd 2000; Zhu 2002; Cantonese: Tse 1991; So and Dodd 1995; To et al. 2013; Taiwanese: Hsu 1987; Tsay 2001; Thai: Tuaycharoen 1977). Cross-linguistic studies on acquisition from Thai, Cantonese, Taiwanese and Mandarin generally agreed that

level tones are in general acquired earlier than contour tones (Thai: Tuaycharoen 1977; Cantonese: Tse 1991; So and Dodd 1995; Taiwanese: Hsu 1987; Tsay 2001; Mandarin: Wan & Yang, 2017).

However, different from the acquisition process, studies on language learning have shown that lexical tones are the most difficult linguistic unit to acquire and be stable, so error patterns will vary depending on international learners' language background (Zeng 1999). Studies reported in Chinese and English showed that international learners might have different learning difficulties depending on their mother tongues (e.g., Yue 1986; Yu 1988; Shen 1989; Leather 1990; Wang 1995; Wang Y. 1995; Chen 1997; Feng 1997; Miracle 1999; Zeng 1999, 2001, 2007; Liu 2006; Zeng and Furukawa 2006; Lin 2007; Guo and Tao 2008; Tsai 2008; Liao 2010; Chen 2011; Ding 2012; Huang 2013; Huang 2014). The relevant studies have found that in production or perception, the most difficult linguistic element in Mandarin is on rising and low-falling tones for Dutch, English, Japanese and Korean learners (e.g., Yu 1988; Leather 1990; Wang 1995; Wang, Y. 1995; Chen 1997; Feng 1997; Miracle 1999; Liu 2006; Lin 2007; Tsai 2008; Liao 2010; Chen 2011; Huang 2013; Huang 2014).

However, for the other international learners, who have similar tone language background such as Vietnam or Thai, the error patterns will be entirely different. For instance, Vietnamese learners who would find that the most difficult tones to learn in both perception and production are high tone and high-falling tone (Zeng 1999; Wu and Hu 2004; Tran 2005; Chen 2011). However, Thai learners show very little difficulties in producing tones, but in terms of perception, the error rate in perceiving rising tone is significantly low, compared to a higher correction rate in perceiving high-falling tone (Chen 2011). In general, if their mother tongues are stress or pitch-accent languages, tone patterns on rising and low-falling tones have proved to be the most difficult features to be learned; however, if their native language are tone languages, which have more complicated tone contours than Mandarin, mastering tones might not be a problem for them.

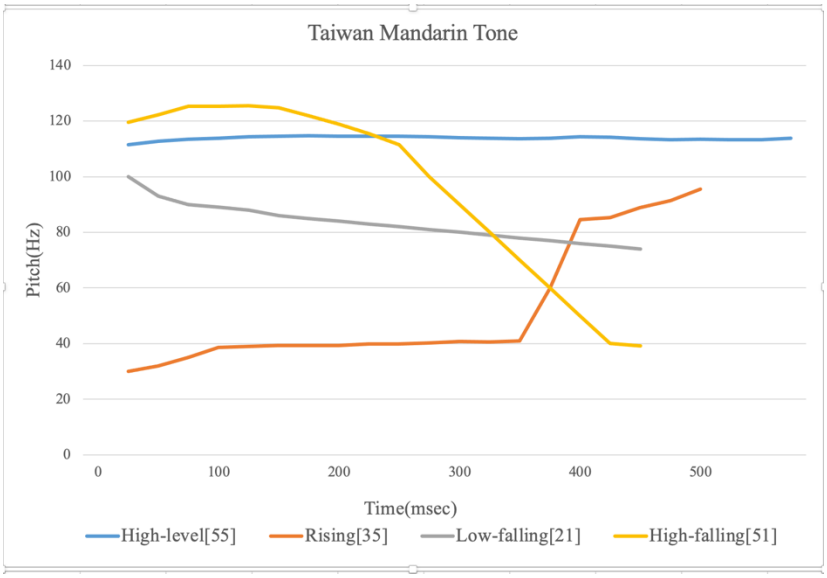
A number of the following studies provided articulatory explanations, markedness, phonetic cues, or theoretical phonological framework to account for universal patterns or language-specific phenomenon on tones. Evidence from the articulatory effort theory lead Ohala and Ewan (1972) and Ohala (1978) to report that rising tone is cross-linguistically longer than falling tones, and needs more energy with muscular control, so rising tone is 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 requires 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, Tseng (1990), Fu and Zeng (2000), and Xu, Tsai and Pfingst (2002) all found that low-falling tone has longer duration than rising tone, followed by high-level tone and/or high-falling tone. In intensity, Ong and Yang (1997), and Xu et al. (2002) agreed that high-falling tone is the strongest in intensity, and high-level tone is the next strongest,

followed by rising tone, and the least strong is the low-falling tone. 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 tone and low-falling tone 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.

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]. Following shows a sample of a male speaker (aged 25) producing citation tones in Taiwan Mandarin.

Figure 1: Mandarin contour tones



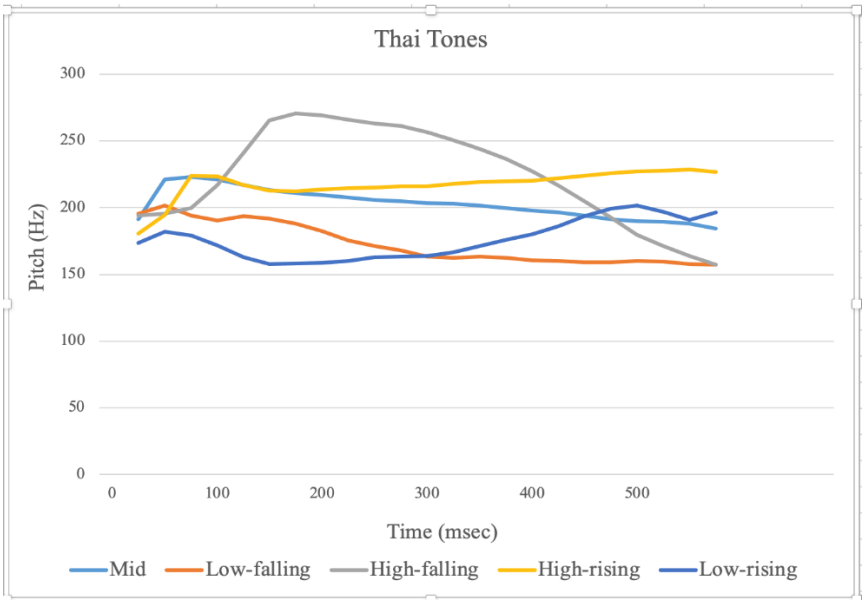
Note that these tone numbers are not intended to indicate underlying sequences, but simply to show the pitches involved in the contour tones. Most syllables in Mandarin carry a tone. However, some light syllables with tone can be produced as unstressed when they are functioning as grammatical particles, or as the second element in a reduplicated kinship term. In these cases, the syllable duration is shortened, and the syllable is spoken with a neutral tone which is phonetically mid-level; that is, it loses its own tone and is pronounced with a mid-level pitch on the surface. In the samples of the dataset, the mean duration is 553.7msec for high-level tone, 404.3msec for rising tone, 321.9msec for low-falling tone and 274.8msec for high-falling tone. As with the occurrence of frequency in the spoken corpora, token counts from a speech corpus (N=604,916), *NCCU Corpus of Spoken Chinese*, showed a rank order in which high-falling tone has the highest frequency of occurrence (39%), followed by low-falling tone (23%), then high-level tone (21%), and finally rising tone, being the least common (17%)[1]. If each tone is expected to occur with approximately equal frequency, the above

[1] The database was collected and provided by Prof. Kawai Chui at National Chengchi University. The site contained four tones plus neutral tone in spontaneous speech. Since the present study focused only

counts should have been distributed evenly. Similarly, in Mandarin acquisition, high-falling tone outnumbered high-level tone, which outnumbers rising tone and low-falling tone (Wan and Yang 2017). Studies from adults and children in general agreed that high-falling tone has greater frequency than rising tone in the world's languages (Li and Thomspon 1977).

There are five tones in Central Standard Thai (e.g., Abramson 1975; Gandour 1979; Hudak 1987; Tingsabadh and Ambrason 1993). The tone system involves mid tone [33], low-falling [21], high-falling [41], high-rising [45], low-rising [14] with citation tones in Central Standard Thai. Figure 2 shows a sample of Thai tones drawn from a female speaker, the source of which is drawn from IPA association.

Figure 2: Thai tones



In this figure, the sample audio file is extracted from IPA (IPA Association 1999) and phonetically measured in Praat (Boersma and Weenink 2019). The duration of each tones is listed in the following: mid tone=550 mesc, low-falling= 575 msec, high-falling=560 msec, high-rising=650 msec, and low-rising=552 msec. There is a large agreement showing the percentage of frequency and the rank order of lexical tones across Thai spoken corpora (Munthuli et al. 2015). All the corpora have agreed that mid tone occurs with the highest frequency, followed by low tone and falling tone. Mid and low tones amounted to more than 50% of all tones. Therefore, compared with Mandarin spoken corpus, Thai speakers showed different preference towards tone patterns. High-falling and high-level tones occurred more often in Mandarin whereas mid and low tones were the most common in Thai.

In addition to frequency counts in the spoken corpora, evidence from language acquisition studies recently has shown the preference for disyllables in relation to the universal tendency in favor of binary foot structure cross-linguistically (e.g., Ota 1999; Demuth and Johnson 2003; Demuth 2006; Demuth and Tremblay 2008; Miyakoda 2012). Wan and Yang (2017) suggested that Mandarin children, caretakers and/or

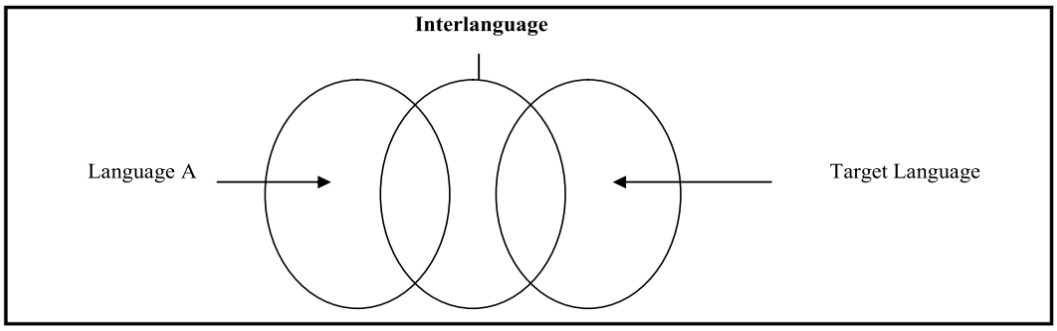
on the four tones, the token and type for the neutral tone were excluded for the discussion.

adults are more likely to utter in disyllables. However, most of the relevant studies on language learning/acquisition have covered the production/perception on monosyllabic tones only. Citation tones in Mandarin are nearly always discussed in isolated words, and only when tone sandhi processes were involved would more than two syllables in sequences be discussed. Recently, the preference for disyllables in language acquisition has shed light on binary foot structure in cross-linguistic studies (e.g., Ota 1999; Demuth and Johnson 2003; Demuth 2006; Demuth and Tremblay 2008; Miyakoda 2012), suggesting a universal tendency.

Issues with regard to how first language influences or interferes second language have drawn many scholars' attention. Instead of focusing on the error patterns as in the traditional analysis, Selinker (1972) suggested that interlanguage can be viewed as the transitional processes between learner's mother tongue and the target language. Brown (1994) and Fauziati (2011) both regarded the learner's system that includes a structural status overlapping with the mother tongue and target language. This fact reflected that the learner is constructing an independent linguistic system progressively and gradually towards the target language system. Since many studies have shown that learners' errors are not fully related to their mother tongue or target languages (e.g., Bailey et al. 1974; Krashen et al. 1978, 2003; Larsen-Freeman 2002; 2003), interlanguage theory is to describe learners' errors, and is viewed as in the process of acquiring the target language.

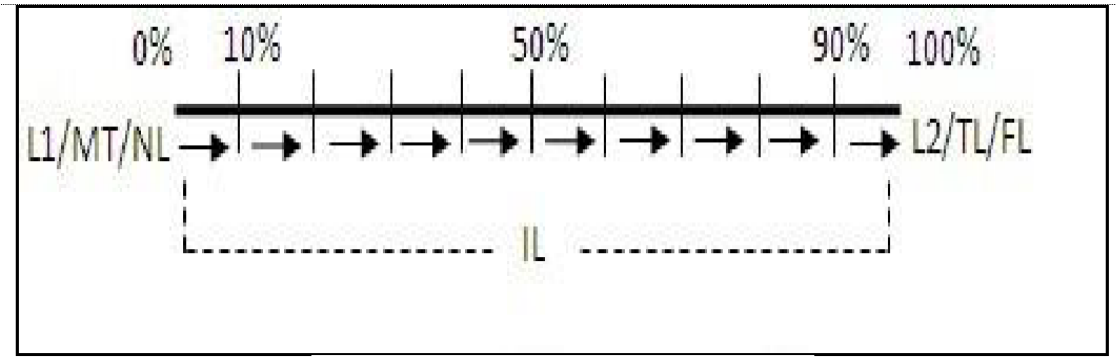
Corder (1981) and Al-Khresheh (2015) suggested the term, 'interlanguage continuum,' in which the learning process continues to move on; learners tend to reconstruct the rules by adding, deleting or shifting linguistic units until the target language system is fully shaped. The figure and illustration are provided below.

Figure 3: The notion of interlanguage (Corder, 1981)



In this figure, 'Language A' represents the learner's mother tongue, and 'Target Language' is the language the learner is learning. There is overlapping from the learner's language system to transfer some linguistic knowledge towards the target language.

Figure 4: Interlanguage Continuum



In this figure when the learner is mapping the input language system to the target language (including second or global language), there will be a transition scale for the learner to be closer gradually to produce the correct input.

With the input language system mapping to the target language (including second or global language), the learner is getting close to produce the correct output.

In the current study, data are drawn from 11 pre-school learners in the intensive summer school at Thai-Chinese International School in Bangkok, Thailand. The data were collected during a three-hour intensive learning session (including the break/snack time). Each class has three sessions, involving 60-min grammatical lecture notes, 60-min Question-Answering and 30-min repetition sessions. During the first session with grammatical lecture notes, the instructor used picture cards or toys to introduce the Chinese grammar, and the children thus had a chance speaking, the voices were overlapped so the source of the multiple learners would not be easily identified (Note that speech diarization is a new cutting-edge technique and is not the main focus in the study). Each child has a chance to talk to the instructor individually for the last two sessions and that was how the recording started to take place. Note that the lexicons or utterances in the question-answering and repetition sessions are drawn from the textbook materials in a naturalistic setting. Tone patterns in the 'repetition' and 'Q-A questions' were both lumped together for discussion since such pre-school children did not make too many changes deviant from the two types; they barely produced a complete sentence when they uttered in Mandarin. The total recorded hours are 63 hours, and the validated hours structured in the phone-aligned Praat system with 17 tiers are 1.5 hours at the current stage[2].

In general, the learners have difficulty in learning a tone language as their global language especially when their mother tongue languages are stress or pitch-accented languages ; however, certain learners whose mother tongue is a tone language will not produce many tone errors (e.g., Leather 1990; Wang 1995; Tseng 1997; Chen 1997;

[2] One might find that 1.5 hours could not provide sufficient data to support the aforementioned relevant issues; however, the released spoken corpus at Academia Sinica contains 3.5-hour speech data with verified phone boundaries (Sinica Phone-aligned Chinese Conversational Speech Database) and 8-hour speech data in segmented speaking turns (Sinica MCDC8). In this preliminary study, the data are presented to show how Thai pre-school learners produce Mandarin tones, and thus the amount has not necessarily reached a level of the production quality for machine learning paradigms for speech recognition.

Miracle 1999, Lin 2007). Therefore, the aim of this paper is to present a preliminary analysis by providing a detailed corpus analysis of interlanguage tonal distribution and patterning. An observational data collection was drawn by 11 pre-school children in Thailand learning Mandarin as their global language at Thai-Chinese International School in Bangkok, Thailand. The novelty of this work is to collect and analyze interlanguage tone patterns through a highly reliable corpus, and to provide Praat acoustic parameters for the assessed error distribution by looking at the frequency and the various patterns involving tone units in Chinese lexicon. These units involved in the errors are classified by segmenting word shapes, based on the number of syllable, into monosyllabic, disyllabic or multi-syllabic. These children's tone patterning along with the instructor's speech were systematically measured by plotting the F_0 variation when they were asked to repeat or respond questions from their Chinese instructor in Mandarin. Evidence from the data would be able to show how Thai children utilize their own linguistic knowledge in learning Mandarin, and how Mandarin four tones interacting with one another.

This paper is organized as follows. The following section includes the methodology for the data collection and the workflow annotating the dataset. Section three presents the preliminary analysis in terms of some theoretical framework. The final section will summarize the study, and will suggest some plans for the future study.

2. Methodology

The entire data collection session containing 63 hours was collected by my research team. Data collection took place between June 11 and June 30, 2018, from an intensive summer Chinese class for pre-school beginners at Thai-Chinese International School, Bangkok, Thailand. Under the study, meaningful words were classified based on the criteria proposed by Vihman (1996), Vihman and McCune (1994), Zhu (2002), and Wan and Yang's (2017) work 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 matches the adult target; 3) when children produced the words more than twice with similar phonological shapes across different uses. For example, in [twej51] 'correct', [tɕ^{h}jow35] 'ball', and [fan51] 'rice', many children would produce them as in [tej51], [t^{h}jow35], and [fa51]. Since more than two segments match the adult form, examples like these that appear more than twice in different contexts are considered meaningful words. However, if the tone is not recognizable, they are classified as non-meaningful words. The current dataset in the study is selected from a corpus accumulated to 63 hours for pre-school Thai children (N=11, 4;1-6;5), and 516 hours for elementary school children (N=13, 7;00-12; 08), and 388 hours for high school students (N=14, 13;03-15;11).

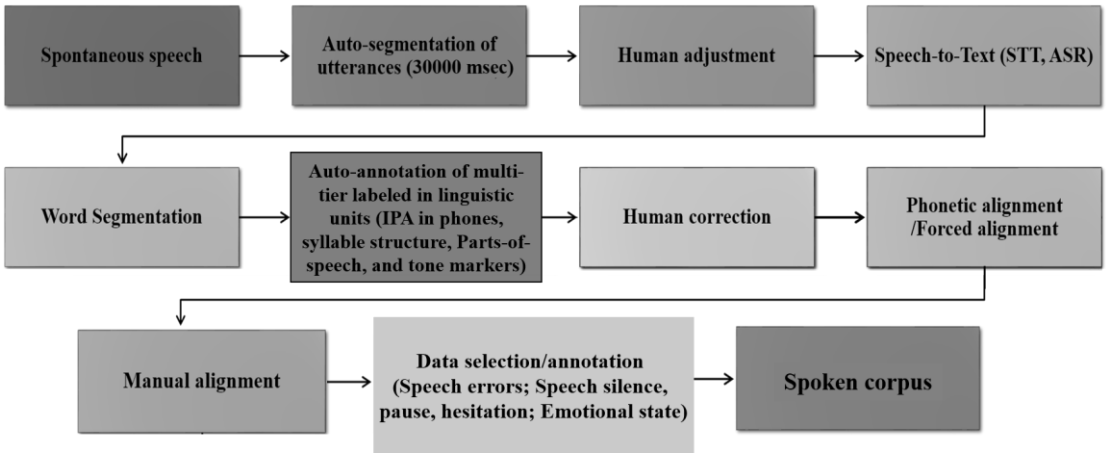
Table 1: Participants

Participants	Gender	Age	Language background
1	M	5	Thai
2	M	6	Thai
3	M	6	Thai
4	M	6	Thai
5	M	6	Thai
6	M	5	Thai
7	M	6	Thai
8	F	5	Thai
9	F	4	French/English/Thai
10	M	5	Thai
11	F	6	Thai

To gain information concerning the tone contour, eleven children (8 boys, 3 girls), who learn Mandarin as their second or third language, participated in the study. The 11 participants were all healthy and had not detected with any hearing or intellectual impairment, and were all willing to utter in Mandarin with the instructor. All sessions were audio-recorded in Zoom (H6 Handy Recorder) with high quality sound files for audio signals in 16-bit 44100 Hz sampling rate.

I have constructed spoken corpora in Mandarin from normal and aphasic adult speakers as well as from normally developing and phonologically disordered children for more than 25 years; the spoken corpora were listed in the appendix. For normal adult speakers, due to the recent improvement of state-of-the-art speech recognition with sequence-to-sequence models, the spoken files are semi-automatically transcribed and annotated. The workflow in processing each sound file is provided as follows.

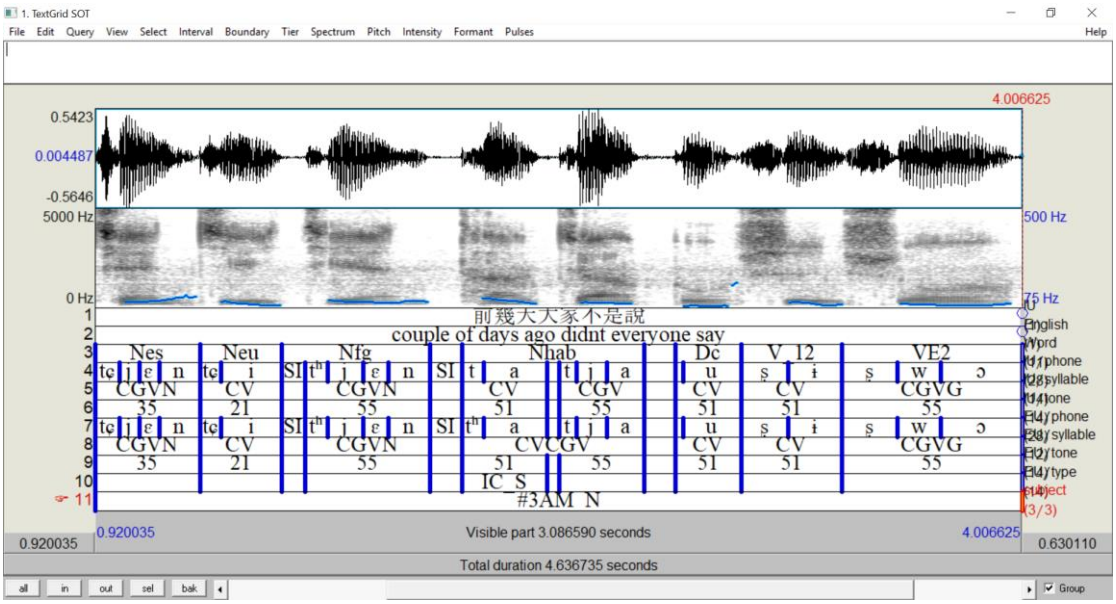
Figure 5: A workflow for normal Mandarin adult speakers



Each audio file involving the spontaneous speech between multi-speakers was pre-processed into auto-segmented frames of 30000 msec at the sampling rate of 16-bit 44100Hz. To avoid a phone or utterance cut in half across different files, the assistant double-checked the beginning and the end points of the frames to make sure that each

frame contained complete linguistic units in a fully contextual utterance. All the obtained frames were first sent to a Speech-to-Text (STT) system for transcription[3]. The output of the STT system was then verified by comparing with an assistant's manual transcription[4]. The entire transcript is then automatically segmented by the CKIP parser, Chinese Knowledge and Information Processing, 2004 (Ma and Chen 2003).[5] The audio file and transcription were processed in Praat with automatically forced alignment to phone level with the accuracy of 85 percent[6], followed by manually adjusting on the alignment of the phones in each utterance. Further editing and correction are made by two research assistants. The boundaries of each phone are then checked manually by research assistants. All the information of Chinese characters, parts of speech, and phonetic units in IPA involving consonants, vowels, syllables and tones are automatically generated in each tier constructed by a hidden phoneme dictionary. The sample of a Mandarin adult speech data is provided as follows.

Figure 6: A sample of time-aligned multi-tier spontaneous speech in Mandarin



In the entire corpus, the first tier showed the utterance in Chinese characters, automatically dragging the English translation in the second tier; parts-of-speech, the third tier; phonetic units in IPA, the fourth tier; syllable structure, the fifth tier, and tone units, the sixth tier. Two well-trained assistants will need to double- or triple-check

[3] The audio files were automatically transcribed in Speech-to-Text systems powered by AI Labs in Taiwan or Google API depending on the recording quality.

[4] The average error rate of the STT systems was 30%. Main influencing factors were the recording environment and the voice quality of the speakers.

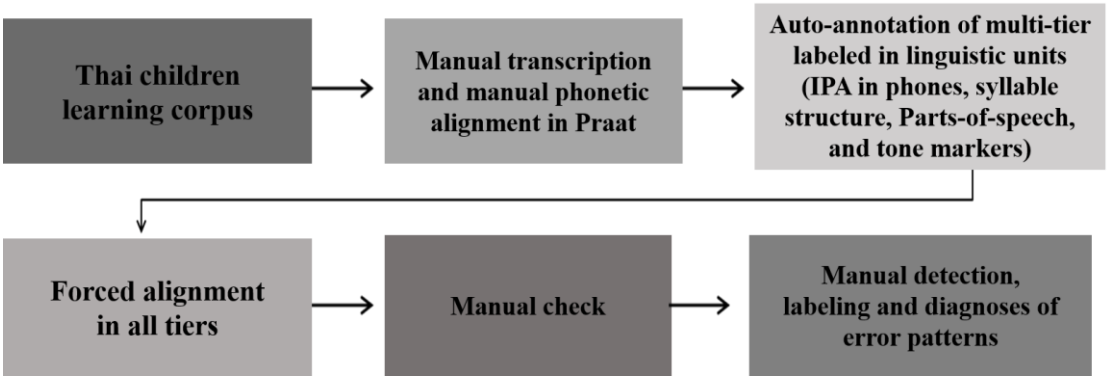
[5] I appreciate to get the open source code from Professor Wei-yun Ma at Academia Sinica before the coding is released to the public.

[6] The current forced aligner is originally based on traditional method HMM with a phoneme dictionary and further refined by DNN models, the system of which is hoping to handle spontaneous speech in challenging noisy environment and to reduce the cost of annotating labeled datasets for trained assistants. I appreciate the original source software from Professor Li-hsin Ning and Professor Jiahong Yuan, and my deepest appreciation goes to Dr. Chain-wu Lee for training the accuracy rate amounted to higher than 85%.

whether the speech errors exist in Tier 7 through Tier 9, and mark the type of errors patterning in the tenth tier if possible. The final tier stated the emotional status of the speaker. Therefore, in this sample, the speaker #3, male, said the utterance [tɕʰjɛn35 tɕi21 **t̲ʰ**jɛn55 **t̲**a51 tɕja55 pu35 ʂi51 ʂwɔ55] 'couple of days ago didn't everyone said," and erroneously said the utterance [tɕʰjɛn35 tɕi21 **t̲ʰ**jɛn55 **t̲ʰ**a51 tɕja55 pu35 ʂi51 ʂwɔ55] in a neutral state of emotion. In this case of initial single-consonant substitution errors, the consonant [tʰ] is perseverated and substituted for the following consonant [t], rendering a meaningless utterance.

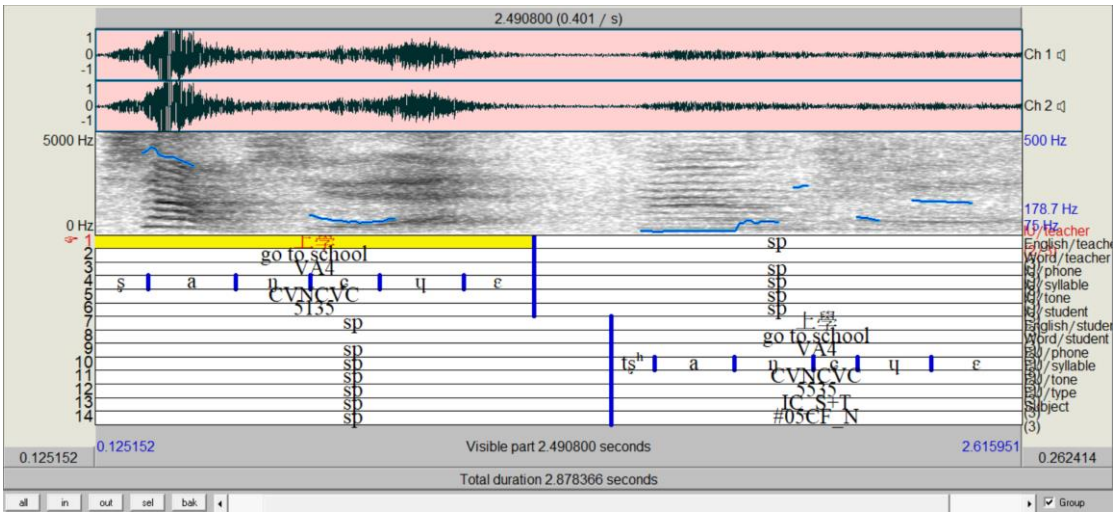
Unfortunately, the current spoken corpus does not yield the same workflow since all the participants in the study are nonnative speakers of Mandarin, and they are pre-school children; the STT system does not successfully transcribe the speech data to the corresponding text automatically. Therefore, a different method is utilized, and the workflow is provided as follows. The automatic speech recognition will need certain quality time to improve in the near future.

Figure 7: A workflow for Thai children learning Mandarin



Each audio file involves the lexicon extracted from the read speech or spontaneous speech from the 11 participants. The data were pre-processed into auto-segmented frames of 30000 msec at the sampling rate of 16-bit 44100Hz. To avoid a phone or utterance cut in half across different files, the assistant double-checked the beginning and the end points of the frames to make sure that each frame contained complete linguistic units in a fully contextual utterance. All the obtained frames sent to Praat were annotated and verified by two linguists manually. The first tier of the phonetic alignment was manually segmented, and the following corresponding tiers, with the multi-tier labeled in different linguistic units and intervals, will be automatically generated.

Figure 8: A sample of time-aligned multi-tier speech dataset involving repetition in Thai pre-school child learning Mandarin



In the data, the first and seventh tiers are manually transcribed and aligned by the assistants; however, the rest of the information involving parts-of-speech, phonetic units in IPA, syllable structure and tone markers in those tiers is automatically generated by the system; 'sp' stands for speech pauses or speech silence in the dataset. The teacher said [**ʂ**aŋ**51** ɕɥɛ35] 'go to school' and the imitative response to a verbal stimulus in a neutral state of emotion from the participant #5, a girl, is [**tʂ^{h}**aŋ**55** ɕɥɛ35]; in this case, the initial target consonant is mispronounced as a [tʂ^{h}] with a wrong tone [55]. In this case, the teacher produced louder in volume than the pre-school child since she was timid repeating the Chinese word. Note that 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.

The usual methodology for collecting tone data is to rely on the transcribers' 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 the studies of Wong, Schwartz and Jenkins (2005) and Wong (2008, 2012, 2013). 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[7]. As stated in Demuth (2006), the challenge with the field of phonological acquisition or learning has been the lack of longitudinal phonetically transcribed data from multiple children, let alone that these pre-school children were just beginning to learn Mandarin. Therefore, the data to be discussed below is hoping to provide a preliminary analysis showing how Thai pre-school children transfer their first tone

[7] The trained transcribers and I have transcribed the data from adults' speech-error corpus, aphasic adult speakers' corpus, children's corpora including normally developing children and phonologically disordered children, from both spontaneous and elicited production experiments. I have collected those data for 25 years, and when we transcribe the data, lexical information is not the feature we rely on. Especially, the majority of the data in adults' speech errors or in Chinese aphasic patients do not carry any lexical information, and many of them are simply related to sound errors, which utter meaningless forms.

system to another tone language.

3. Results and Discussion

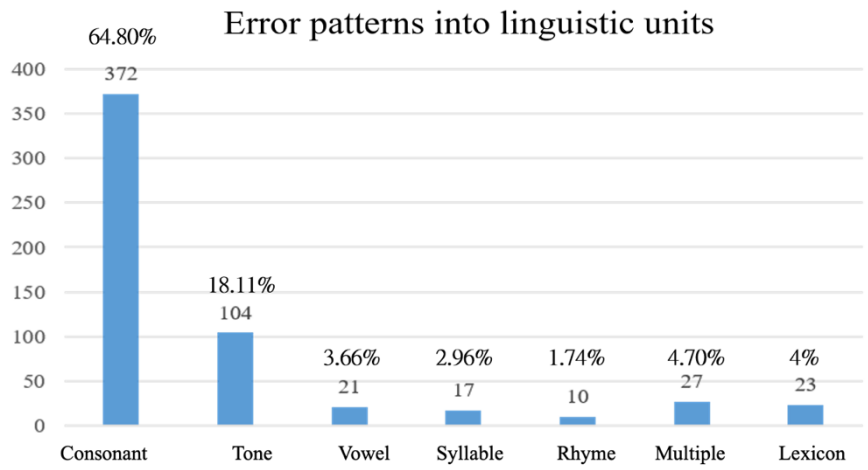
In the study, for a 90-min length of validated dataset, it yields 1234 Chinese lexicon with 777 meaningful lexical items (63%) whereas 457 lexical items (37%) are not able to be identified, thus rendering meaningless output. For instance, low-rising tone [kʰa14] in Thai means 'legs', and a child consistently pronounced [tɕʰa51] for [tɕjaw21] 'legs' in Mandarin, which render meaningless form. When there is a phonetic distance between the target language and mother tongue, I decide to consider 'meaningful words' only in this study. Of 777 words, 271 words are monosyllables; 451, disyllables; 20, tri-syllables; 3, four-syllables, and 32 words in neutral tone syllables. Among the dataset, the following shows the errors categorized into each linguistic unit.

Table 2: Error patterns into linguistic units

Consonant	Tone	Vowel	Syllable	Rhyme	Multiple	Lexicon	Total
372 (64.8%)	104 (18.11%)	21 (3.66%)	17 (2.96%)	10 (1.74%)	27 (4.7%)	23 (4%)	574

In the table and the following figure, consonant errors evidently occur more often than the other types, and Thai children seem not to produce many tone errors. This finding has been confirmed in the Thai adults' learning data (Chen, 2011).

Figure 9: Error patterns into linguistic units



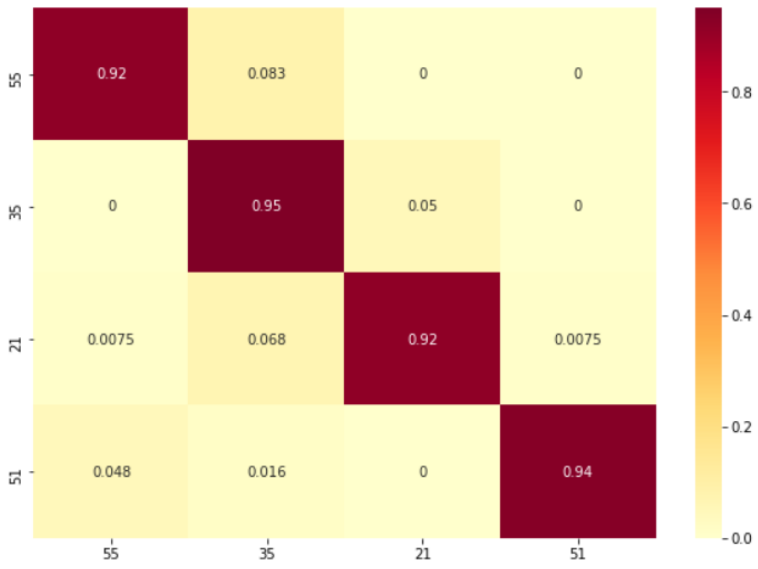
The following table and confusion matrix summarize the distribution of tone patterns between target language (TL) and mother tongue (MT) in monosyllables, disyllables and multi-syllables from the 11 pre-school Thai children's production system. The following tones refer to the four-tone system in Mandarin. For those tones produced in Thai tones are not the focus in this paper.

Table 3: TL-MT pair in Mandarin monosyllabic tones

TL \ IT	55	35	21	51
55	25	2	0	0
35	0	21	1	0
21	1	10	138	1
51	4	1	0	67

In this table, of 271 tokens, 25 tokens in high-level tone are produced correctly, 21 tokens, rising tone; 138 tokens, low-falling tone and 67 tokens in high-falling tone. Wilcoxon test of distinct and shared tones between target and responded monosyllabic words using a 95% confidence interval shows a statistically significant difference ($p<.005$). In high-level tone, 2 tokens are mistakenly produced to rising tone, and in rising tone, 1 token produced to low-falling. Low-falling and high-falling tones have wider variabilities; in low-falling tone, 10 tokens produced to rising tone, and 1 token to either high-level or high-falling tone. In high-falling tone, 4 tokens were produced to high-level tone and 1 token to rising tone. The 11 pre-school Thai children showed that low-falling tone is the most common, followed by high-falling tone, high-level tone is the third common, and rising tone is the least common. The main reason is these Thai children are just beginning to learn Mandarin, and the most common lexicon is [wɔ21] 'I' in their utterance. So the occurrence of frequency from the database is generally related to the input verbal stimulus given by the instructor. In addition, all those erroneously produced tones are true Mandarin tones, showing that Thai learners do not transfer their mother tongue to the target language. The following confusion matrix shows the accuracy rate.

Figure 10: Confusion matrix in monosyllables



In this matrix, it shows that Thai pre-school children have no problem producing monosyllabic tones in Mandarin since all four tones amount to 90 percent accurate with the rank order of the following: Rising > High-falling > High-level = Low-falling. The

error rate in other target-error pairs is relatively low.

Table 4: TL-MT pair in Mandarin disyllabic tones

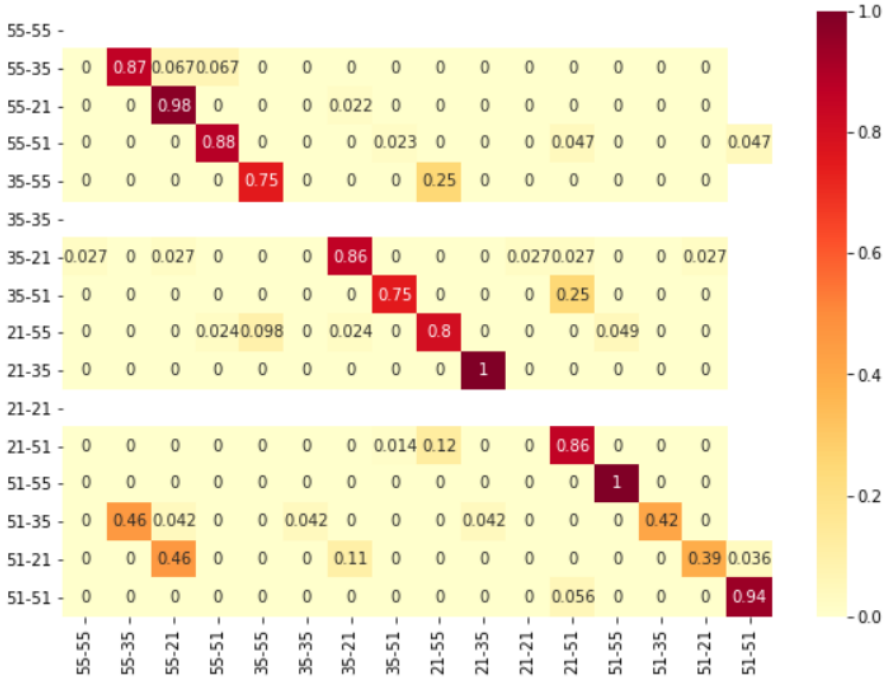
Error / Target	55-55	55-35	55-21	55-51	35-55	35-35	35-21	35-51
55-55								
55-35		17	1	1				
55-21			58				1	
55-51				50				1
35-55					8			
35-35								
35-21	1		1				42	
35-51								4
21-55				1	5		1	
21-35								
21-21								
21-51								1
51-55								
51-35		11	1			1		

Error / Target	21-55	21-35	21-21	21-51	51-55	51-35	51-21	51-51
55-55								
55-35								
55-21								
55-51				3				3
35-55	3							
35-35								
35-21			1	1			1	
35-51				1				
21-55	43				3			
21-35		9						
21-21								
21-51	12			81				
51-55					3			
51-35		1				13		

At this current stage, in this table, it shows that the pattern of [21-51] (N=81) are produced more common, and the following order is [55-21] (N=58), [55-51] (N=50),

[21-55] (N=43), and [35-21] (N=42). Wilcoxon test of distinct and shared tones between target and responded disyllabic words using a 95% confidence interval shows a statistically significant difference (p<. 005). The 11 pre-school Thai children show that tonal pattern involving the combination of falling tones, low-falling + high-falling are the most common, followed by the combination of high-level + low-falling, the one of high level+ high-falling is the third common, and so far those participants have not yet been able to produce all the 15 combinations. Again, the frequency of the tone combination simply reflects the patterns in the instructor's given lexicon. The following confusion matrix shows the accuracy rate of each combination. Note that the accuracy rate of disyllabic tones is comparatively lower than that of monosyllabic tones.

Figure 11: Confusion matrix in disyllables



If one looks at the matrix alone, it shows that the accuracy rate of rank order higher than 80 percent is the following: [21-35] (100%, N=9) = [51-55] (100%, N=3) > [55-21] (98%, N=58) > [51-51] (94%, N=22) > [55-51] (88%, N=50) > [55-35] (87%, N=17) > [35-21] (86%, N=42) =[21-51] (86%, N=81) > [21-55] (80%, N=43). However, if one compares the token frequency with the accuracy rate, it might explain the difference in that the tone pattern [21-51] is the most common and the accuracy rate is 86 percent whereas there are only 2 tokens of tone pattern [51-55], but the accuracy rate reaches 100%. Therefore, if one combines the token frequency with the accuracy rate, the data show the following information: [21-51]: N=81, 86%; [55-21]: N=58, 98%; [55-51]: N=50, 88%; [21-55]: N=43, 86%; [35-21]: N=42, 86%. Since the number is not sufficient to run a solid statistical analysis, the figure might only suggest that Thai children are more familiar with certain tone patterns, and possibly tone combination involving the falling contour is their preference. Note that all those patterns of tone combinations are given evenly from the classroom instructors.

Table 5: TL-IT pair in Mandarin tri-syllabic tones

TL \ IT	55-55-51	55-51-51	51-51-21	51-35-51	51-21-55	51-21-35	51-21-51
55-55-51		2					
55-51-51		3					
51-51-21			2				
51-35-51				4			
51-21-55					2		
51-21-35						5	
51-21-51				2			

There is not sufficient data in the tri-syllabic tones and it is thus not able to make a plausible generalization. Note that the instructor in class has inputted more lexicon in disyllables, followed by monosyllables, and the number of tri-syllables and multi-syllables will still need more time to accumulate.

The fact that these Thai children prefer to produce disyllables might have largely been influenced by the instructor's input tokens, and this is similar to the study of first language acquisition. Wan and Yang (2017) found that Mandarin children in Taiwan are more likely to produce more disyllabic lexicons due to the fact that the caretakers have greater input lexicons on disyllables; in the children's tokens, disyllables (N=6,172, 78%) occur more than three times than monosyllables (N=1,729, 22%), and in the caretakers' speech as well as in linguistic environment, the speech corpus (N=604,916), *NCCU Corpus of Spoken Chinese,* showed that Taiwan Mandarin speakers have a preference to produce more disyllables (N=368,997, 61%) than tri-syllables (N=96,789, 16%), monosyllables (N=90,738, 15%), and multi-syllables (N=48,392, 8%). Therefore, data from Thai and Mandarin acquisition might confirm the universal preference in cross-linguistic studies for binary foot structure (e.g., Ota 1999; Demuth and Johnson 2003; Demuth 2006; Demuth and Tremblay 2008; Miyakoda 2012).

In the further study when the validated hours reach 20 hours, which is estimated to yield 25000 tokens, a different method suggested by Wong, Schwartz and Jenkins (2005) and Wong (2008, 2012, 2013) must be conducted. The usual methodology for collecting tone data is to rely on the transcriber's intuitions as to the categorization of the tone in the spoken utterance with lexical information. However, this methodology is subject to the transcriber's bias. The following example shows a sample of the entire pitch range from the start to end points in a disyllabic word: [ʂi35 ər51] 'twelve'.

Figure 12a: A sample of pitch listing the disyllabic tone movement

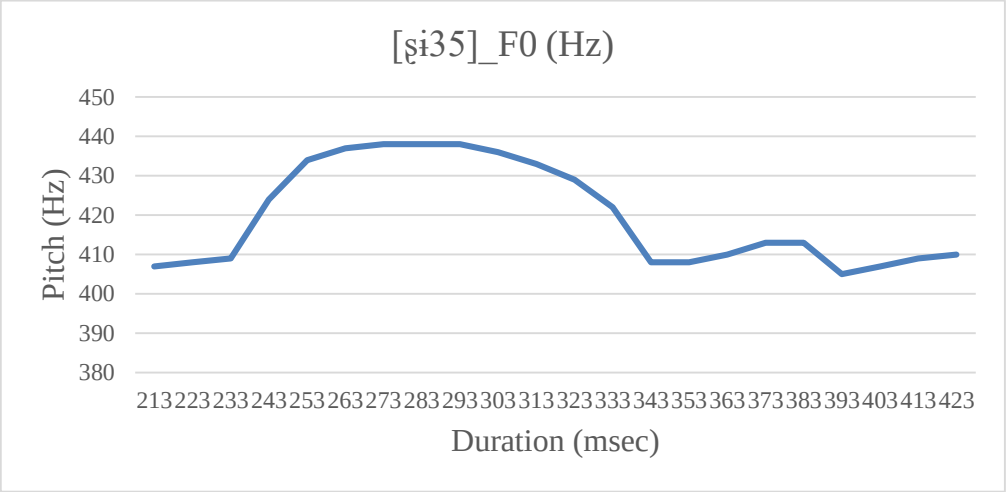
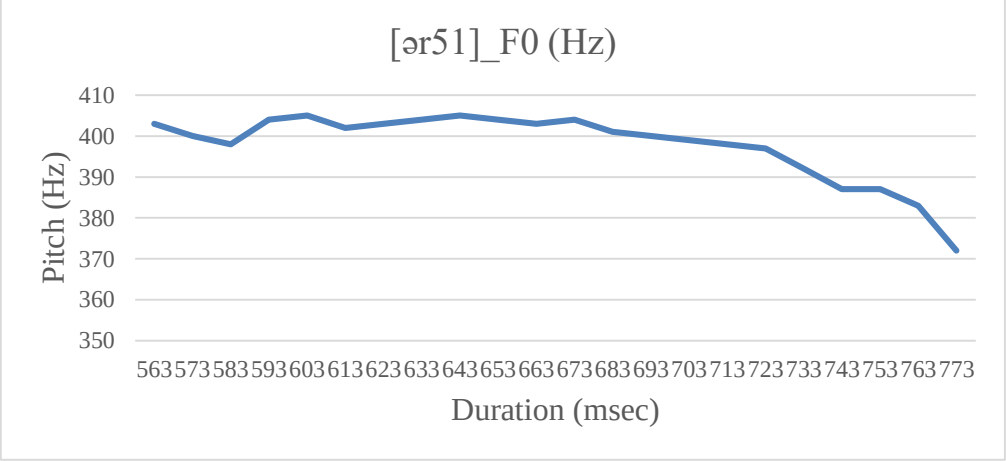
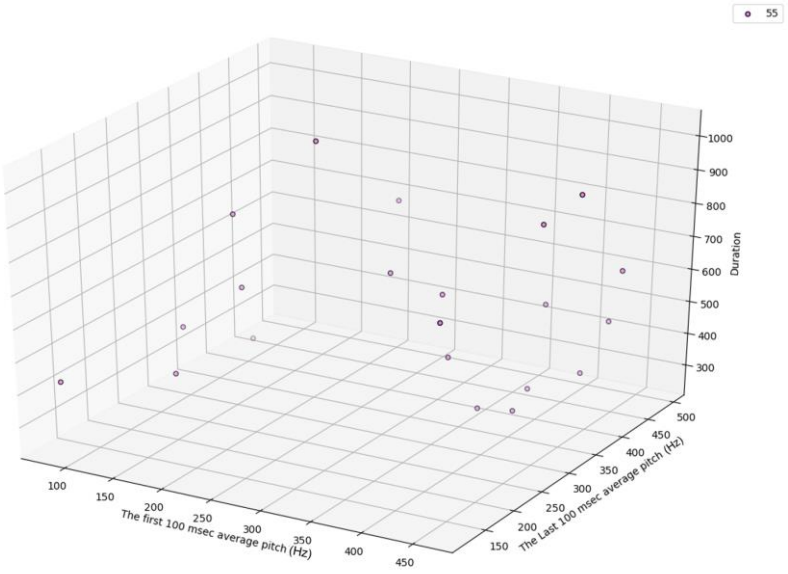


Figure 12b: A sample of pitch listing the disyllabic tone movement



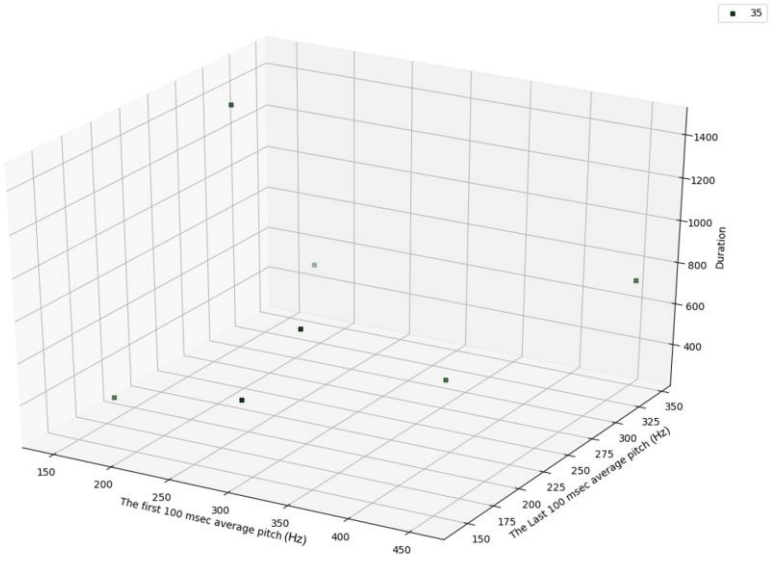
In these two figures, it shows the tone pitch contour ranging in every 10 msec of the disyllabic word. The following four three-dimension figures list the first and last 100 msec mean pitch with the duration of tones in Mandarin monosyllables from the dataset (x= the first-100 msec mean pitch; y = the last-100 msec mean pitch; z = the duration of tone).

Figure 13: Three dimension plots of high-level tone in categorical perception



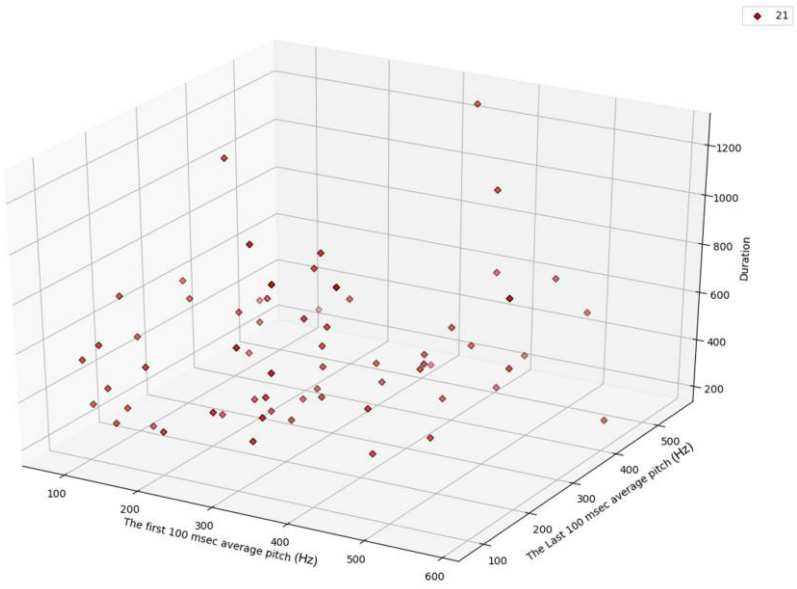
In this figure, 25 tokens in high-level tone are produced by Thai pre-school children, and perceived in the 'correct' category judged from the transcribers' evaluation. Among the tokens, the duration in length ranges from 300msec to 800msec, the first 100-msec mean pitch ranges from less than 100 Hz to 425 Hz, and the last 100-mesc mean pitch, from 160 Hz to 450 Hz, meaning that high-level tone from the start to end points could be within a possible range such as 100 Hz through 160 Hz, or from 425 Hz through 450 Hz, but within a wider range stretching between 100 Hz and 450 Hz, or between 160 Hz and 425 Hz. If one looks at the distribution in the three-dimensional plots, it is rather difficult to capture a generalization in which the scattered dots all point to the high-level tone category. The similar finding is shown in the following plots.

Figure 14: Three dimension plots of rising tone in categorical perception (x=the first 100 msec average pitch in Hz, y=the last 100 msec average pitch in Hz, z=duration in msec)



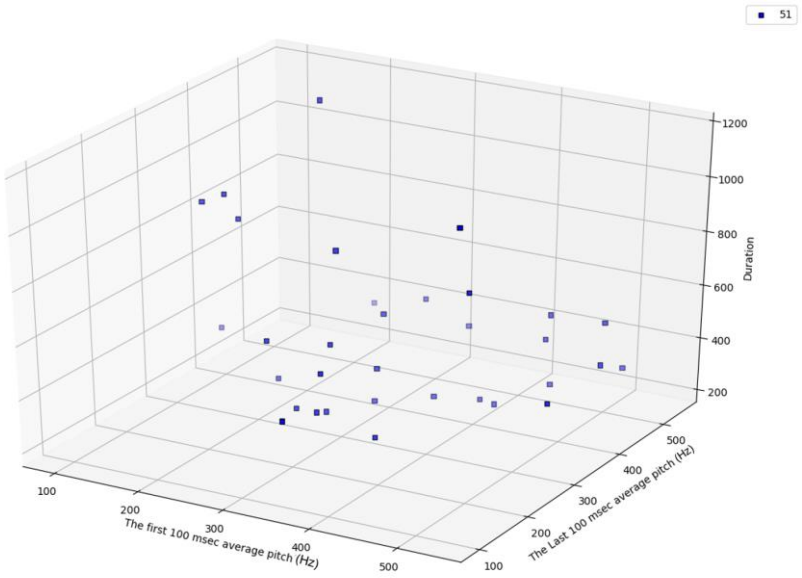
In this figure, 21 tokens in rising tone are produced by Thai pre-school children and perceived in the 'correct' category judged from the transcribers' evaluation. Among the tokens, the duration in length ranges from 400 msec to 1300 msec, the first 100-msec mean pitch ranges from less than 20 Hz to 340 Hz, and the last 100-mesc mean pitch, from 210 Hz to 325 Hz, meaning that rising tone from the start to end points could be within a possible range such as 100 Hz through 160 Hz, or from 425 Hz through 450 Hz, but within a wider range stretching between 100 Hz and 450 Hz, or between 160 Hz and 425 Hz. Since the tokens of rising tone are relatively rare, and the dots of rising tone are scattered in the three-dimensional plots, it is insufficient to make a fine judgment.

Figure 15: Three dimension plots of low-falling tone in categorical perception (x=the first 100 msec average pitch in Hz, y=the last 100 msec average pitch in Hz, z=duration in msec)



In this figure, 138 tokens in low-falling tone are evidently produced the most often by Thai pre-school children. The 'correct' category judged from the transcribers' evaluation shows that the duration in length ranges from 120 msec to 1200 msec, the first 100-msec mean pitch ranges from less than 40 Hz to 580 Hz, and the last 100-mesc mean pitch, from 200 Hz to 580 Hz. Judged by the duration and the mean pitch of the start and end points, it is impossible to draw a fine line of categorizing the best low-falling-tone exemplars.

Figure 16: Three dimension plots of high-falling tone in categorical perception (x=the first 100msec average pitch in Hz, y=the last 100msec average pitch in Hz, z=duration in msec)



In this figure, 67 tokens in high-falling tone are produced by Thai pre-school children and perceived in the 'correct' category judged from the transcribers' evaluation. Among the tokens, the duration in length ranges from 200 msec to 1050 msec, the first 100-msec mean pitch ranges from 80 Hz to 450 Hz, and the last 100-mesc mean pitch, from 300 Hz to 550 Hz, showing a wider range of distribution, so it is still insufficient to make a fine boundary in the scattered pattern of three-dimensional plots.

One might wonder the pitch range seems to be wider in the Thai children learning Mandarin; however, research from Keating and Bhur (1978) had long showed that some participants evidence remarkable ranges of F0 values with utterance as low as 30 Hz in F0 to as high as 1500Hz with American children. The following table shows the information of children's age, gender, range of F0 and type of utterances.

Table 6: Ranges of F0 values in American children (Keating and Bhur 1978)

Child	Sex	Age (weeks)	Range of F0 (Hz)	Type of utterance
L.S.	M	33	40-850	Babbling
		66	90-1150	Babbling
J.S.	M	67	250-2500	Babbling and Words
F.R.	M	125	66-750	Sentences
		169	30-1100	Sentences
G.R.	M	69	60-850	Words
		128	120-1050	Sentences
J.B.	F	66	30-1500	Words
		109	50-1150	Sentences
R.C.	F	33	80-500	Babbling
		69	60-900	Babbling

These four three-dimensional scattered patterns might confirm the findings in Wong, Schwartz and Jenkins (2005) and Wong (2008, 2012, 2013). The transcribers seem to categorically rely on their lexical information to judge whether the tone errors exist; namely those transcribers largely relied on the lexical understanding to help perceive better in tone languages. Therefore, whenever they need to make a decision to see whether those lexical items are understandable, their judgments on good or bad exemplars might have been influenced as to whether or not they understand the spoken utterances from those Thai children. In the next study, a phonetically perceptual task in a more rigorous experimental control in a low-pass filter is required for those transcribers to eliminate lexical information but only reserve tone information from those spoken tokens.

4. Conclusion

Based on the preliminary analysis, the data might not be able to make a robust claim on the tone error patterns; however, the 11 pre-school Thai children do show interlanguage tonal distribution, and they seem to have no difficulty producing Mandarin tones since they have more complicated tone contours in their mother tongue, and Thai is also a tone language. In addition, the children are more likely to utter with disyllabic tones, but disyllables are not the main syllable structure in Thai. This shows the preference for disyllables in relation to the universal tendency in favor of binary foot structure cross-linguistically. Since those Thai children uttered [wɔ21] 'I' most often, low-falling tone outnumbers the rest of the tones. When more validated hours of data are confirmed, it is necessary to calculate the type of the data instead of the tokens of them.

Different from the traditional analysis on the error patterns, this study has suggested that interlanguage can be viewed as the transitional process between Thai and Mandarin, and those tone error patterns are not fully related to their mother tongue, possibly suggesting that Thai children might have constructed an independent linguistic system gradually towards Mandarin. The study has provided a preliminary analysis through a highly reliable corpus, and provided Praat acoustic parameters for the tone distribution by looking at various tone combinations measured by plotting the F0 variations in Mandarin. In the near future, a phonetically perceptual task in a low-pass filter for reserving tone information will need to be conducted in the hope of eliminating the transcribers' bias to rely on the lexical information.

Acknowledgment:

I am very grateful for the helpful insights and constructive suggestions made by the anonymous reviewers. I appreciate Tzu-yi Tseng and Anthony Lin for their labeling and annotation on this project. Thanks go to Professor Wei-yun Ma for the coding of CKIP parser, and Professor Li-hsin Ning and Professor Jiahong Yuan for the traditional HMM coding for the phonetic forced alignment in Mandarin. I am also grateful for obtaining free access to the NCCU corpus of spoken Chinese provided by Professor Kawai Chui. Special thanks go to Dr. Chain-wu Lee for his continuous cutting-edge high-tech programming support in constructing the entire corpora in the Phonetics and

Psycholinguistics Lab at National Chengchi University. All remaining errors of analysis or interpretation are my own. This research was funded to the author by MOE New-South-Bound policy (MOE 107F107016) and MOST grants (MOST 107-2410-H-004-127-, MOST 108-2410-H-004 -098 -MY2) to the author in Taiwan.

References

Abramson, Arthur S. 1975. Thai tones as a reference system. *Haskins Laboratories: Status Report on Speech Research SR* 44:127-136.

Al-Khresheh, Mohammad Hamad. 2015. A review study of interlanguage theory. *International Journal of Applied Linguistics and English Literature* 4.3:123-131.

Bailey, Nathalie, Carolyn Madden, and Stephen D. Krashen. 1974. Is there a "natural sequence" in adult second language learning? *Language Learning* 21.2:235-243.

Boersma, Paul and Weenink David. 2019. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.53, retrieved from http://www.praat.org/

Brown, H. Douglas. 1994. *Principles of Language Learning and Teaching* (3rd ed.). New Jersey: Prentice Hall Regents.

Chao, Yuen Ren. 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, Yuen Ren. 1968. *A grammar of spoken Chinese.* Berkeley and Los Angeles: University of California Press.

Chen, Chun Mei. 2011. Effects of corrective feedback in the prosodic development of L2 learners. *The Asian Conference on Language Learning 2011 Official Proceedings*. 48-60. ISSN: 2186-4691.

Chen, Li-Mei, and Raymond D. Kent. 2009. Development of prosodic patterns in Mandarin-learning infants. *Journal of Child Language*. 36.1:73-84.

Chen, Qinghai. 1997. Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers Association* 32:21-39.

Clumeck, Harold Victor. 1977. *Studies in the acquisition of Mandarin Phonology.* Ph.D. Dissertation. University of California, Berkeley.

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

Corder, Stephen Pit. 1981. *Error analysis and interlanguage*. Oxford University Press.

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

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

679 Demuth, Katherine, and Mark Johnson. 2003. Truncation to subminimal words in
680 early French. *Canadian Journal of Linguistics* 48:211-241.
681 Demuth, Katherine, and Annie Tremblay. 2008. Prosodically-conditioned variability
682 in children's production of French determiners. *Journal of child language* 35.1:
683 99-127.
684 Ding, Hongwei. 2012. Perception and production of Mandarin disyllabic tones by
685 German learners. *Speech Prosody 2012*.
686 Duanmu, San. 2007. *The Phonology of Standard Chinese*. 2nd Edition. Oxford: Oxford
687 University Press.
688 Fauziati, Endang. 2011. Dynamicity of Interlanguage Errors: A Case study of
689 Indonesian students learning English as a foreign language. *Indonesian Journal*
690 *of Applied Linguistics* 1.1:25-40.
691 Feng, Li Ping. 1997. Distribution of and comparison on phonetic errors between
692 English and Japanese students of beginning-level Mandarin. *The Fifth*
693 *Conference on Chinese Language Teaching*. Bejing: Bejing University Press.
694 Fu, Qian-Jie, and Fan-Gang Zeng. 2000. Identification of temporal envelope cues in
695 Chinese tone recognition. *Asia Pacific Journal of Speech, Language, and*
696 *Hearing* 5:45–57.
697 Gandour, Jack. 1979. "Perceptual dimensions of tone: Thai." *Southeast Asian*
698 *linguistic studies* 3: 277-300.
699 Guo, Lijuan, and Liang Tao. 2008. Tone production in Mandarin Chinese by
700 American students: A case study. In *Proceedings of the 20th North American*
701 *Conference on Chinese Linguistics (NACCL-20)*1: 123-138.
702 Hsu, Joseph H. 1987. *A study of the various stages of development and acquisition*
703 *of Mandarin Chinese by children in Chinese milieu.* National Science
704 Council project report.
705 Hsu, Joseph H. 2003. *A study of the stages of development and acquisition of*
706 *Mandarin Chinese by Children in Taiwan.* Taipei: Crane Publishing.
707 Huang, Hui-Chung. 2014. *An Analysis of the Strategies of Teaching Mandarin Tones*
708 *to Japanese Learners of Chinese at the Basic Level*. MA thesis, Chung Yuan
709 Christian University, Taoyuan.
710 Huang, Yi-Ning. 2013. *The Tone Performance of Disyllabic Words in Mandarin*
711 *Speech Flow---For the Intermediate-Level-Learners of Native English*
712 *Speakers*. MA thesis, National Taipei University of Education, Taipei.
713 Hudak, Thomas John. 1987. Internal rhyme patterns in classical Thai
714 poetry. *Crossroads: An Interdisciplinary Journal of Southeast Asian Studies*
715 3.2/3:94-105.
716 IPA Association. 1999. *Handbook of the International Phonetic Association: A guide*
717 *to the use of the International Phonetic Alphabet*. The UK: Cambridge
718 University Press.
719 Kaplan, George A, and Kaplan E. 1971. The Prelinguistic Child. In J. Eliot (ed.),
720 *Human Development and Cognitive Process.* New York: Holt, Rinehart,
721 Winston. 359-381.

Kent, Raymond D. 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.

Krashen, Stephen D. 2003. *Explorations in language acquisition and use: The Taipei lectures.* Portsmouth, NH: Heinemann.

Krashen, Stephen D, James Butler, Robert Birkbaum, and Judith Robertson. 1978. Two studies in language acquisition and language learning. *International Review of Applied Linguistics in Language Teaching* 39.40:73–92.

Larsen-Freeman, Diane. 2002. Making sense of frequency. *Studies in Second Language Acquisition* 24:275-285.

Larsen-Freeman, Diane. 2003. *Teaching language: From grammar to grammaring*. Boston: Heinle.

Leather, Johnathan. 1990. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. *New sounds* 90:72-97.

Liao, Wan-Ju. 2010. The Tone Analysis of English Native Speaker--American Beginners' Study. MA thesis, National Kaohsiung Normal University, Kaohsiung.

Li, Charlie N., and, Sandra A. Thompson. 1977. Acquisition of tone in Mandarin-speaking children. *Journal of Child Language* 4.2:185-199.

Lin, H. 2007. *Mandarin Tonal Acquisition of Novice Japanese Learners*. MA thesis, National Taiwan Normal University, Taipei.

Lin, Yen-Hwei. 2007. *The sounds of Chinese*. Cambridge: Cambridge University Press.

Liu, J. N. 2006. *Study on the Tone Errors of Intermediate-Level Korean Students of Mandarin*. MA thesis, Tianjin Normal University, Tianjin.

Locke, John L. 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, John L. 1983. *Phonological Acquisition and Change.* New York: Academic Press.

Ma, Wei-Yun, and Chen Keh-Jiann. 2003. Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff. *Proceedings of ACL, Second SIGHAN Workshop on Chinese Language Processing*.168-171.

Mehler, Jacques, Peter Jusczyk, , Ghislaine Lambertz, Nilofar Halsted, Josiane Bertoncini, and Claudine Amiel-Tison. 1988. A precursor of language acquisition in young infants. *Cognition* 29.2:143-178.

Miracle, W. Charles. 1989. Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers Association* 24.3:49-65.

Miyakoda, Haruko. 2012. Markedness within phonological theory: Focus on acquisition and disorder data. *Acoustical Science and Technology* 33.3:142-146.

Munthuli, A., C. Tantibundhit, C. Onsuwan, K. Kosawat, and C. Wutiwiwatchai.

2015. Frequency of occurrence of phonemes and syllables in Thai: Analysis of spoken and written corpora. In *Proceedings of 18th International Congress of Phonetic Sciences*.

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

Ohala, John. J., and William G. Ewan. 1972. Speed of pitch change. *Journal of the Acoustical Society of America* 53*:*345.

Ong, S. & C. H. Yang. 1997. On the energy and duration of the four tones of Mandarin speech. *Journal of Technology* 12.1: 125-129.

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

Selinker, Larry. 1972. Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching* 10.1-4: 209-232.

Shen, Xiao-nan Susan. 1989. Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association* 242:7-47.

So, Lydia KH, and Barbara J. Dodd. 1995. The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language* 22.3: 473-495.

Tingsabadh, MR Kalaya, and Arthur S. Abramson. 1993. Thai. *Journal of the International Phonetic Association* 23.1: 24-28.

To, Carol KS, Pamela SP Cheung, and Sharynne McLeod. 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.

Tran, T. K. L. 2005. *Error Analysis of Mandarin Tones from Vietnamese Learners*. MA thesis, National Taiwan Normal University, Taipei.

Tsai, Ping-Lun. 2008. *Perception and Production of Chinese Lexical Tone by Adult English Speaking Learners*. MA thesis, National Tsinghua University, Hsinchu.

Tsao, Feng-Ming. 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, Jane. 2001. Phonetic parameters of tone acquisition in Taiwanese. In M. Nakayama (ed.), *Issues in East Asian Language Acquisition* 7:205-226.

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

Tseng, Chiu-yu. 1990. *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Taipei: Institute of History and Philology.

Tseng, Chiu-yu. 1999. An analysis on the perception of Mandarin Chinese tones by Chinese language learners from South -East Asia. *The Second Conference of Chinese Language Teaching in South East Asia*. 92-113.

Tseng, Jin-Jin. 2001. [jiunianyiguan kecheng zhong de minnanyu jiaocai jiaofa] Minnanese Textbooks and Teaching Methods in the Nine-year Education Policy. *Guowen Tiandi* 17.1:8-13.

Tseng, Jin-Jin. 2007. [yindunixiya xuesheng huayu fayin pianwu fenxi] Error Analysis of Chinese Pronunciation by Indonesian Students. Paper presented at

the *8th International Conference of TCSL*, Beijing.
Tseng, Jin-Jin, and Furukawa Chiharu. 2006. [yi renzhi xinlixue wei jichu de shuweihua yu shengdiao jiaoxue chutan: yi chuji riji xuesheng wei duixiang] Research on Digital Chinese Tone Teaching Based on Cognitive Psychology: A Case Study of Elementary Level Japanese Students. Paper presented at the *3th International Conference of TCSL*, Taipei.
Tuaycharoen, Pintip. 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, Marilyn May. 1996. *Phonological development: The origins of language in the child*. Blackwell Publishing.
Vihman, Marilyn May, and Lorraine McCune. 1994. When is a word a word? *Journal of child language* 21.3: 517-542.
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.
Wan, I-Ping, and Han-Chieh Yang. 2017. Tone Acquisition at the One-Word Stage in Taiwan Mandarin: A corpus study. *Journal of Child Language Acquisition and Development* 5.3:165-182.
Wong, Puisan. 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, Puisan. 2012. Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics* 40.1: 141-151.
Wong, Puisan. 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, Puisan, Richard G. Schwartz, and James J. Jenkins. 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.
Wang, Y. 1995. On the topic of Americans learning Mandarin tones. *Language Teaching and Research* 3:126-140.
Wang, Y. 1995. Ye tan meiguoren xuexi hanyu shengdiao. *Yuyan Jaoxue Yu Yanjiu* 2:126-140. (On American learners' tone acquisition. Language Teaching and Research 2:126-140.)
Menji, W. U., and H. U. Mingguang. 2004. An analysis of Vietnamese students errors in Chinese Tones. *Chinese Teaching in the World* 68:81-87.
Xu, Li, Yuhjung Tsai, and Bryan E. Pfingst. 2002. "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses." *The Journal of the Acoustical Society of America* 112.1:247-258.
Yang, Han-Chieh. 2013. *Tone acquisition in Taiwan Mandarin.* MA thesis. National Chengchi University, Taipei.

851 Yip, Moira. 2002. *Tone*. Cambridge, UK: Cambridge University Press.
852 Yu, A. Q. 1988. Discussion on the teaching of tone. *The 1st International Conference*
853 *on Chinese Language Teaching*. Beijing: Beijing Language and Culture Press.
854 Yue-Hashimoto, A. O. 1986. Shengdiao jiaofa de Shangque. *Diyijie Guoji Hanyu*
855 *Jiaoxue Taolunhui Lunwenxuan* 229-235.
856 Zhu, Hua. 2002. *Phonological development in specific contexts: Studies of Chinese-*
857 *speaking children*. Clevedon, England: Multilingual Matters Limited.
858 Zhu, Hua, and B. Dodd. 2000. The phonological acquisition of Putonghua (modern
859 standard Chinese). *Journal of Child Language* 273:3-42.

860 Appendix

Major Corpora in Taiwan Mandarin in the Phonetics and Psycholinguistics Laboratory at National Chengchi University							
Corpus types	Participants	Cumulative hours	Recording date	Location	Progress rate	Structure database	PRAAT multi-tier labeled corpus
Thousand hours of NLP corpus (dialogue: multiple speakers)	>50 Taiwanese studying overseas	1000+ hrs	1995-1999; 1999-2002	SUNY-Buffalo, NY, USA; Taiwan	86%	V	55 hours
Experimental corpus of speech error （single word: a speaker）	22 college students	0.9K hrs	2003-2005	Chengchi University Taipei	100%	V	N/A
Aphasia corpus （dialogue: two speakers）	18 patients	1K hrs	2003-2009	Taiwan University Hospital	90%	V	N/A
Experimental corpus of aphasia （single word: a speaker）	10 patients	0.3K hrs	2006-2010	Taiwan University Hospital	100%	V	N/A
Child acquisition corpus （dialogue: mother and children）	24 children (0;7-6;11)	1.6K hrs	2012-2016	Children's home (Taipei and New Taipei)	10%	under construction	1%
Children experimental spoken corpus (single word/description: a speaker）	260 children (0;7-6;11)	260 hrs	2012-2014	Recording: RCMBL; Corpus data: PPL	50%	Under construction	N/A

Spoken corpus of Phonologically disordered children（dialogue: two speakers）	18 children (3;0-6;11) (VGH patients)	0.2K hrs	2014-2016	Taipei Veterans General Hospital (VGH)	1%	N/A	N/A
Experimental corpus of children with phonological disorder (single word: a speaker）	20 children (3;0-6;11) (VGH patients)	20 hrs	2014-2016	Research Center for Mind, Brain & Language (RCMBL)	64%	Under construction	N/A
Mandarin corpus of Thai learners	11 toddlers 13 children 14 teenagers	967 hrs	2018-2019	Thai-Chinese International School	40% (Toddlers)	Under construction	5% (Toddlers)

中文摘要

本文主要是對於泰國曼谷中華國際學校 11 位幼童學習華語提出初步聲調中介語研究。本案所有學習語料庫建構於 Praat 聲學聲檔中，語料庫自動分成 17 層，其中涵蓋授課老師與參與幼童的口語語音，並且採用機器學習方式半自動處理聲學語料（人工處理僅佔 30%，系統自動標註語言學標記佔 70%）。本研究將幼兒詞彙中的音節數量（單音節、雙音節及多音節）搭配聲學基頻(F0)做一初步觀察。目前結果顯示為泰國幼童本身母語為五調聲調多音節語言，比華語複雜，因此他們口語產製華語聲調並沒有任何困難，而且他們對於雙音節詞彙有相當高的使用比例。傳統研究提及的第二語言偏誤，在本文中則是建議該以中介語作為學習外語的過渡中介期，主因在於這群泰國幼童產生的聲調偏誤類別，聲學基頻參數顯示並非來自於泰語母語聲調，而是偏向近似華語的聲調調型。由此可見，泰國幼童或許刻意建構一個獨立聲調新系統，作為學習新語言的中介過渡期。未來研究將會在華語聲調辨識上執行另一項語音辨識實驗，亦即將所有聲檔以低頻波過濾(low-pass filter)所有的語意詞彙，以求更精準的譯寫正確率。