Perception and Production of Mandarin Tones by German Speakers

Hongwei Ding[1], *Oliver Jokisch*[2] *and Rüdiger Hoffmann*[2]

[1]School of Foreign Languages, Tongji University, China
[2]Laboratory of Acoustics and Speech Communication, Dresden University of Technology, Germany
hongwei.ding@tongji.edu.cn, {oliver.jokisch;ruediger.hoffmann}@ias.et.tu-dresden.de

Abstract

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. Besides common mistakes found in previous literature, German speakers have much smaller f0 range than Chinese native speakers. Findings can provide implications for cross language studies and teaching practices.

Index Terms: Mandarin tones, Mandarin and German

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 [8]. A number of cross-linguistic studies have been conducted to examine if and how non-tonal speakers process the tones differently [10, 13]. These studies, however, were mainly focused on English learners [8, 13], the examination of German learners has received little attention. This paper aims to extend the investigation to the 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 [7]; German truncates falling accents, falling tones become gradual and simply end earlier in comparison to British English [6]. 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 [2] 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.

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 also provided implications [4, 5], 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.

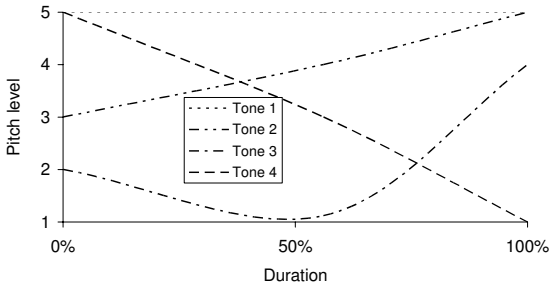

Figure 1: Canonical pitch forms of 4 tones produced in isolation

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 with higher levels. 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 about 12 months with two hours of courses every week. Their age range is between 18-55. All the subjects have no speech, language, 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 x 4 tones) monosyllabic words uttered by a female professional Chinese native speaker. Because most of the participants were not acquainted with Mandarin tones, in-

structions were provided for about 45 minutes until all of them had acquired the knowledge of the tone contours with some practising exercises. They were 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 x 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 44.1Hz. 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 x 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 44.1kHz. The recording gain level was set to ensure similar recording level 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 [1] software on the computer.

To accommodate the pitch range differences among female and male, German and Chinese speakers, f0 was normalized for each speaker across four tones. The f0 values obtained from each speaker were converted to their logarithms, using a formula commonly adopted for such purposes [10]:

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

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 [2].

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 f0 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 f0 changes within tones.

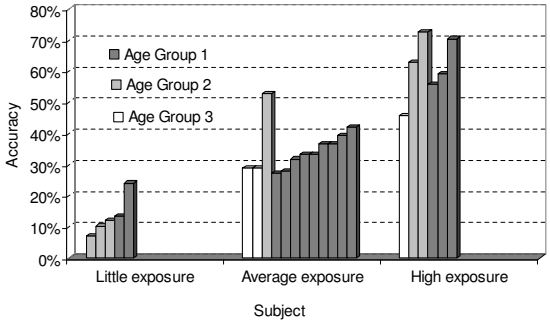

Figure 2: The identification accuracy of Mandarin tones by German beginners

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 Figure 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 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, all confusion 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 more advanced 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 Figure 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 [$p < 0.0001$].

3.3. Perception experiment

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

Tone 1 and tone 4 are identified more accurately than tone

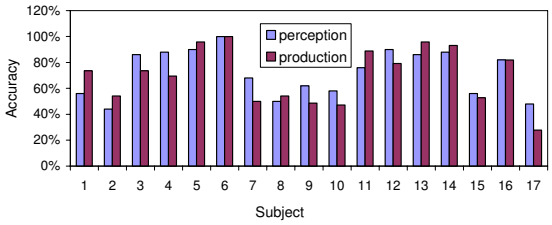

Figure 3: The relationship between perception and production

Table 1: *Perception results of German learners*

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

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.

It is noted 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 Figure 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 [3], the accuracy of tone 3 perception (65%) is thus higher than tone 2 (56%), and tone 2 is more frequently misperceived as tone 3 (46%) than tone 3 is incorrectly perceived as tone 2 (43%). That means glottalization serves as a cue to the identification of tone 3 for the German subjects.

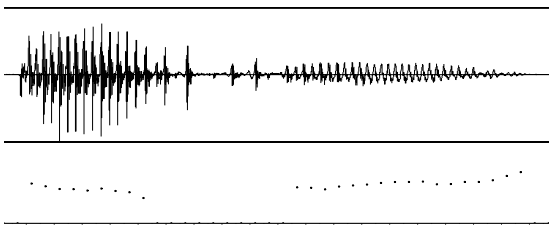

Figure 4: Waveform and pitch contour of creaky 3rd tone "da3"

3.4. Production experiment

The overall results of the production evaluation is 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, but it does exist in German speech [11]. 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	**233 (76%)**	54 (18%)	8 (2%)	11 (4%)
Tone 2	38 (12%)	**238 (78%)**	17(6%)	13 (4%)
Tone 3	5 (2%)	158 (51%)	**140 (46%)**	3 (1%)
Tone 4	4 (1%)	44 (15%)	4 (1%)	**254(83%)**

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 [14]. Therefore we can plot the f0 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 f0 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 Figure 5, the T values remain relative constant for Mandarin speakers with ΔT=0.27 from beginning to the end, f0 contour of German speakers has a little rise with ΔT=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 f0 range of ΔT=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.

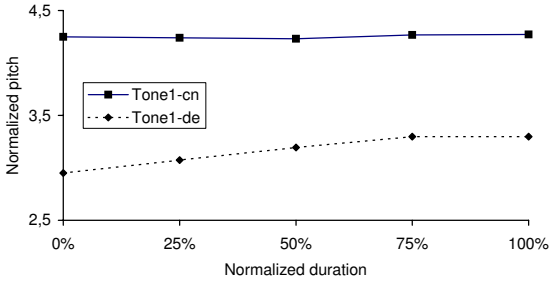

Figure 5: Pitch contours of Tone 1 ('de' means German speakers, 'cn' means Chinese speakers)

The mid-rising tone (tone 2) is presented in Figure 6, the starting point for native speakers is 3, but for the German speakers is 2.2. Although f0 range of native speakers (ΔT=1.7) is smaller than the German speakers (ΔT=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 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 f0

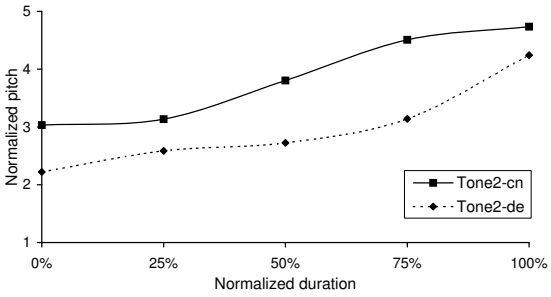

Figure 6: Pitch contours of Tone 2 ('de' means German speakers, 'cn' means Chinese speakers)

range (ΔT=2.4 in contrast to ΔT=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%).

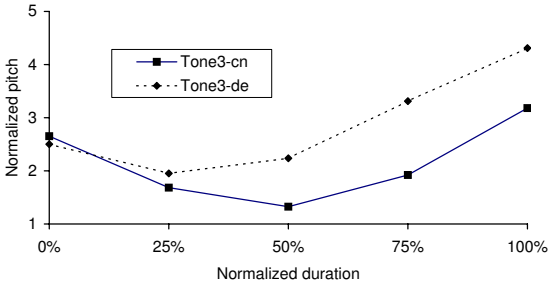

Figure 7: Pitch contours of Tone 3 ('de' means German speakers, 'cn' means Chinese speakers)

The high-falling tone is presented in Figure 8, both German and native speakers demonstrate falling pitch contours. But the native speakers have a much larger f0 range (ΔT=3.4 in contrast to ΔT=1.4), German learners just truncate the highest part and simply end earlier, as is observed in comparison to English [6]. Because the average pitch range is not large enough, the production of tone 4 can only achieve 83% accuracy.

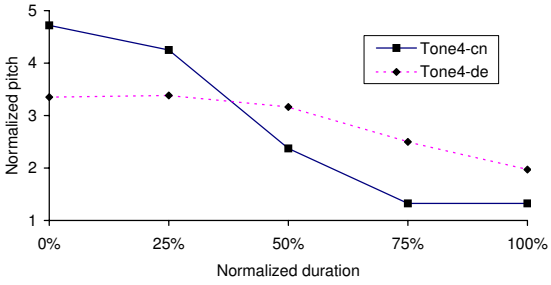

Figure 8: Pitch contours of Tone 4 ('de' means German speakers, 'cn' means Chinese speakers)

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 errors for these German subjects.

4. Conclusion

The variation of Chinese lexical tones in words and sentences is very complicated [12] 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. 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 for language teaching and statistical guidelines to build an interactive Mandarin Chinese learning system, which is also jointly attempted by [9]. With the help of speech technology, the pronunciation mistakes can be presented to learners visually (by comparing pitch contours) and audibly (by listening to the sound), 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 for disyllabic and trisyllabic words as well as for some typical sentence structures.

5. References

[1] Boersma, Paul & David Weenink. http://www.praat.org.

[2] Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. University of California Press, Berkeley.

[3] Ding, Hongwei, Oliver Jokisch & Rüdiger Hoffmann. 2006. *The effect of glottalization on voice preference*. In Proc. Speech Prosody 2006, 851-854.

[4] Ding, Hongwei, Oliver Jokisch & Rüdiger Hoffmann. 2006. *F0 analysis of Chinese accented German speech*. In Proc. ISCSLP 2006, Huo, Qiang et al. (ed.), 49-56.

[5] Ding, Hongwei, Oliver Jokisch & , Rüdiger Hoffmann. 2007. *Perception and analysis of Chinese-accented German vowels*. In J. Archives of Acoustics, 32(1), 89-100.

[6] Grabe, Esther. 1998. *Pitch accent realization in English and German*. In Journal of Phonetics, 26(9), 129-143.

[7] Jilka, Matthias. 2005. *Exploration of different types of intonational deviations in foreign-accented and synthesized speech*. In Proc. Interspeech 2005, 2393-2396.

[8] Leather, Jonathan. 1990. *Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers*. In Leather, Jonathan & Allan James (ed.). New sounds 90, 72-97. University of Amsterdam.

[9] Mixdorff, Hansjörg et al. 2009. *Towards a computer-aided pronunciation training system for German learners of Mandarin*. In Proc. SLaTE 2009

[10] Peabody, Mitchell, Stephanie Seneff & Chao Wang. 2004. *Mandarin tone acquisition through typed interactions*. InSTIL/ICALL 2004, 173-176.

[11] Rathcke, Tamara & Jonathan Harrington. 2007. *The phonetics and phonology of high and low tones in two falling f0-contours in standard German*. In Proc. Interspeech 2007, 982-985.

[12] Shih, Chilin. 2000. *A Declination model of Mandarin Chinese*. In Botinis, Antonis (ed.) Intonation: analysis, modelling and technology, 243-268. Kluwer: Dordrecht.

[13] Wang, Yue. 2003. *Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training*. In J. Acoust. Soc. Am., 113(2), 1033-1043.

[14] Xu, Yi. 1998. *Consistency of tone-syllable alignment across different syllable structures and speaking rates*. In Phonetica, 55, 179-203.