

## An Investigation of Tone Perception and Production in German Learners of Mandarin

Hongwei DING<sup>(1)</sup>, Rüdiger HOFFMANN<sup>(2)</sup>, Oliver JOKISCH<sup>(2)</sup>

<sup>(1)</sup> *Tongji University*  
*School of Foreign Languages*  
1239 Si Ping Road, 200092 Shanghai, P.R. China  
e-mail: Hongwei.Ding@tongji.edu.cn

<sup>(2)</sup> *Dresden University of Technology*  
*Institute of Acoustics and Speech Communication*  
Helmholtzstr. 10, 01062 Dresden, Germany  
e-mail: {Ruediger.Hoffmann, Oliver.Jokisch}@tu-dresden.de

(received January 3, 2011; accepted March 7, 2011)

This study investigates the possible errors related to Mandarin tone perception and production by German speakers. In a preliminary test, 23 German listeners should identify the tones of 186 monosyllables. Results show that exposure to Mandarin Chinese can help to discriminate lexical tones as highly expected. In the main experiment, 17 German subjects were asked to take part in a perception and production test. Stimulus of perception involves 48 monosyllables uttered by a standard professional Chinese speaker; acoustic measures were conducted to analyze the production of 72 monosyllables for each subject. It is found that German speakers have much smaller  $f_0$  range than Chinese native speakers. Findings can provide implications for cross language studies and teaching practices.

**Keywords:** Mandarin tones, German speakers, tone perception, tone production.

### 1. Introduction

Because of the increasing contact between Europe and China, German learners demonstrate a growing interest in speaking Chinese. For speakers whose native languages are non-tonal, tones can present a great difficulty (LEATHER, 1990). A number of cross-linguistic studies (PEABODY *et al.*, 2004; WANG, 2003) have been conducted to examine if and how non-tonal speakers process the tones differently. These studies, however, were mainly focused on English learners (LEATHER, 1990; WANG, 2003); the examination of German learners has received little attention. This paper aims to extend the investigation to German learners.

Both English and German are intonation languages, which employ prosodic variations at the sentence level to express communicative intentions. Compared with English, German tone inventory is much simpler, and less steeper than American speakers (JILKA, 2005). German truncates falling accents, falling tones become gradual and simply end earlier in comparison to British English (GRABE, 1998). On the other hand, as a tone language, Mandarin Chinese uses tonal variations at lexical level, and is characterized by rich and steep pitch movements.

Mandarin includes four lexical tones, Chao (1968) 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 Fig. 1.

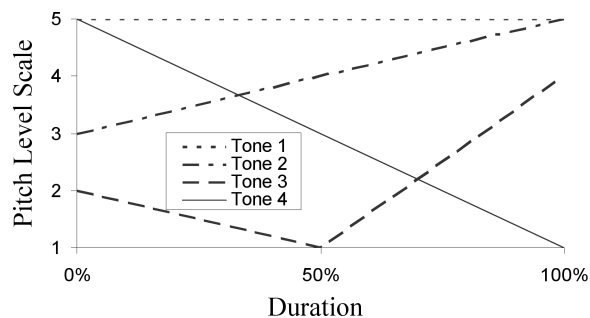


Fig. 1. Canonical pitch forms of 4 tones produced in isolation.

It is obvious that large pitch changes in very short time is characteristic in Mandarin; however, gradual pitch movements are typical for German intonation. Our previous acoustic experiments (DING *et al.*, 2006b; 2007) also provided implications, that Chinese speakers usually transfer the Chinese realization of lexical tones to their German production with larger pitch movements, resulting thus in foreign-accented German. On the other hand, German speakers also try to realize Chinese tones with their German production, which can deviate from that of native Chinese speakers.

## 2. Method

One preliminary experiment of tone identification was carried out to find whether German beginners can discriminate the tonal differences of Mandarin correctly. In the main experiment, perception and production studies were conducted by learners of intermediate level. The perception study was investigated in terms of identification of the tonal categories. The production study was conducted in comparing the acoustic parameters between German learners and Chinese native speakers.

### *2.1. Subjects*

In the preliminary test, participants are 23 German learners in Chinese beginner's courses. Their age ranges from 21 to 67, there are 12 females and 11 males. The test was carried out at the very beginning of the course before they began to learn Chinese. In the main experiment, subjects are 17 German adults, including 7 females and 10 males, who had learned Chinese for more than 12 months with one two-hours' lesson each week. Their age range is between 18 and 55. All the subjects have no speech or hearing disorders.

In the production study, another four Mandarin speakers are also included. The ages of the two female and two male Mandarin speakers are between 24 and 43, they are native Chinese speakers with standard pronunciation.

### *2.2. Stimuli and procedure of the perception test*

The stimuli for the preliminary identification test include 184 (46 different syllables  $\times$  4 tones) monosyllabic words uttered by a female professional Chinese native speaker. Because most of the participants were not acquainted with Mandarin tones, instructions were provided for about 45 minutes until all of them had acquired the knowledge of the tone contours with some practicing exercises. They were then asked to choose the most probable lexical tone for the syllable they had heard.

In the main experiment, the stimuli for the perceptual study consist of 48 (12 different syllables  $\times$  4 tones) monosyllabic words uttered by the same professional speaker. Just like the preliminary experiment, the listeners were asked to attach a tone mark to the syllable in the perceptual test.

For both preliminary and main experiments, pinyin transcription was presented in the test paper. The syllables were arranged in a random order of tones, so that no tonal rhythm existed in the perception process. All the sound files were played at a sampling rate of 44.1 Hz. The stimuli were presented at a comfortable listening level.

### *2.3. Stimuli and procedure of the production test*

The reading material includes 72 (18 different syllables  $\times$  4 tones) monosyllables. These 18 syllables are designed so as to cover different syllable structures and different vowel types.

The speech samples were recorded using a digital tape recorder (Tascam), and a low noise microphone (Sennheiser MD421). The recordings were digitalized at sampling rate of 44.1 kHz. The recording gain level was kept similar among the subjects with no clipping.

The 72 monosyllables in pinyin transcription and with tones were presented to speakers in a randomized order for reading.

#### 2.4. 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 are so ambiguous, it is difficult to put them into a certain category. Four Chinese native speakers were organized to evaluate the production. If there is any disagreement, the decision favors the majority. Then the acoustic analysis was carried out with PRAAT (BOERSMA, WEENINK) software on the computer.

To accommodate the pitch range differences among female and male, German and Chinese speakers,  $f_0$  was normalized for each speaker across four tones. The  $f_0$  values obtained from each speaker were converted to their logarithms, using a formula commonly adopted for such purposes (WANG, 2003):

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

where  $H$  and  $L$  are the highest and lowest  $f_0$  for a given speaker, and  $X$  is any given point of a pitch contour. The output ( $T$ ) is a value between 0 and 5, which is similar to the 5-point pitch scale for Mandarin tones proposed by CHAO (1968).

To adjust for differences in speaking rate, duration was normalized per tone across speakers. The longest pitch contour was first determined; all other pitch contours for that tone were then stretched by interpolation.

The  $f_0$  of each utterance was estimated at five positions of vocalic segments (0% (initial), 25%, 50%, 75%, 100% (final)), so that the vocalic segment was divided into four consecutive sections of equal duration to permit a componential analysis of  $f_0$  changes within tones.

### 3. Results

#### 3.1. Preliminary identification experiment

The accuracy of the tonal identification ranges from 7% to 73% of beginners. Subjects are divided into 3 groups according to exposure degree to Mandarin Chinese: (1) *Subjects with little exposure* who have never heard Chinese spoken before; (2) *Subjects with average exposure* who have had some contact with Chinese colleagues or friends; (3) *Subjects with high exposure* who have made trips to China. They are further divided into three groups according to age: (1) *Age Group 1* are learners from 21–35; (2) *Age Group 2* are learners from 36–50; and (3) *Age Group 3* are learners from 51–67. The results can be observed in Fig. 2.

The results show that not all German beginners are “deaf” to Mandarin tones. A further investigation has revealed that listeners who have high exposure to Chinese can achieve an accuracy over 45% even before they begin to learn Chinese, those who have little exposure can only identify less than 25% of

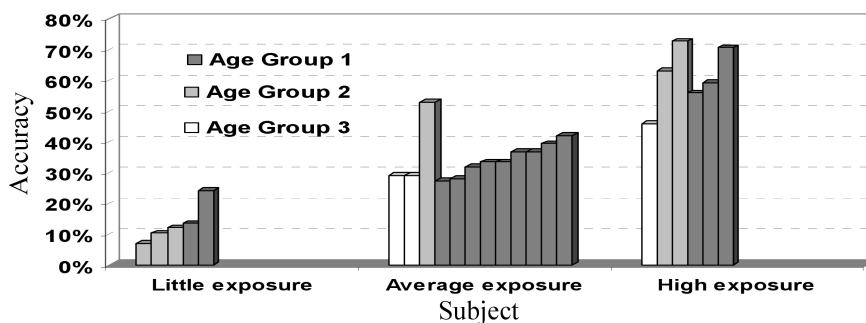


Fig. 2. The identification accuracy of Mandarin tones by German beginners.

the tones correctly. Actually the chance of accurate identification is about 25% by choosing one tone category out of four. It means subjects with less than 25% accuracy can hardly discriminate the lexical tones, their tonal confusion also exhibits no regular patterns, and all confusions among these 4 tones are possible. We can conclude that exposure to Mandarin Chinese can facilitate tonal identification. But whether there is any discrepancy in perception and production between German learners of intermediate level and Chinese native speakers is the focus of our main experiment.

### 3.2. The relationship between perception and production

A comparison of perception and production confusion patterns is shown in Fig. 3. The perception test includes 48 tokens, while the production test includes 72 tokens of each speaker, so that the accuracy of perception and production is represented in terms of percentage here. As shown in the figure, the accuracy of tone perception is highly correlated [ $r = 0.83$ ] with the tone production, and the correlation is significant [ $p < 0.0001$ ]

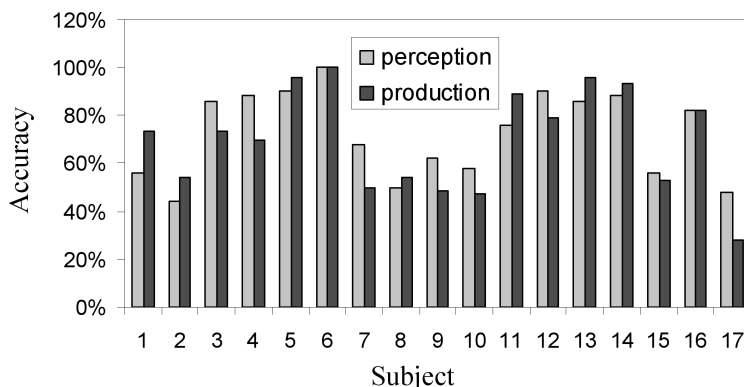


Fig. 3. The relationship between perception and production.

### 3.3. Perception experiment

For the perceptual identification tasks, there are 48 tokens from 17 speakers, resulting in a total of 816 tokens (48 tokens  $\times$  17 listeners), with 204 tokens for each tone. Table 1 shows the perceptual pattern of tones.

**Table 1.** Perception results of German learners.

Target	Perceived			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	<b>171 (84%)</b>	15 (7%)	8 (4%)	10 (5%)
Tone 2	28 (14%)	<b>114 (56%)</b>	46 (22%)	16 (8%)
Tone 3	18 (9%)	43 (21%)	<b>132 (65%)</b>	11 (5%)
Tone 4	15 (7%)	14 (7%)	4 (2%)	<b>171 (84%)</b>

Tone 1 and tone 4 are identified more accurately than tone 2 and tone 3. The accuracy of tone 1 and tone 4 are both 84%, while the accuracy the tone 2 and tone 3 are 56% and 65% respectively.

Because a high-level tone (tone 1) is not a normal pattern for German native speakers, it can be easily identified. A high-falling tone appears only in final position of a declarative sentence in German, tone 4 becomes unfamiliar to German listeners when it occurs in isolation, and could be easily identified as a monosyllable in the experiment.

We have also noticed in our experiment that if a tone 3 is a modal tone, this tone 3 is more frequently misperceived as tone 2 than the correspondent tone 2 is incorrectly perceived as tone 3. But a creaky and breathy tone 3 as shown in Fig. 4 is usually correctly perceived as tone 3. As the professional female speaker involved in the perception test is characterized by having a glottalized period in the middle of many third tones (DING *et al.*, 2006a), the accuracy of tone 3 perception (65%) is thus higher than tone 2 (56%), and tone 2 is more frequently misperceived as tone 3 (22%) than tone 3 is incorrectly perceived as tone 2 (21%). That means glottalization serves as a cue to the identification of tone 3 for the German subjects.

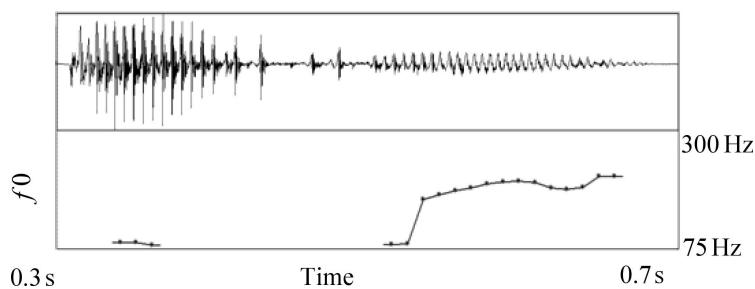


Fig. 4. Waveform and pitch contour of creaky 3rd tone “da3”.

## 3.4. Production experiment

The overall results of the production evaluation are presented in Table 2. It is observed that the production of tone 4 achieves the best result (83%), tone 2 (78%) the second, tone 1 the third (76%), the worst is tone 3 (46%). A high-falling tone is although not usual in isolation; it does exist in German speech (RATHCKE, HARRINGTON, 2007). This falling tone is less confused with other tones; the subjects can correctly produce it with some efforts.

**Table 2.** Production results of German learners.

Target	Identified by Chinese native speakers			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	<b>233 (76%)</b>	54 (18%)	8 (2%)	11 (4%)
Tone 2	38 (12%)	<b>238 (78%)</b>	17(6%)	13 (4%)
Tone 3	5 (2%)	158 (51%)	<b>140 (46%)</b>	3 (1%)
Tone 4	4 (1%)	44 (15%)	4 (1%)	<b>254 (83%)</b>

Although there are different syllable structures and there exist some speaking rate difference among speakers, tone contours will not be systematically changed, which is also demonstrated in (XU, 1998). Therefore we can plot the  $f_0$  contours of all types of syllables with the average values of the speakers. The confusion of misproduced tones will be analyzed with the illustration of the pitch contours and pitch height of the German speakers compared with the native norms. The pitch contours of all subjects are normalized for  $f_0$  and duration. Pitch values are represented on a 5-point scale as  $T$  values.

As shown in the following figures, the German speakers display a different pitch contour and pitch height than the native Mandarin speakers, which serves as an explanation for the mispronunciation of the tones.

The high-level tone (tone 1) is shown in Fig. 5, the  $T$  values (derived from formula (1)) remain relative constant for Mandarin speakers with  $\Delta T = 0.27$  from beginning to the end,  $f_0$  contour of German speakers has a little rise with  $\Delta T = 0.35$ , so that some high-level tones (18%) would be identified as rising tones. Moreover, the German speakers keep their tone level a little lower than the native speakers with an  $f_0$  range of  $\Delta T = 1.1$ . It indicates that German speakers' production of level tone is influenced by their non-tonal intonation system: it is unusual to start a tone so high; it is also difficult to keep the tone level than to raise the pitch, as if they were reading the monosyllabic words in an enumerating way.

The mid-rising tone (tone 2) is presented in Fig. 6, the starting point for native speakers is 3, but for the German speakers is 2.2. Although  $f_0$  range of native speakers ( $\Delta T = 1.7$ ) is smaller than the German speakers ( $\Delta T = 2.0$ ), the pitch contours are quite different. The native speakers show a straightforward rising of the pitch, while the German speakers have a visually spanned continuum

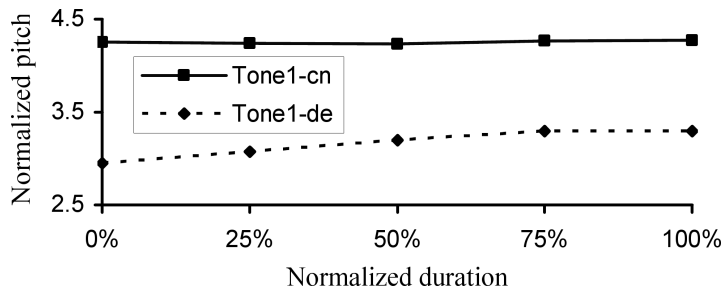


Fig. 5. Pitch contours of Tone 1 ('de' means German speakers, 'cn' means Chinese speakers).

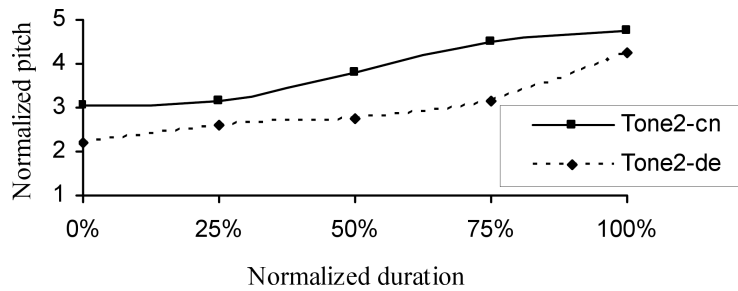


Fig. 6. Pitch contours of Tone 2 ('de' means German speakers, 'cn' means Chinese speakers).

of lower pitch before rising, which leads to the unambiguous production of tone 2 as tone 3.

The low-falling-rising tone (tone 3) is illustrated in Figure 7, both German and native speakers start at approximately the same point, German speakers have a higher ending (4.3) than Chinese native speakers (3.2), so that they display a larger  $f_0$  range ( $\Delta T = 2.4$  in contrast to  $\Delta T = 2.0$ ). But the native speakers have a deeper falling ( $T = 1.3$ ) than the German speakers ( $T = 2.0$ ). Lack of the deep falling, the pitch contour of tone 3 resembles that of tone 2, which accounts for the worst production of tone 3. More tone 3 are mispronounced as tone 2 (51%) than are rightly pronounced as tone 3 itself (46%).

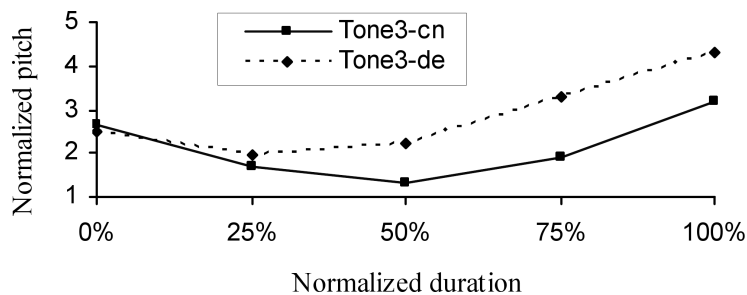


Fig. 7. Pitch contours of Tone 3 ('de' means German speakers, 'cn' means Chinese speakers).



The high-falling tone is presented in Fig. 8, both German and native speakers demonstrate falling pitch contours. But the native speakers have a much larger  $f_0$  range ( $\Delta T = 3.4$  in contrast to  $\Delta T = 1.4$ ), German learners just truncate the highest part and simply end earlier, as is observed in comparison to English (GRABE, 1998). Because the average pitch range is not large enough, the production of tone 4 can only achieve 83% accuracy.

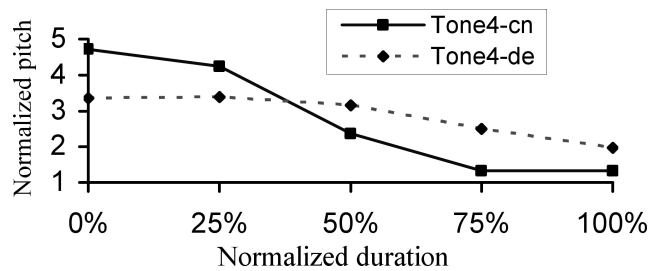


Fig. 8. Pitch contours of Tone 4 ('de' means German speakers, 'cn' means Chinese speakers).

It seems that the German subjects tend to apply their intonation system to the reading of Chinese tones. In summary, to start tone 1 and tone 4 high proves to be unusual for them, to raise the pitch right away for tone 2 and to have a deep lowering of tone 3 are the main difficulties for the German subjects.

#### 4. Conclusion

The variation of Chinese lexical tones in words and sentences is very complicated (SHIH, 2000) and difficult to learn. With a preliminary test, however, we suggest that the tonal variation at lexical level in isolation can be learned by German adult speakers if they have enough exposure to Chinese language. The results also present a high correlation between the accuracy of perception and production by intermediate learners. A further examination of pitch height and pitch contour reveals that German learners have difficulty in changing pitch movement rapidly in a very short time because of typical gradual pitch movement in German, which leads to some confusion in production.

The aim of this investigation is to provide implications and statistical guidelines to build an interactive Mandarin Chinese learning system. An effective learning strategy should be supported by existing technological progress such as speech recognition and speech synthesis with audio-visual feedback. The reproduction of the learners should be recognized and analyzed in terms of acoustic parameters. The difference will be presented to learners visually (by comparing pitch contours) and audibly (by listening to the sound); the correct production can also be synthesized with the voice of the speaker for imitation. In many repeated sessions, the brain of the learners can overcome the constraints of mother tongue and develop a system to recognize and produce Chinese tonal categories

successfully. Further investigations should be carried out into disyllabic and trisyllabic words as well as into some typical sentence structures.

### Acknowledgments

We would like to thank two anonymous reviewers for their valuable comments. The first author is partly supported by the fund of Tongji University for promoting academic exchange with Germany.

### References

1. BOERSMA P., WEENINK D., <http://www.praat.org>.
2. CHAO Y.R. (1968), *A grammar of spoken Chinese*, University of California Press, Berkeley.
3. DING H., JOKISCH O., HOFFMANN R. (2006a), *The effect of glottalization on voice preference*, Proceedings of Speech Prosody 2006, HOFFMANN R., MIXDORFF H. [Eds.], pp. 851–854.
4. DING H., JOKISCH O., HOFFMANN R. (2006b), *F0 analysis of Chinese accented German speech*, Proceedings of ISCSLP 2006, HUO Q. *et al.* [Eds.], pp. 49–56.
5. DING H., JOKISCH O., HOFFMANN R. (2007), *Perception and analysis of Chinese-accented German vowels*, Archives of Acoustics, **32**, 1, 89–100.
6. GRABE E. (1998), *Pitch accent realization in English and German*, Journal of Phonetics, **26**, 2, 129–143.
7. JILKA M. (2005), *Exploration of different types of intonational deviations in foreign-accented and synthesized speech*, Proceedings of Interspeech 2005, pp. 2393–2396.
8. LEATHER J. (1990), *Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers*, New sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second-Language Speech, LEATHER J., JAMES A. [Eds.], pp. 72–89, University of Amsterdam.
9. PEABODY M., SENEFF S., WANG C. (2004), *Mandarin tone acquisition through typed interactions*, Proceedings of InSTIL/ICAL Symposium: NLP and Speech Technologies in Advance Language Learning Systems 2004, pp. 173–176.
10. RATHCKE T., HARRINGTON J. (2007), *The phonetics and phonology of high and low tones in two falling f0-contours in standard German*, Proceedings of Interspeech 2007, pp. 982–985.
11. SHIH C. (2000), *A Declination model of Mandarin Chinese*, [in:] *Intonation: analysis, modelling and technology*, Botinis A. [Ed.], pp. 243–268, Kluwer, Dordrecht.
12. WANG Y. (2003), *Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training*, Journal of the Acoustical Society of America, **113**, 2, 1033–1043.
13. XU Y. (1998), *Consistency of tone-syllable alignment across different syllable structures and speaking rates*, Phonetica, **55**, 4, 179–203.