Mandarin Tone Perception and Production by German Learners

Hussein Hussein [1,2], Hue San Do [1], Hansjörg Mixdorff [1], Hongwei Ding [3], Qianyong Gao [4], Guoping Hue [4], Si Wei [4] and Zhao Chao [4]

[1] Department of Computer Sciences and Media, Beuth University of Applied Sciences, Berlin, Germany

[2] Laboratory of Acoustics and Speech Communication, Dresden University of Technology, Dresden, Germany

[3] School of Foreign Languages, Tongji University, Shanghai, P.R. China

[4] Department of EEIS, University of Science and Technology of China, Hefei, Anhui, P.R. China

{hussein, hsdo, mixdorff}@beuth-hochschule.de,

hongwei.ding@tongji.edu.cn,{qygao, gphu, siwei, chaozhao}@iflytek.com

Abstract

This paper reports on the continued activities to develop a computer-aided phonetic learning system for German learners of Mandarin. In the current study we used a corpus which consists of disyllables and sentences that were produced by first-year German students and native speakers of Mandarin. Half of the German group had attended an additional phonetic seminar. The first experiment was a tone identification test of Mandarin disyllables and production test thereof, followed by a sentence production experiment in which the subjects produced Chinese sentences in both reading and speaking mode. Prosodic features (rhythmic and intonational) of Mandarin tones between German learners and Chinese native speakers were compared. The results based on the annotations of an expert and analysis of prosodic features of Mandarin tones show that German learners who attended the weekly seminar were able to pronounce the tones better than those who did not attended the weekly seminar.

Index Terms: Computer-Aided Language Learning (CALL), Mandarin tones, prosodic analysis

1. Introduction

The growing interest in speaking a foreign language in a globalized world stimulates activities towards computer-aided language learning (CALL). CALL is a tool to facilitate individualized language learning and pronunciation training, for example [1]. The pronunciation training might be the most difficult to be transferred to a computer because providing useful and robust feedback on learner errors is far from being a solved problem [2]. In this paper we report on the on-going development of a Mandarin training system for German learners within a three-year project funded by the German Federal Ministry of Education and Research [2][3].

It is commonly known that Mandarin is a tone language and hence the tonal contour of a syllable changes its meaning [4]. The most important acoustic correlate of tone is *F0*. Mandarin has four syllabic tones and a neutral tone in unstressed syllables. In citation forms of monosyllabic words the tonal patterns are very distinct (see Figure 1), but when several syllables are connected, *F0* contours observed vary considerably due to tonal coarticulation. We observed that the acquisition of tonal patterns of poly-syllabic words is much more difficult than of mono-syllabic words [2].

German is a non-tone language. Mandarin differs from German significantly on the segmental as well as the suprasegmental level and poses a number of problems to the German learners.

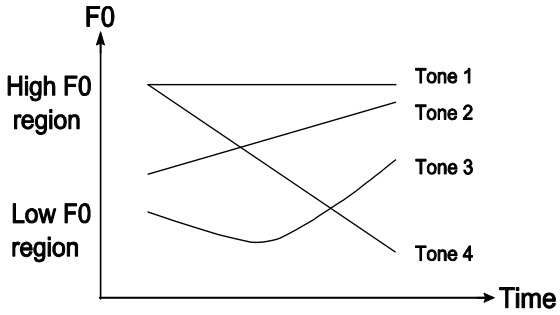

Figure 1: *Prototypical F0 contours of Mandarin tones.*

When teaching Chinese tones to German learners we assume that in order to produce a certain tone contour correctly, one must be able to discriminate and identify them in the first place. We therefore performed a tone analysis of the perception and production tests of disyllables. Furthermore, the correctness of syllable tone was compared between reading and speaking mode and between German learners who did and did not attend an additional phonetic seminar. The rhythmic and intonational features of the Mandarin tones for native speakers and German learners of Mandarin were compared to identify the differences and similarities.

2. Experiment Method

2.1. Corpus Design and Data Collection

The data used in this experiment consists of recordings from 13 first year German students of Chinese Studies at the East Asian Seminar of Free University Berlin (*FUB*) (henceforth "*DE*") and six native speakers of Mandarin from Tongji University, Shanghai, China (henceforth "*CN*"). The data was recorded with a sampling frequency of 16 kHz and a resolution of 16 bit. In addition to the regular classes of Mandarin language training (eight hours per week), seven of the German students had attended a weekly seminar of two hours as additional training and used a computer-aided

phonetic pronunciation training system at home. About two-third of the seminar was dedicated to phonetic, one-third to grammar and conversation exercises. The phonetic exercises comprised discrimination, identification and imitation of mono- and disyllables, contrastive exercises with minimal pairs of differing initials or finals as well as reading from the text book, constantly monitored and corrected by the teacher. At the time of the recording the students had completed 12 weeks of Mandarin language training.

The data collected from the German learners of Mandarin consist of four parts:

1. Identification of disyllables (henceforth "*Perception*"): In the first part, the students first listened to ten Chinese disyllables and identified the tone combinations by showing prepared tone cards. These cards had the tones marked as tone 1, 2, 3, 4 and 0 for the neutral tone.

2. Production of disyllables (henceforth "*Production 1*"): In the second part, the same ten Chinese disyllables were presented to the students in Pinyin transcription on a computer screen and they had to read them aloud (in Pinyin transcription the tone is shown by diacritical symbols). The disyllables chosen for both identification and production mode consisted of real Chinese words with different tonal combinations. The students who had attended the weekly seminar (henceforth "*WS*") had encountered the words in the pronunciation training system before whereas the rest of the testing group (henceforth "*WOS*") had not.

3. Production of sentences (reading mode) (henceforth "*Production 2*"): In the third part, the students were asked seven questions in Chinese and they read seven sentences as answers aloud. Thereafter, the students asked three questions. The sentences and questions were presented to them on a computer screen both in Chinese characters and Pinyin transcription.

4. Production of sentences (speaking mode) (henceforth "*Production 3*"): Subsequently, the same questions were asked again in Chinese, but this time only a German cue (consisting of two or three German words) was presented on the screen and the students had to give the same Chinese answers and ask the same questions as in part 3.

Parts 3 and 4 consist of 20 sentences, each sentence contains both monosyllabic and disyllabic words, with a minimum of three and a maximum of nine syllables. The sentences were chosen from the textbook "The New Practical Chinese Reader I" which the students use in their regular Chinese class. Therefore, all students had encountered the sentences before.

The data collected from the six Chinese native speakers (three females and three males) consists of ten sentences. The same ten sentences as *Production 2* were presented to the Chinese Native Speakers in Chinese characters and then read aloud without preceding questions.

2.2. Data Evaluation

The data collected were annotated and analyzed as follows:

1. Expert (German teacher of Mandarin): The expert (the second author) listened to the data of the production tests 1-3 (disyllables, sentence reading and speaking mode) several times and annotated the tones she perceived using the numbers 1-4 and 0 for the neutral tone.

2. Ten native speakers of Mandarin from Tongji University, Shanghai, China listened to the German data of the *Production 2* and *Production 3* (sentence reading and speaking mode) and rated degree of foreign accent and intelligibility on a scale from one to five, five being the best score, i.e. native-like competence. The native speakers were between 22 and 28 years of age.

3. The data from all four tests were evaluated statistically regarding correctness of tone whether there was a difference between seminar students (*WS*) and non-seminar students (*WOS*). For the disyllable identification and production test the correctness of tone and common confusion partners tone-wise were analyzed. As for the sentence production tests (*Production 2* and *Production 3*), tone correctness was firstly compared between reading and speaking mode and secondly tone correctness between *WS* and *WOS* students.

Furthermore, the German data from *Production 2* and *Production 3* was analyzed regarding its prosodic features like syllable duration and parameters of *F0* contour of Mandarin tones and was compared to the data collected from Chinese native speakers.

2.3. Data Analysis

The rhythmic and intonational features of the Mandarin tones on the syllable level were calculated for a contrastive analysis between German learners and native speakers of Mandarin. Therefore, the German and Chinese data were forced-aligned on the syllable and phone-levels using the Automatic Speech Recognition (ASR) system in a forced alignment mode. The used ASR system is part of an automated proficiency test of Mandarin [5]. The ASR system used acoustic models which were trained on data from native speakers of Mandarin. The label files from the forced alignment were converted to the *Praat* TextGrid format [6] and combined in a single TextGrid file containing syllable and phone labels. The boundaries of syllables and phones were then hand-corrected.

The *F0* contours were calculated using the *Praat* algorithm with a step of 10msec and different standard settings of the minimum and maximum parameters of *F0* for male (100 and 350 Hz) and for female speakers (120 and 450 Hz). The *F0* contours were corrected using the *Praat Pitch-Editor*.

The *F0* contour reflects the tone on the syllable level. In order to reduce the variation of the speaker's *F0* range among female and male speakers and between German learners and native speakers of Mandarin, the *F0* contours were normalized for each speaker. The normalized *F0* contour was calculated as in the following formula [7]:

$$Y = 5 \frac{\log(X) - \log(L)}{\log(H) - \log(L)} \quad (1)$$

where Y is a normalized fundamental frequency value, X is the raw fundamental frequency value, H and L are the highest and lowest *F0* value for a given speaker. The value of Y is between zero to five, which is similar to the five-point pitch scale for Mandarin tones. No round-off process was used in the normalization of *F0* contour.

3. Results

The annotations produced by the expert were used as a reference for judging the correctness of tones produced by German learners of Mandarin. Accent and intelligibility of entire utterances were evaluated by ten native speakers of Mandarin from Tongji University, Shanghai.

3.1. Analysis of Correctness of Mandarin Tone

We performed a tone analysis of the perception and production tests of disyllables. Furthermore, the correctness of syllable tone was compared between reading and speaking mode (*Production 2* and *Production 3*) and between *WS* and *WOS*.

3.1.1. Perception and Production 1 Tests

Table 1 shows the correctness of tone when the German students identified disyllables (*Perception*). Tone 2 has the lowest correctness (< 50%) and, as expected, is most frequently confused with tone 3 which also has a relatively low correctness but is mostly confused with Tone 4 and less with tone 2. Tone 4 has the highest correctness which is probably related to the similar tone contour of German syllables carrying stress. Tone 1 is mostly confused with tone 2 and 4, but rarely with tone 3. The neutral tone is mostly confused with tone 4. When comparing testing groups, *WS* clearly perceive all tones more accurately than *WOS*.

When looking at the correctness of disyllables produced by German students (*Production 1*) the results are quite different (see table 1). Tone 1, 2 and 4 have similar correctness (> 70%) whereas tone 3 is half of the time pronounced as tone 2. This problem is common among German learners who have difficulties with the falling-rising third tone. Unlike in the identification test, German students confuse both tone 1 and 4 more often with the neutral tone.

Table 1. *Correctness and confusion partners of tones in the Perception and Production 1 tests (in %).*

	Perception					*Production 1*				
Tone	T1	T2	T3	T4	T0	T1	T2	T3	T4	T0
T1	67	15	4	13	0	71	6	4	6	13
T2	12	49	33	6	0	10	71	10	3	6
T3	0	17	56	25	2	4	50	38	4	2
T4	9	8	11	71	2	5	9	5	71	9
T0	8	15	0	23	54	15	0	0	23	62

3.1.2. Production 2 and Production 3 Tests

Table 2 shows the correctness rate of tone production in sentences for reading (*Production 2*) and speaking mode (*Production 3*). Surprisingly, correctness rates for *Production 2* are only slightly better than for *Production 3*. When Pinyin transcription or Chinese characters are presented, the neutral tone has the highest correctness, followed by tone 2 which is more often confused with tone 1 and 3. When only a German cue is presented, however, the neutral tone is produced more often. Consistent with the production of disyllables, tone 1 is often confused with tone 4, and tone 3 is commonly pronounced as tone 2. Notably, when reading aloud complete sentences in contrast to mere disyllables, German students more often produce a neutral tone for all four tones, even more so when neither Pinyin transcription nor Chinese characters are available. The neutral tone is similar to unstressed syllables in German sentence intonation and apparently interferes with the tone pronunciation in Chinese sentences which has been observed in class as well.

When looking separately at the correctness between *WS* and *WOS* for the *Production 2* (see table 3), *WS* clearly perform better. Correctness for *Production 2* is only slightly higher than for *Production 3* for both *WS* and *WOS*. *WS* pronounce tone 1 less correctly when they speak freely (table 4). *WOS* pronounce tone 3 less correctly and confuse it more often with tone 2 and 0 in the speaking mode. Without Pinyin and Chinese characters, *WOS* pronounce tone 4 as the neutral tone. Learning effects have to be taken into account since *Production 3* consists of the same sentences as *Production 2*.

Table 2. *Correctness and confusion partners of tones in the Production 2 and Production 3 tests (in %).*

	Production 2					*Production 3*				
Tone	T1	T2	T3	T4	T0	T1	T2	T3	T4	T0
T1	69	0	0	8	23	63	2	2	13	19
T2	10	70	10	3	7	5	69	8	5	13
T3	6	18	62	3	12	5	20	55	1	18
T4	6	2	1	64	27	7	1	1	60	32
T0	9	0	0	5	86	4	2	0	5	89

Table 3. *Correctness and confusion partners of tones for WS and WOS for Production 2 (in %).*

	WS					*WOS*				
Tone	T1	T2	T3	T4	T0	T1	T2	T3	T4	T0
T1	82	0	0	7	11	54	0	0	8	38
T2	6	81	8	1	4	15	57	11	6	10
T3	6	16	69	0	8	6	19	52	7	15
T4	4	0	1	78	17	8	4	1	49	38
T0	9	0	0	6	86	10	0	0	3	87

Table 4. *Correctness and confusion partners of tones for WS and WOS for Production 3 (in %).*

	WS					*WOS*				
Tone	T1	T2	T3	T4	T0	T1	T2	T3	T4	T0
T1	71	4	4	14	7	54	0	0	13	33
T2	1	82	7	3	8	9	55	8	7	20
T3	3	17	67	1	12	7	25	42	1	25
T4	4	0	1	76	19	11	1	0	43	45
T0	0	4	0	7	89	10	0	0	3	87

3.2. Contrastive Analysis of Prosodic Features of Mandarin Tone

In order to identify the differences and similarities between native speakers and German learners of Mandarin the prosodic properties (rhythmic and intonational features) of the Mandarin tones on the syllable level were compared.

3.2.1. Comparison of Syllable Duration

The mean and standard deviation (SD) values of syllable duration depending on the Mandarin tones for *DE* of both *Production 2* and *Production 3* and *CN* are shown in Table 5. The table shows that the mean of syllable duration depending on the tone produced by the German learners is longer than the syllable duration of the native speakers of Mandarin. This indicates that learners of a language speak more slowly.

Table 5. *Mean and standard deviation of syllable duration depending on the tone for German Data (Production 2 and Production 3) and Chinese Data.*

Tone	*DE-Production 2*		*DE-Production 3*		*CN-Production 2*	
	mean	SD	mean	SD	mean	SD
T1	0.33	0.09	0.30	0.08	0.26	0.07
T2	0.30	0.07	0.29	0.06	0.26	0.06
T3	0.27	0.09	0.27	0.11	0.22	0.07
T4	0.32	0.09	0.32	0.12	0.28	0.06

DE show the same tendency concerning syllable duration as *CN* except for tone 4: tone 1 being the longest and tone 3 the shortest. German students tend to produce longer syllables when reading Pinyin transcription or Chinese characters than when presented with a German cue. Independent samples Kruskal-Wallis test shows that syllable duration is significantly different depending on the group (*WS*, *WOS* and *CN*), as well as the tone ($p < .001$).

3.2.2. Comparison of F0 Contour of Mandarin Tone

Table 6 shows the parameters of the normalized *F0* contours of syllables depending on the tones for Chinese native speakers, *WS* and *WOS* students. The Chinese native speakers were used as reference for evaluation of *WS* and *WOS*.

The mean value of *F0* contour of tone 1 by German learners is lower than by Chinese native speakers. But the *F0* range of tone 1 is greater by German learners. It indicates that the German learners are not able to start the tone with a high-level and to keep the same level of *F0* contour [7]. The *F0* range of tones 2, 3 and 4 by German learners is smaller than by native speakers of Mandarin. This indicates that the German learners are not able to raise the *F0* contour of the rising tone (tone 2) enough like the native speakers. It is difficult for German learners to change the *F0* contour of the falling-rising tone (tone 3) which usually confused with tone 2 [2][7]. The German learners are not able to start with a high-level and decrease the *F0* contour enough like the native speakers in the falling tone (tone 4).

The mean of *F0* contour of tone 1 by *WS* is greater than by *WOS*. But the *F0* range of tone 1 by *WS* is smaller which indicates that the *WS* students are able to keep the high-level of pitch for tone 1. *F0* range of tone 3 and tone 4 by *WS* is greater than by *WOS*. This indicates that the *WS* students could decrease and increase the *F0* contour of tone 3 and could start tone 4 with a high-level and keep a longer decline of *F0* contour more than German learners of *WOS*.

Table 6. *Mean, standard deviation and range of normalized F0 subcontour of syllables depending on the Mandarin tones for CN, WS and WOS (Production 2 and Production 3).*

	Tone	*F0* mean	*F0* SD	*F0* range
CN	Tone 1	2.83	0.28	1.13
	Tone 2	1.59	0.63	1.89
	Tone 3	1.71	0.62	1.87
	Tone 4	2.80	0.80	2.43
WS	Tone 1	2.80	0.32	1.18
	Tone 2	1.60	0.48	1.50
	Tone 3	1.69	0.44	1.39
	Tone 4	2.60	0.72	2.16
WOS	Tone 1	2.60	0.40	1.34
	Tone 2	1.73	0.50	1.52
	Tone 3	1.69	0.39	1.26
	Tone 4	2.17	0.53	1.69

These findings are consistent with the correctness rates in section 3.1: *WOS* produce perceptible more false tones than *WS* which is partially related to the insufficient rise and fall of *F0*. The Kruskal-Wallis test shows that SD and range of *F0* contour are significantly different between groups *WS*, *WOS* and *CN* ($p < .001$) whereas the *F0* mean value -as could be expected- is not. The same test shows significant differences for mean, SD and range of *F0* contour depending on the tone of the syllable ($p < .001$).

3.3. Comparison of Entire Utterance

The mean accent and intelligibility ratings are 3.21 and 3.68 for the *WS* group and 2.80 and 3.35 for the *WOS* group, respectively. The SD values of accent and intelligibility ratings are 0.82 and 0.65 for the *WS* group and 0.84 and 0.79 for the *WOS* group, respectively. The accent and intelligibility for *WS* are greater than by *WOS*. The high scores of the utterance-wise judgments could be related to the higher tonal accuracy produced by *WS* as described in the previous sections. Independent samples Mann-Whitney U-tests suggest that these differences are highly significant ($p < .001$ for both accent and intelligibility). A stronger accent does not necessarily impede communication. When generating phonetic exercise for Chinese learners it should be helpful find out more precisely which kind of tonal inaccuracy does impede intelligibility.

4. Conclusions

German learners at beginner level have difficulties identifying Mandarin tone in general, especially discriminating the second, third and the neutral tone. By tone production, tone 3 poses the biggest challenge, while tone correctness rates are slightly higher when Pinyin transcription and Chinese characters are provided. The lower correctness rate when confronted with mere German cues is probably related to the increased memory load, i.e. German learners cannot recall tone features accurately. These findings are in line with the results from *F0* contour analysis. The syllable duration depending on the tone by German learner is longer than by Chinese native speakers.

The results of all four tests show that Seminar students (*WS*) clearly performed better than non-seminar students (*WOS*). Further tests should verify if these results are related to an actual pronunciation problem (*WOS* are not able to pronounce tones accurately) or whether German learners rather have difficulties memorizing and retrieving tonal features.

5. Acknowledgements

This work is funded by the German Ministry of Education and Research grant 1746X08 and supported by DAAD-NSC (Germany/Taiwan) and DAAD-CSC (Germany/China) project related travel grants for 2009/2010.

6. References

[1] EURONOUNCE "Intelligent Language Tutoring System with multimodal feedback functions", Dresden University of Technology, Dresden, Saxonia, Germany. http://www.euronounce.net/.

[2] Mixdorff, H., Külls, D., Hussein, H., Gong, S., Hu, G. and Wei, S., "Towards a Computer-Aided Pronunciation Training System for German Learners of Mandarin", Proc. of SLaTE 2009, Wroxall Abbey Estate, Warwickshire, England, 2009.

[3] Hussein, H., Mixdorff, H., Do, H. S., Wei, S., Gong, S., Ding, H., Gao, Q. and Hu, G., "Towards a Computer-Aided Pronunciation Training System for German Learners of Mandarin - Prosodic Analysis", Proc. of Workshop on Second Language Studies, Tokyo, Japan, September 2010.

[4] Wang, W. S.-Y., "Phonological Features of Tone", International Journal of American Lingustics, Vol. 33, 2, pp. 93-105, 1967.

[5] Wang, R. H., Liu, Q. F. and Wei, S., "Putonghua Proficiency Test and Evaluation", Advances in Chinese Spoken Language Processing, Chapter 18, Springer press, pp. 407-430, 2006.

[6] Boersma, P. and Weenink, D., "*Praat* doing Phonetics by Computer", version 5.0.42, www.praat.org.

[7] Ding, H., Jokisch, O., Hoffmann, R., "Perception and Production of Mandarin Tones by German Speakers", Proc. of 5th Conference on Speech Prosody, Chicago, USA, May 2010.