Comparison of Pitch Variation and Pitch Range in L1 and L2 Mandarin Chinese

Man Gao
Dalarna University
mao@du.se

Abstract
The objective of this work is to investigate speakers’ fundamental frequency in L1 and L2 Mandarin speech. A comparison of various measures including F0 mean, maximum and minimum F0 and F0 range was made between native and L2 speakers’ five different types of speech. Two native Mandarin Chinese speakers and two L2 Chinese speakers whose native language is Swedish were invited to read disyllabic words in isolation, statements, questions, text and hold a conversation with the author. Acoustic measures were performed on the speech data for analyzing the speakers’ F0 uses. The findings seem to suggest that L2 Mandarin deviates from L1 Mandarin systematically but in a different manner depending on the speech genre. The results are likely to be related to different sound structures and the cultural norms of reading various speech genres in Mandarin and Swedish, as well as the influence one’s native language has on one’s L2 language production.

Index Terms: Mandarin Chinese, pitch range, pitch use, L2 Mandarin Chinese, Swedish

1. Introduction
The lexical tone of Mandarin Chinese has been a topic of interest in many linguistic research areas such as speech perception and production, language acquisition, and cross-language studies. It is well established that Mandarin Chinese has a vastly different phonological system from Indo-European languages such as English; one crucial distinction is Chinese lexical tone. Prior studies ([1], [2], [3]) have found that Mandarin Chinese speakers have a wider pitch range than English speakers, which may be attributable to the Chinese tonal system. The term ‘pitch range’ in this paper is employed in a linguistic sense to refer to the range of pitches between high and low F0 that are commonly used in speech. A systematic review of the literature shows that F0 use (i.e., pitch range and pitch level) can vary across languages ([4], [5], [6]), in different social groups ([7]), and also between dialects of the same language ([8], [9]).

It is generally agreed that, in the field of second language (L2) acquisition, Mandarin tones are one of the major obstacles for learners of Chinese from non-tonal language backgrounds. Therefore, there is a very extensive literature on the topic of L2 acquisition of Mandarin tones. However, this research is focused on: 1) pronunciation and perception accuracy of tones; 2) types of errors in pronunciation and perception of tones; and 3) categorical perception of tones. Investigation of pitch range and pitch variation of L2 Mandarin Chinese is highly limited. A few studies ([10], [11], [12], [13]) have found that beginner level learners of Mandarin present a narrower pitch range than native speakers, although these observations were mostly based on an analysis of words that include only one or two syllables. There is no direct study investigating the pitch range and pitch variation of L2 Chinese in text or conversations, to the best of the author’s knowledge.

There is, however, a body of work that investigates pitch range and pitch variations of L2 speech in other languages. These suggest that “transfer of pitch range (from L1 to L2) is common” ([14], pp. 310). Two studies ([15], [16]) found that Dutch speakers’ L2 English displays a narrower pitch range than native speakers’ L1 English. Similarly, [17] reported that both Venezuelan and German L2 English learners have a narrower pitch range than English native speakers ([18]). A possible explanation for these distinctions might be that English has a wider pitch range than the native languages of the L2 learners under investigation. More recently, [19] examined the F0 patterns of the L2 English produced by speakers from a tone language background - Vietnamese. It was found that Australian English exhibits a broader F0 range and F0 variation than beginner level L2 English. [19, pp. 113] attributed this finding to the influence of L1 and therefore rejected the hypothesis that “tone languages have an overall larger F0 range”. However, the study by [20], who compared speech by native English speakers, native Chinese speakers and Chinese L2 English learners, yielded somewhat contradictory results. Compared with native English speech, Chinese L2 English displays a wider pitch range at the phoneme level but a similar range at the sentence level. Previous studies summarized above have established that L2 English pitch use deviates from native English speech: native English generally has wider pitch range than that displayed by L2 speakers. In view of all that has been mentioned so far, one may suppose that there are also systematic differences of F0 use between native Mandarin Chinese and L2 Mandarin speech.

The present study explores the pitch profile of L2 Mandarin produced by Swedish informants. This investigation is original in two respects: first, it examines the pitch profile of L2 Mandarin, something which has rarely been addressed before; and, secondly, it is the first attempt to investigate the effect of a pitch accent language (Swedish) on the pitch profile of the L2 production of a tone language (Mandarin). The aims of this study are to compare the pitch profile of native Mandarin Chinese and Swedish L2 Mandarin Chinese and assess the pitch variations of different types of speech samples produced by L1 and L2 Chinese speakers. This study was exploratory in nature, with the purpose of establishing a basis for a larger-scale work that surveys the speaking fundamental frequency of L2 Chinese speakers from different language backgrounds.
2. Method

2.1. Subjects
A total of four male informants were recruited for the recording. Two were native speakers of Chinese, whose Mandarin was rated as grade A of level 2 or above in the National Proficiency Test of Putonghua\(^1\). The other two informants were native speakers of Swedish (Götaland dialect) who speak Mandarin Chinese as their second language at home. These four informants were between 24 and 39, with a mean age of 29.3 years. None of them had received any previous musical training, nor, according to their self-reporting, did they have any speech disorders.

2.2. Speech Material and Procedure
The speech material used in this study includes five different types: 1) Disyllabic words in isolation. A number of disyllabic words were selected to include all possible Chinese syllable types, all tonal combinations, and as many consonant and vowel combinations as possible; 2) Disyllabic words from 1) above embedded in two types of carrier sentences: statements and questions; 3) A narrative text of 108 syllables composed of simple Chinese characters and words; 4) The well-known Chinese children’s story ‘The Shepherd Boy & the Wolf’; and 5) Conversation with the author of this paper about the experience of learning Chinese (for Swedish informants) and about travelling abroad (for Chinese informants). All informants produced the speech material 1) to 2) twice and the speech material 3) to 5) once. All informants were instructed to read or speak at a normal speech rate.

A field memory recorder (Foxtec FR-2LE) and a high-quality microphone (AKG CK93) were used for the recording. The informant was seated in a quiet room with about 15 cm between the microphone and his mouth. All speech data were recorded at a 44.1 kHz sampling rate and a bit depth of 32.

The speech material was presented to the informants in typed Chinese characters. Each informant was given sufficient time to familiarize themselves with the material. For the two Swedish informants, pinyin (the Chinese phonetic alphabet) was provided for some words in the type 1 material.

2.3. F0 Analysis
F0 analysis of the collected speech data was conducted using Praat ([21]) and the script Prosody Pro ([22]). The F0 information was extracted with automatic vocal cycle markings, which is subject to subsequent manual corrections. Some speech data had to be excluded from the F0 analysis because of creakiness (especially the syllables that carry Tone 3), misreading, and long pauses within a sentence. For disyllabic words produced in isolation, data including pronunciation errors were also removed because the comparison of F0 is made among syllables carrying four lexical tones. Because ‘The Shepherd Boy & the Wolf’ included many characters that were unfamiliar to the Swedish informants, their readings were completed with many pauses and repetitions that were subsequently removed from further analysis. The number of tokens and sentences included in the F0 analysis and comparison are listed as follows:

- Disyllabic words in isolation – L1: 116 words; L2: 80 words.
- Statements – L1: 52 sentences; L2: 35 sentences.
- Questions – L1: 52 sentences; L2: 35 sentences.
- Simple text: one reading with a total of 9 sentences for every informant.
- Conversation: the first 15 sentences were analyzed and compared for every informant.

Table 1 below lists the six types of measures used in this study when comparing L1 Chinese with L2 Chinese. For a more reliable comparison between L1 and L2 informants, all measures were converted to the logarithmic scale – semitone.

<table>
<thead>
<tr>
<th>Measure Type</th>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum F0</td>
<td>F0max</td>
<td>Highest F0 value in a token; average max. F0 across all tokens of the same type</td>
</tr>
<tr>
<td>Minimum F0</td>
<td>F0min</td>
<td>Lowest F0 value in a token; average min. F0 across all tokens of the same type</td>
</tr>
<tr>
<td>Span of F0</td>
<td>F0 range</td>
<td>F0max-F0min within a token; average span of F0 across all tokens of the same type</td>
</tr>
<tr>
<td>Average F0</td>
<td>F0 mean</td>
<td>Average of F0 values in a token; the mean of average F0 across all tokens of the same type</td>
</tr>
<tr>
<td>kth percentile of F0</td>
<td>k% F0</td>
<td>F0 value at which k percent of the F0 data set is below</td>
</tr>
<tr>
<td>10-90% percentile range of F0</td>
<td>10-90% F0</td>
<td>Difference between the 90th and 10th percentiles of F0</td>
</tr>
</tbody>
</table>

The F0 range is computed using the Hz to semitone equation established by [23]. For all other measures, the semitone of a F0 value is computed using the following formula:

\[
F0 \text{ (st)} = 12 \log_2 \left( \frac{F0 \text{ (Hz)}}{\text{1/2Hz}} \right)
\]

3. Results
The paper now turns to the presentation of the results obtained. Only a limited statistical analysis has been performed due to the small number of informants recruited to this study.

F0max, F0min and F0mean of all five speech types are presented in Table 2 below. An invariable pattern between L1 and L2 Chinese is lacking. For monosyllables (the first syllable in the disyllabic word), statements and questions, L1 Chinese generally displays a slightly higher F0max, but the pattern for F0min is mixed. For the free speech data, L2 Chinese has a higher F0max and a lower F0min. As for the simple text data, divisions, and Chinese-language teachers in China are required to reach level 2, grade A or above.

\(^{1}\) The National Proficiency Test of Putonghua is a standardized test to evaluate Chinese native speakers' spoken fluency in Mandarin Chinese (Putonghua). There are a total of six
L1 Chinese has F0max similar to L2 Chinese, but a higher F0min. The differences observed for F0max and F0min are quite small in Semite. As for the measure of F0mean, the difference between L1 and L2 Chinese is generally small. However, the L2 informant M2 appears to have a higher F0mean than the other three informants, except for the question data type. Both L1 informants have highest F0mean when reading questions and the lowest value when conversing freely.

Table 2: Comparison of F0 max and F0 min.

<table>
<thead>
<tr>
<th>Speech type</th>
<th>F0max (st)</th>
<th>F0min (st)</th>
<th>F0mean (st)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
<td>M2</td>
<td>M1</td>
</tr>
<tr>
<td>Mono-syllable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1(CH)</td>
<td>87.3</td>
<td>89.2</td>
<td>80.9</td>
</tr>
<tr>
<td>L2(SV)</td>
<td>84.7</td>
<td>89.7</td>
<td>81.9</td>
</tr>
<tr>
<td>Statement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1(CH)</td>
<td>90.9</td>
<td>93</td>
<td>73.8</td>
</tr>
<tr>
<td>L2(SV)</td>
<td>88.4</td>
<td>91.2</td>
<td>76.8</td>
</tr>
<tr>
<td>Question</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1(CH)</td>
<td>93.7</td>
<td>94.2</td>
<td>76.7</td>
</tr>
<tr>
<td>L2(SV)</td>
<td>90.2</td>
<td>90.9</td>
<td>76.9</td>
</tr>
<tr>
<td>Text (reading)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1(CH)</td>
<td>91.5</td>
<td>91.5</td>
<td>75.7</td>
</tr>
<tr>
<td>L2(SV)</td>
<td>90.9</td>
<td>91.6</td>
<td>73.3</td>
</tr>
<tr>
<td>(Free) Speech</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1(CH)</td>
<td>87.7</td>
<td>90.3</td>
<td>73</td>
</tr>
<tr>
<td>L2(SV)</td>
<td>88.9</td>
<td>91.7</td>
<td>72</td>
</tr>
</tbody>
</table>

The 10% F0 and 90% F0 are plotted as yellow lines on these figures as reference points. Visual comparison between the distribution of F0 from L2 Chinese statement data (the top row) and L1 data (the bottom row), the F0 points of L1 data cover a much wider area than those of the L2 data. The 10-90% F0 of L1 Chinese are also slightly bigger (11.70 st and 11.41 st) than that of the L2 data (7.77 st and 10.61 st). For one of the L2 informants (M2), his high F0 points resemble the high points in the L1 data (the bottom row in Figure 2), but his low F0 points are spread over a large area on the scale from 80 st to 70 st.

Figure 1 below presents the F0 range for individual tones (carried by the first syllable of the disyllabic words), sentences, questions, reading text and free speech. It is found that the native Chinese data generally displays a wider F0 range than the L2 Chinese data, with the exception of text reading and free speech, which presents the opposite pattern. Visual comparison of the width of the boxes in Figure 1 shows that the F0 ranges from L2 speech have larger within-group variance than those from L1 speech, especially for sentences (both statements and questions), text and speech data.

For sentences, reading text and free speech data types, the F0 measure points of the same data type (40 to 55 measure points per sentence) were plotted onto one graph for each informant, as shown in Figure 2 to Figure 4.

In Figure 3 the same pattern is also observed for questions. Here the F0 points are clustered in a much narrower area for the L1 data (the top row), as compared with the lower row for the L1 data. The L2 informants have a much smaller 10%-90% F0 (8.71 st and 8.46 st) than the L1 informants (12.58 st and 12.67 st). Similarly, the low F0 points of the L2 informant M2 are scattered along the scale, which is not the case with the other informants.

The pattern observed in the statement and question data is not found with the text reading data or the free speech data. Figure 4 below only shows the distribution of F0 in free speech data, but the results for text reading are strikingly similar. Visual inspection shows that the distribution of F0 is about the same between L1 and L2 Chinese, but the L2 informant M2 consistently presents F0 points that are higher on the pitch scale than the others.
4. Discussion

This study examines whether there exists systematically different F0 uses, in particular maximum and minimum F0, F0 range, and variation of F0, between native Mandarin speech and L2 Mandarin produced by Swedish speakers. The study also examines if various types of speech data (disyllabic words, statements, questions, text reading and free speech) have an influence on speaking fundamental frequency. The overall results show that the L2 Chinese use of F0 varies from L1 Chinese in a number of respects, but that the variation pattern is affected more by the particular speech type. Therefore, the discussion is organized according to the different types of speech data.

At the syllable level, the data of isolated words suggests that it is the lexical tone that determines the F0 pattern. Except for the high level tone (Tone 1), which presents a similar pitch range between L1 and L2 Chinese, L1 Chinese displays a wider pitch range than L2 Chinese for the other three contour tones. The high falling tone (Tone 4) in L2 Chinese in particular shows a drastic deviation from the native Chinese data. This finding confirms the results of existing studies ([10], [12]), which show that American learners of Chinese have a narrower pitch range than native Chinese speakers at the syllable or word level. Even though Swedish is a pitch accent language unlike English, the two Swedish accents only contrast the meaning of a limited set of words ([24]). In Central Standard Swedish, for example, the pitch level is not utilized to differentiate these two accents but the number of F0 peaks. Therefore, having Swedish as the mother tongue for L2 Chinese speakers may affect their F0 patterns in Chinese, at the syllable level at least, in the same way as English.

At the sentence level, the overall pattern is the same for monosyllables: the pitch range of L1 Chinese data is wider than L2 data, which is well supported from measurements of F0max, F0 range and 10-90% F0. Additionally, the L2 informant M2 exhibits a much greater F0 variation than the other informants. The L1 data show that questions have a higher F0max than statements while this is not the case for L2 Chinese. This is consistent with observations made in early work by [10], who also analyzed the pitch range of sentence and observed that the utterance sentences provided by native Chinese speakers have a wider pitch range than American learners of Chinese. The intonation of utterances has been extensively studied for many languages. It is very well established that almost all languages use a raised F0 to signal questions ([25]). Research ([26]) shows that the intonation of questions in Swedish has a raised topline and the focus widens the total range of the F0 curve and in Mandarin Chinese that the F0 of a complete sentence and a sentence-final syllable are raised to a higher level in questions than in statements ([27]). Considering the similarity in sentence intonation between Mandarin Chinese and Swedish, the observed differences of various F0 measures between L1 and L2 Chinese could also be attributed to the effect of L2 speakers’ native language on their Chinese speech.

The text reading and free speech data show a highly similar F0 profile when comparing L1 and L2 Chinese. Contrary to the pattern observed at syllable and sentence level, measurement of F0min, F0max and F0 range shows a broader pitch range for L2 Chinese data. However, the 10-90% F0 measure indicates that L1 and L2 Chinese are highly similar, except that data from the L2 informant M2 has an overall higher pitch level than data from the other informants. The results of these two sample types corroborate previous findings in two aspects. Firstly, previous work has shown that text reading and free speech resemble each other in terms of speaking fundamental frequency and that, therefore, text reading speech can represent natural speech ([28]). The current study provides further evidence from Mandarin Chinese in support of this claim. Secondly, in another study ([3]), it was found that American English speakers’ F0 uses are the same as Mandarin speakers when reading a text, but they use a much higher F0max, a lower F0min, and a broader F0 range than Mandarin speakers when telling a story, which is opposite to the pattern observed from words in isolation data. Though the current study compares L1 and L2 Mandarin Chinese, instead of speech from two different languages, the results are surprisingly in line with those from [3]. A possible explanation which they offer ([3, pp. 1059]) for these results may be due to what they describe as the “cultural conventions” of reading different genres in English and Mandarin, which can perhaps account for findings from this study as well. Though it is Swedish informants’ Mandarin speech that is under examination, the influence of their native language’s pitch range on their L2 is to be expected ([14]). For the text reading and free speech genres in this study, however, only 9 and 15 sentences respectively per informant were analyzed as opposed to the 35 and 52 sentences for statements and questions per informant. With a small sample size to begin with, caution must be applied to this type of speech material as the findings might not properly represent the informants’ L2 Mandarin.

5. Conclusion

The aim of the present pilot study was to examine the idea that L2 Mandarin Chinese may have different speaking F0 from L1 Mandarin speech. The results show that L2 Mandarin produced by Swedish speakers differs from native Mandarin Chinese in two different patterns depending on the type of speech material being used. For words in isolation and sentences, L2 Mandarin has a narrowed pitch range and a generally less extreme F0max and F0min, but the opposite pattern is found for text reading and free speech data. Possible causes for these differences are the effect of L2 speakers’ native language and the customs with which L2 speakers approach various types of speech genres. Since this is only a pilot study with 4 informants and limited text reading and free speech data, the results reported here are only preliminary in nature. Notwithstanding the potential limitations of the findings, the study certainly adds to our understanding of the F0 profile in L2 Mandarin Chinese.
6. References


