

# Perception and Production of Mandarin Disyllabic Tones by German Learners

Hongwei Ding

School of Foreign Languages, Tongji University, China

hongwei.ding@tongji.edu.cn

## Abstract

This study investigates the possible deviations related to Mandarin disyllabic tone perception and production by German speakers. Nine German subjects are involved in both perception and production tests. The stimuli of tone perception consist of 40 disyllabic words uttered by a standard professional Chinese speaker. These words are also employed as reading material in the production experiment. Statistical analysis and acoustic measures have been carried out to find out the tonal performance of the German learners. Results show that their perception performance is best with the falling tone and worst with the rising tone, although there are some differences between the first and the second syllables. However the production performance exhibits different patterns, each subject has his or her preferred tonal combination, often independent of the target tones. It is further observed that these German subjects are reluctant to produce the second syllable of a disyllabic word in a falling tone. Findings can provide some implications for cross linguistic study and second language acquisition.

**Index Terms:** Mandarin disyllabic words, German learners, learning Chinese tones

## 1. Introduction

It has been demonstrated that German intonation is more monotonous and less lively than English [5], while Mandarin Chinese is characterized by rich and steep pitch movements [2]. It is interesting to examine how German speakers process Mandarin tones differently than Mandarin native speakers.

As a tone language, Mandarin includes four lexical tones, Chao [1] used a letter notation system to distinguish among these tones, in which '1' represents the lowest and '5' represents the highest pitch level of the speaker's pitch range. When pronounced in isolation, these four tones have shapes that ideally look like those seen in Figure 1.

In our previous investigation of monosyllabic tones [3], de-

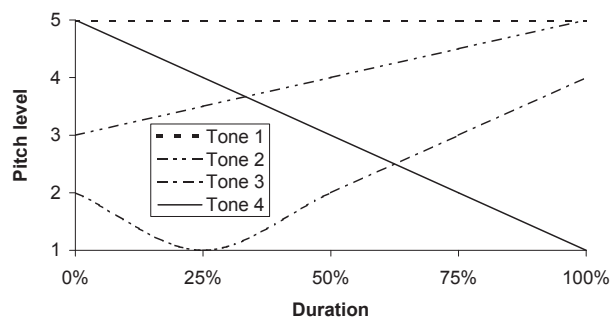


Figure 1: Canonical pitch forms of 4 tones produced in isolation

viations of these four tones produced in isolation can be summarized as following: German speakers

- can hardly keep high tone level;
- can not raise their pitch straightforward for rising tone;
- have difficulty to dip the low tone so deep as the natives;
- have more gradual and less steep falling tones.

However in the real world speech, people seldom produce tones in isolation. Syllables will be grouped into words, phrases and utterances. To investigate continuous speech, words are more practical than syllables. While more than 80% words are disyllabic, this paper aims to extend our previous investigation [3] of isolated tones of German learners to disyllabic words.

## 2. Method

This study aims to address the following questions:

- Are there any correlations between the perception and production of Mandarin tones of the German learners?
- Are there any differences in perceptual and production performance in the first and second syllables respectively?
- Are there any particular patterns for their perception and production performance?

### 2.1. Subjects

Subjects were selected from a Chinese course for German students at Tongji University in Shanghai. They came from different parts from Germany with the age between 22-26. They had learned Chinese for more than 1 year with 4 hours per week in Germany, and had been in China for 2 months. All of them had already about 150 hours of Chinese courses. Nine average students were chosen as subjects from this class, two of them were female and seven were male. All of them participated in the perception and production tests. None had difficulty in hearing and speaking.

### 2.2. Stimuli and procedure

The stimuli were disyllabic words taken from A-grade Word of "The Syllabus of Graded Word and Characters for Chinese Proficiencies", which consist of the most frequent words. 40 words were selected to cover all 20 combinations with each tonal combination occurring twice. Considerations were also taken to incorporate different vowels, consonants and their combinations, because different categories of vowels and consonants can also have effects on tonal contours of the syllables [7]. Most of the words were familiar to these students, some were new words. The purpose to mix familiar words with unfamiliar ones is to avoid their tonal performance merely out of memory. The experiments are divided into perception test and production test.

### 2.2.1. Perception test

These 40 disyllabic words were produced by a professional Chinese male speaker, the utterances were recorded with 16 kHz and 16 bit in a quiet room as perception stimuli. In the perception test, the subjects were asked to listen to the recording and write out the most probable lexical tone marks for the syllables they had heard. A test was done beforehand to adjust the duration of the pauses between words, loudness, etc., so that the recording was played to all the subjects only once in a quiet room at a comfortable listening level and at an appropriate pace. The words were arranged in a random order in terms of tonal combination, so that a certain tonal rhythm could be avoided in the perception process. In the test paper both pinyin transcriptions and Chinese characters of each word were presented. The students only had to write the tone diacritics above the corresponding vowels they had perceived.

### 2.2.2. Production test

The same subjects were then involved in the production experiment. The recording was conducted one by one in a quiet room with 16 kHz and 16 bit. The recording gain was set to ensure a similar level across subjects with no clipping. In the production experiment, the same words as those in the perception test were presented to the subjects. In addition to the characters and pinyin transcriptions, tone diacritics were also indicated above the vowels. The subjects were asked to read these words with the right lexical tones at a normal rate.

With nine subjects we have achieved 360 words both in perception and production.

### 2.3. Analysis

While the test papers of perception could be simply evaluated with the choices of the participants, the statistics of production was very complicated. Because some productions were so ambiguous, it was quite difficult to put them into a certain category. Three Chinese native speakers were organized to evaluate the production. If there was any disagreement, the decision favored the majority. Then the acoustic analysis was carried out with PRAAT software on the computer.

Each word was annotated with initial and final. To adjust for differences in speaking rate, duration was normalized per segmental annotation across speakers. F0 of each utterance was estimated at 10 equal points of each annotation using the script developed by Xu [9].

To accommodate the pitch range differences across males and females, German and Chinese speakers, f0 was normalized for each speaker across the tones. F0 values obtained from each speaker were converted to their logarithms, using a formula commonly adopted for such purposes [6]:

$$T(X) = 5 \frac{\lg X - \lg L}{\lg H - \lg L} \quad (1)$$

where H and L are the highest and lowest f0 for a given speaker, and X is any given point of a pitch contour. The output (T) is a value between 0 to 5, which is similar to the 5-point pitch scale for Mandarin tones proposed by Chao [1].

## 3. Results

The results are presented in term of accuracy of perception and production.

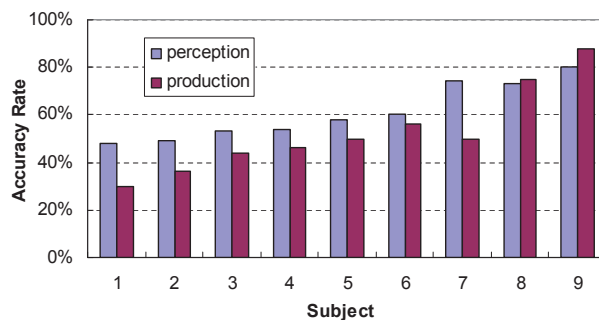


Figure 2: Perception and production of individual subject

### 3.1. Overall Accuracy

We have achieved 720 syllables (40 disyllabic words x 2 syllables x 9 subjects) both in perception and production tests. The overall accuracy of tonal perception and production is illustrated in Table 1.

Table 1: Overall accuracy of perception and production

	Accuracy		
	1st syllable	2nd syllable	sum
perception	58%(208)	56%(201)	<b>57%(409)</b>
production	58%(210)	48%(171)	<b>53%(381)</b>
sum	<b>58%(418)</b>	<b>52%(372)</b>	<b>55%(790)</b>

The overall accuracy rates of first syllable are a little better than those of the second syllable both in production and perception. Among the subjects tonal perception accuracy ranges from 48% to 80%, production accuracy ranges from 30% to 88%. The difference of production is greater than that of perception.

### 3.2. Correlation between perception and production

A comparison of perception and production confusion patterns of the subjects is shown in Figure 2. Accuracy was evaluated in syllables rather than in words.

The accuracy of tone perception is highly correlated [ $r=0.88$ ] with that of tone production, and the correlation is significant [ $p < 0.001$ ].

In order to gain a clear picture of the results, we present the performance of first and second syllables separately in the following perception and production tests

### 3.3. Perception experiment

#### 3.3.1. First syllable

Only lexical tones 1 - 4 can occur in first syllable, one tone can be followed by 5 different tones (including neutral tone), which results in 90 tokens (5 combinations x 2 times x 9 subjects) altogether. The target tone in the recording could be perceived by subjects as one of these four lexical tones, the perception result of first syllable is presented in Table 2.

Tone 4 and 3 are identified more accurately than tone 1 and 2. The pitch contours of the first syllable of the Mandarin speaker for perception are illustrated in Figure 3. The confusion between tone 1 and 2, tone 3 and 4 could be attributed to the similar end pitch height, while the confusion between tone 2 and 3 may due to the similar beginning pitch level.

Table 2: Accuracy of first syllable in perception

Target	Perceived			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	<b>52%(47)</b>	21%(19)	14%(13)	12%(11)
Tone 2	18%(16)	<b>50%(45)</b>	21%(19)	11%(10)
Tone 3	6%(5)	13%(12)	<b>67%(60)</b>	14%(13)
Tone 4	9%(8)	8%(7)	21%(19)	<b>62%(56)</b>

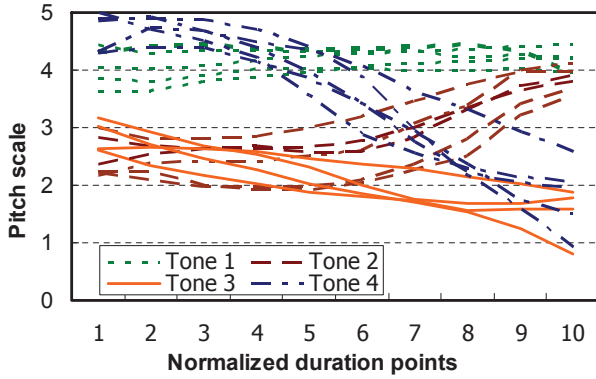


Figure 3: Pitch contours of 1st syllable of Mandarin speaker

The pitch level is the T value calculated with equation (1), time duration is normalized at ten points, each line is the average of two individual values because of two occurrences of the same combination in the stimuli. Each tone contour should be represented by five different curves because of 5 different following tones. Because of the tone sandhi rule, the first tone 3 is changed to tone 2, we have 6 rising tone (tone 2) contours and 4 low tone (tone 3) contours.

### 3.3.2. Second syllable

All 5 tones (including neutral tone) can occur as the second syllable in disyllabic words, but each syllable can only have 4 different combinations. For all tones occurring in the second syllable we can still get 360 tokens (5 tones x 4 combinations x 2 times x 9 subjects), which should be identical to the number of the first syllables. The identification result can be observed in Table 3.

Table 3: Accuracy of second syllable in perception

Target	Perceived				
	Tone 1	Tone 2	Tone 3	Tone 4	Tone 0
Tone 1	<b>61%(44)</b>	13%(9)	7%(5)	17%(12)	3%(2)
Tone 2	17%(12)	<b>47%(34)</b>	21%(15)	10%(7)	6%(4)
Tone 3	14%(10)	10%(7)	<b>47%(34)</b>	29%(21)	0%(0)
Tone 4	8%(6)	6%(4)	13%(9)	<b>61%(44)</b>	13%(9)
Tone 0	13%(9)	4%(3)	4%(3)	17%(12)	<b>63%(45)</b>

Occurring in the second syllable, neutral tone, tone 4 and tone 1 were better identified than tone 2 and tone 3.

With the same procedure we plot the pitch contour in the second syllable of the professional Mandarin speaker for perception in Figure 4. Similarly the confusion between tone 2 and 3 may be due to the beginning pitch height, confusion between tone 3 and 4 to similar end pitch level.

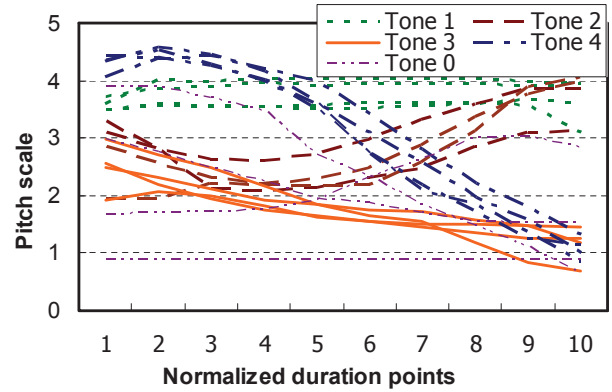


Figure 4: Pitch contours of 2nd syllable of Mandarin speaker

We must point out that tone 3 of our speaker only dips down and does not rise, which can be found in many standard speakers, and there is no glottalization in the middle of his third tone.

## 3.4. Production experiment

The production result is presented in the same way as that of perception.

### 3.4.1. First syllable

The production result of the first syllable is shown in Table 4.

Table 4: Accuracy of the 1st syllable in production

Target	Produced			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	<b>66%(59)</b>	9%(8)	18%(16)	8%(7)
Tone 2	22%(20)	<b>48%(43)</b>	19%(17)	11%(10)
Tone 3	12%(11)	20%(18)	<b>54%(49)</b>	13%(12)
Tone 4	20%(18)	8%(7)	7%(6)	<b>66%(59)</b>

Better accuracy was achieved in tone 1 and 4, confusion between tone 2 and 3, tone 2 and 1, tone 4 and 1 can be observed.

### 3.4.2. Second syllable

Different accuracy patterns can be observed in the second syllable from that of the first syllable in Table 5. Tone 1 and tone 3 achieved better accuracy than tone 2 and 4.

Table 5: Accuracy of the 2nd syllable in production

Target	Produced				
	Tone 1	Tone 2	Tone 3	Tone 4	Tone 0
Tone 1	<b>71%(51)</b>	0%(0)	26%(19)	3%(2)	0%(0)
Tone 2	26%(19)	<b>47%(34)</b>	19%(14)	7%(5)	0%(0)
Tone 3	18%(13)	4%(3)	<b>67%(48)</b>	11%(8)	0%(0)
Tone 4	33%(24)	6%(4)	33%(24)	<b>28%(20)</b>	0%(0)
Tone 0	32%(23)	8%(6)	32%(23)	3%(2)	<b>25%(18)</b>

When we explore the individual subject further, we find that each subject employed a particular pattern to deal with different tonal combinations. We have two arguments to plot finals to represent the disyllabic words: 1) It is observed in our data that

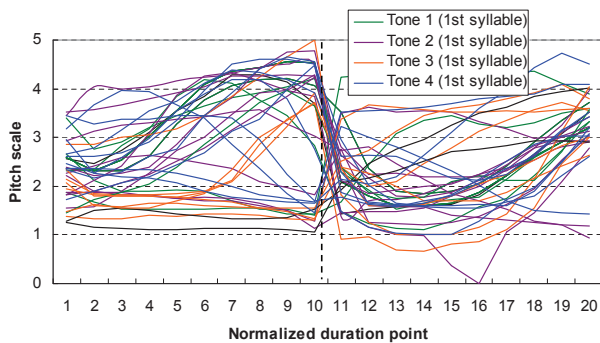


Figure 5: Pitch contours of all utterances of speaker 3

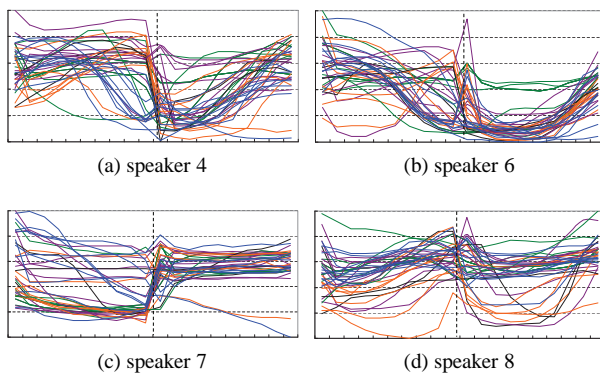


Figure 6: Pitch contours of all utterances

the finals are the main carriers of tonal alignments, which is also reported in [8]; 2) All utterances of different structures can be plotted in one chart for comparison. In each graph in Figure 5-6, Y-axis is the T value calculated with equation (1), x-axis is the normalized time point, the dashed line is the boundary between finals. In Figure 5 many words produced by speaker 3 were perceived as two continuous rising tones by Mandarin native listeners.

Most disyllabic words were perceived in Figure 6

- (a) as tone 1 + tone 2 ( or tone 3) produced by speaker 4
- (b) as tone 4 + tone 2 ( or tone 3) produced by speaker 6
- (c) as tone 3 + tone 1 produced by speaker 7
- (d) as tone 1 + tone 1 produced by speaker 8

#### 4. Discussion

With the results we have actually provided positive answers to the questions put forth at the beginning. When we go further to compare the results of disyllabic with the monosyllabic [3] words, we find:

- In the perception experiment tone 4 has achieved overall best accuracy, and tone 2 the worst accuracy, which coincides with the results in the isolated tones. But tone 1 has obtained less accuracy than tone 3 in disyllabics, which needs to be further investigated.
- Neutral tone perception is actually not easy due to its unstable pitch contour. It is unknown whether the good perceptual performance is due to its shorter duration or

its occurrence in familiar words. More unfamiliar words should be tested on this issue.

- In the production performance no similar results can be found in disyllabic as in monosyllabic words. The subjects seem to employ their own preferred tonal combinations to deal with disyllabics, and their patterns are quite different. One common phenomenon is that the second syllable is seldom produced as a falling tone, which should be further explored with more database and different reading styles.

Because of different material and subjects, the results in this investigation are not totally same as those in [4], but the main findings are quite similar. We have cooperatively conducted some perceptual experiments, and a computer-aided pronunciation training system based on students of Chinese major has been accomplished in Germany [4], we would like to pursue further investigations of German students of non-Chinese major who are studying in China in search for new extension.

#### 5. Conclusions

In addition to practical applications, we also intend to reveal different pitch movement patterns between intonation language German and tone language Chinese. Since disyllabic words play an important role in Mandarin, further investigations should be carried out with more subjects and more unfamiliar words.

#### 6. Acknowledgements

The author is sponsored by Shanghai Pujiang Program (11PJC099) and Innovation Program of Shanghai Municipal Education Commission (12ZS030) for this research work.

#### 7. References

- [1] Chao, Y. R., "A grammar of spoken Chinese", University of California Press, Berkeley, 1968.
- [2] Ding, H., Jokisch, O. and Hoffmann, R., "F0 analysis of Chinese accented German speech", in Proc. ISCSLP 2006, Huo, Q. et al. [Ed], 49-56, 2006.
- [3] Ding, H., Hoffmann, R. and Jokisch, O., "An investigation of tone perception and production in German learners of Mandarin", in Archives of Acoustics, 36(3): 509-518, 2011.
- [4] Hussein, H., Do, H.S., Mixdorff, H., Ding, H. et al. "Mandarin tone perception and production by German learners", in Proc. SLATE, Venice, 2011.
- [5] Jilka, M., "The contribution of intonation to the perception of foreign accent", Doctoral Dissertation, University of Stuttgart, 2000.
- [6] Peabody, M., Seneff, S., Chao Wang, C., "Mandarin tone acquisition through typed interactions", in InSTIL/ICALL, 173-176, 2004.
- [7] Shih, C., "A Declination model of Mandarin Chinese", in Intonation: analysis, modelling and technology, Botinis, A. [Ed], Kluwer: Dordrecht. 243-268. 2000.
- [8] Xu, Y., "Consistency of tone-syllable alignment across different syllable structures and speaking rates", in Phonetica, 55, 179-203, 1998.
- [9] Xu, Y., Online: <http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/>, ProsodyPro.praat. 2005-2011.