Evaluation of Benefits from a Computer-Aided Pronunciation Training System for German Learners of Mandarin Chinese

Hue San Do[1], *Hussein Hussein*[1,2], *Hansjörg Mixdorff*[1], *Oliver Jokisch*[2], *Hongwei Ding*[3],*Qianyong Gao*[4], *Si Wei*[4] *and Guoping Hu*[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, China

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

{hsdo, hussein, mixdorff}@beuth-hochschule.de, oliver.jokisch@tu-dresden.de, hongwei.ding@tongji.edu.cn, {qygao, siwei, gphu}@iflytek.com

Abstract

The paper reports on the benefits of a computer-aided phonetic learning system for German learners of Mandarin. In the current study seven German first-year students of Mandarin Chinese participated in a test run of the phonetic training software. The students took four training units of 30 minute each within a week where they practiced their pronunciation and vocabulary with the software. A test was conducted before and after training in which the students read aloud both Mandarin Chinese disyllables of different tone combinations and words as well as sentences which they have encountered before when using the software. The corpus consisted of 25 tokens which were used in both tests, i.e. before and after the training with the software. The speech signals were recorded and then annotated by an expert regarding syllable components (initial, final and tone). The correctness of the syllable components and tone combinations, confusion partners and *F0* parameters of Mandarin tones were compared before and after the training. Ten native speakers of Mandarin rated the degree of foreign accent and intelligibility. The results based on the annotations of an expert, analysis of *F0* parameters of Mandarin tones and rating of accent and intelligibility show that the German learners yielded more accurate results of initial, final and tone after having practiced with the phonetic training tool.

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

1. Introduction

In a globalized world the growing demand for foreign language competence stimulates activities towards computer-aided language learning (CALL). CALL is a tool to facilitate individualized language learning and pronunciation training, for example [1]. Within this area, 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 the current 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 Ministry of Educations and Research. The results in this paper present the final stage of development of training system for German learners of Mandarin.

Modern Mandarin (*Putonghua*) differs from German significantly on the segmental as well as the supra-segmental levels and poses a number of problems to the German learner. Mandarin comprises a relatively small number of about 400 different syllables which are formed by combining 22 consonant initials (including glottal stop) and 38 mostly vocalic finals. Many of the phonemes building initials and finals have exact or close counterparts in the German language. Errors usually arise from phonemes of Mandarin without correspondences in German [3].

Mandarin is a tonal language. Tone is very important to distinguish Mandarin syllables, i.e. the tonal contour of a syllable changes its meaning. The tone distinction in Mandarin is the most complex problem for German learners. Mandarin has four syllabic tones and a neutral tone. Mandarin tone can be represented by prototypical *F0* contours [4] as shown in Figure 1 [5]. The acquisition of tonal patterns of poly-syllabic words is much more difficult than mono-syllabic words [2].

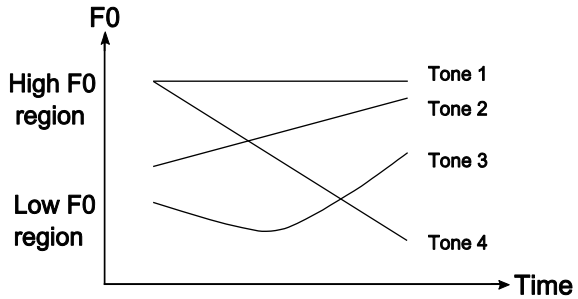

Figure 1: *Typical F0 patterns of four basic tones.*

In order to detect that using a computer-aided phonetic learning system for German learners of Mandarin (henceforth "*CALL-Mandarin system*") well be improve the phone and tone pronunciation, an analysis of syllable components before (henceforth "*Pretest*") and after (henceforth "*Posttest*") training with the software was implemented to calculate the correctness of syllable components and to detect confusion partners of tones. The intonational features of Mandarin tones

as well as the rating of accent and intelligibility for *Pretest* and *Posttest* were compared.

2. Framework of CALL-Mandarin System

The *CALL-Mandarin system* contains of phone and tone recognizers from the partner in the project (*iFlyTek* company, Hefei, China). Figure 2 shows the Graphical User Interface (GUI) of the computer-aided pronunciation training system for German learners of Mandarin. It provides the user with a list of 15 lessons with vocabulary, phrases and sentences as well as tone pair drills. The framework contains a set of reference which comprises utterances produced by native speakers of Mandarin. The software records the user's voice and shows the obtained data (*F0* contour and energy envelope) in real time as an audiovisual feedback. The position of the imitation signal is adjusted in relation to the start position of the reference signal in order to compare the reference and imitation signals. The user can repeat the imitation several times. The alignment of the imitation signal to the reference signal and the repetition process of imitation of the reference signal are intended to help the learner improve the quality of his pronunciation. Further information about components, application and functionality of the *CALL-Mandarin system* was described in [6].

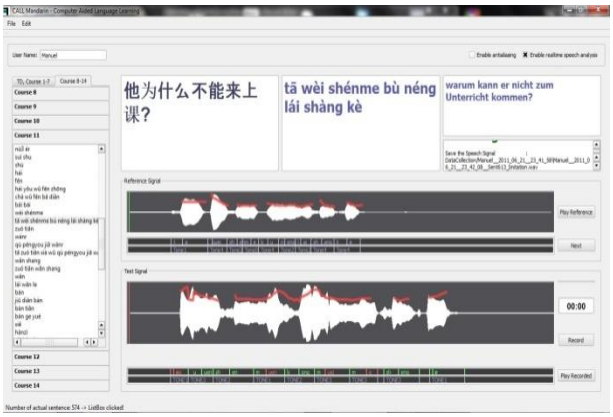

Figure 2: *Graphical User Interface of the CALL-Mandarin system.*

3. Experiment Method

3.1. Corpus Design and Data Collection

The data used in this experiment consists of recordings from seven first-year German students of Chinese Studies at the East Asian Seminar of the Free University Berlin. The German students were between 20 and 23 years of age and had completed two semesters (28 weeks) of Chinese language training at the time of the recording.

In this experiment seven testing subjects (TS), six females and one male, participated in a test run of the phonetic training software. The TS took four training units of 30 minute each on four days within a week where they practiced their pronunciation and vocabulary with the software. At the beginning of the training they were given an introduction how to use the software and advised to start with the tone pair drills and then continue with the chapters 1-14. During the training sessions an expert (German teacher of Chinese) was present to assist, answer questions and help with technical problems. Shortly before and after the training sessions a *Pretest* and *Posttest* were conducted. The data was recorded with sampling frequency of 16 kHz and a resolution of 16 bit. The corpus consists of 25 tokens, made up of two parts in the *Pretest* and *Posttest*, respectively:

Production of disyllables (henceforth "*Production 1*"): Ten Mandarin Chinese disyllables of different tone combinations were presented in Hanyu Pinyin transcription with tone markers to the testing subjects on a computer screen. The Chinese disyllables were shown in succession and read aloud by the TS without time constriction. Every tone combination occurred just once and was intentionally chosen from the vocabulary of existing Chinese words that were unknown to the testing subjects.

Production of disyllables, short phrases and sentences (henceforth "*Production 2*"): 15 words, phrases and sentences chosen from the vocabulary of the "New Practical Chinese Reader 1" were presented in Chinese characters only to the testing subjects on a computer screen. Again, the tokens were shown in succession and read aloud by the TS without time constriction. They consisted of four disyllable words, six phrases and five sentences made up of three, four to eight syllables, respectively, which the testing subjects had encountered during the training sessions. The reason for choosing items from the training material was to make sure that TS have encountered the characters before. Different from phonetic writing system and Pinyin transcription, you can not pronounce Chinese characters you have not seen before because the grapheme-phoneme correspondence is not as consistent and transparent as in alphabetic writing systems.

The tokens and settings used in the *Posttest* were the same as in the *Pretest*.

After the *Posttest* the testing subjects filled out a short questionnaire in which they evaluated the software regarding its functions: general user-friendliness, audiovisual feedback by showing the *F0* contour and energy curve in real-time, results of phone and tone recognition, recording and playback. In addition, they could assess their own progress and gave ideas for improvement.

3.2. Data Evaluation

The collected data was annotated and processed by:

1. Expert (German teacher of Mandarin) who listened to the data several times and annotated the syllables regarding initials, finals and the tones she perceived using Hanyu Pinyin transcription and the numbers 0-4 to mark the tones.

Based on the annotations, the correctness of the syllable components was compared between *Pretest* and *Posttest* for *Production 1* and *Production 2* separately. Similarly, the correctness of the tone combinations in *Production 1* was compared between *Pretest* and *Posttest*.

2. Ten native speakers of Mandarin (five from Tongji University, Shanghai, China and five from Dresden University of Technology, Dresden, Germany) listened to the German data and rated the degree of foreign accent and intelligibility on a scale from one to five, five being the best score, i.e. native-like competence.

3.3. Data Analysis

The forced-alignment on the syllable and phone-levels using the Automatic Speech Recognition (ASR) system was implemented. The used ASR system is part of an automated proficiency test of Mandarin [7]. The label files from the forced alignment were converted to the *Praat* TextGrid format [8] and combined in a single TextGrid file containing syllable and phone labels.

The *F0* contours were calculated using the *Praat* algorithm [8] 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* 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, the *F0* contours were normalized for each speaker. The normalized *F0* contour was calculated as in the following formula [9]:

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

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.

4. Results

The annotations made by the expert were compared to the original tokens which were used as reference to calculate the correctness of pronunciation.

4.1. Analysis of Syllable Components

Figure 3 shows the relatively high correctness for initials and finals in both *Pretest* and *Posttest* and only a slight difference between reading Hanyu Pinyin transcription or Chinese characters (correctness > 90%). Tone correctness, on the other hand, was lower in both *Production 1* and *Production 2* compared to segmental syllable features, and slightly lower when tone markers were not available as in *Production 2*. Yet the correctness increased for both modes more after the training, especially when reading Chinese characters.

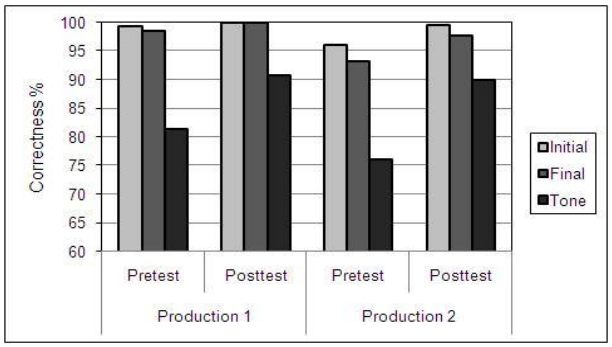

Figure 3: *Correctness of initial, final and tone for Pretest and Posttest in Production 1 and Production 2.*

4.2. Analysis of Mandarin Tone

We performed a more detailed tone analysis regarding tone confusion partners and tone combination.

4.2.1. Tone Confusion

Table 1 shows the correctness of single tones when Chinese disyllables are presented in Hanyu Pinyin transcription (*Production 1*). The correctness for all tones is relatively high already in the *Pretest* except for tone 3 which is mostly confused with tone 2 (51%) [10]. In the *Posttest*, the correctness for all tones increases, especially for tone 3 which is less pronounced as tone 2.

If tone markers are not available and only Chinese characters are presented (*Production 2*), tone correctness, in general, is lower in both *Pretest* and *Posttest* except for tone 3, as shown in table 2. Tone 3 is less often confused with tone 2 but tones 1, 2, 4 and neutral tone are more often pronounced incorrectly in the *Pretest*.

Table 1. *Correctness and confusion partners of tones for Pretest and Posttest in the Production 1 (in %).*

Tone	*Production 1*									
	Pretest					*Posttest*				
	T1	T2	T3	T4	T0	T1	T2	T3	T4	T0
T1	100	0	0	0	0	100	0	0	0	0
T2	0	93	7	0	0	0	95	2	0	2
T3	0	51	49	0	0	0	26	74	0	0
T4	0	6	6	86	3	6	0	0	94	0

Table 2. *Correctness and confusion partners of tones for Pretest and Posttest in the Production 2 (in %).*

Tone	*Production 2*									
	Pretest					*Posttest*				
	T1	T2	T3	T4	T0	T1	T2	T3	T4	T0
T1	78	3	3	3	12	87	1	2	5	4
T2	0	85	10	1	3	0	95	4	1	0
T3	3	20	71	2	5	0	17	83	0	0
T4	7	7	6	64	17	1	2	1	92	4
T0	11	4	0	2	83	4	0	0	0	96

4.2.2. Tone Combinations

The tone combinations in this section were considered correct when both tones were pronounced correctly. Tone combinations involving tone 3 either in front or end position, pose the biggest challenge on German learners. The combinations T2-T3 and T3-T2 were the most difficult to pronounce and were mastered better in the *Posttest*. However, the combinations T4-T2 and T4-T3 were less correct after the training as shown in table 3.

Table 3. *Correctness of tone combination for Pretest and Posttest in the Production 1 (in %).*

Tone Combination	*Production 1*	
	Pretest	*Posttest*
T4-T2	71.43	57.14
T2-T3	28.57	85.71
T1-T2	100.00	100.00
T2-T4	100.00	100.00
T3-T2	0.00	57.14
T3-T1	57.14	100.00
T4-T3	71.43	42.86
T3-T4	57.14	85.71
T2-T1	100.00	85.71
T1-T4	85.71	100.00

4.3. Comparison of F0 Contour of Mandarin Tone

Table4 shows the parameters of the normalized *F0* contours of syllables depending on the tones for *Pretest* and *Posttest* of *Production 1* and *Production 2*.

The mean value of *F0* contour of tone 1 by *Pretest* is greater than by *Posttest*, but the standard deviation (SD) and range of *F0* by *Posttest* is smaller. It indicates that the students after training with the software are able to keep the same level of *F0* contour for tone 1. The *F0* range of tones 2, 3 and 4 by *Posttest* is greater than by *Pretest*. This indicates that the German learners after training with the software are able to start with a low-level and to raise *F0* contour enough in tone 2. After training the students can decrease and raise the *F0* contour of the falling-rising tone (tone 3) more than before training. The German learners of Mandarin after training are

able to start with a high-level and decrease the *F0* contour enough in tone 4.

Table 4. *Mean, standard deviation and range of normalized F0 subcontour of syllables depending on the Mandarin tones for Pretest and Posttest of Production 1 and Production 2.*

Tone	*Production 1* & *Production 2*					
	Pretest			*Posttest*		
	F0 mean	*F0* SD	*F0* range	*F0* mean	*F0* SD	*F0* range
T1	3.03	0.32	1.24	2.94	0.29	1.13
T2	2.18	0.52	1.71	1.67	0.53	1.75
T3	1.98	0.50	1.67	1.57	0.54	1.81
T4	2.54	0.55	1.76	2.39	0.74	2.26

4.4. Comparison of Entire Utterance

The mean and standard deviation of accent and intelligibility ratings for data of *Production 1* & *Production 2* in *Pretest* and *Posttest* are presented in table 5. The accent and intelligibility after training are better than before training. The high scores of the utterance-wise judgments could be related to the higher tonal accuracy produced after training as described in the previous sections. The correlation results between accent and intelligibility for *Pretest* are .965 and .936 and for *Posttest* are .915 and .863 for *Production 1* and *Production 2*, respectively (Correlation is significant at the 0.01 level). Independent samples Mann-Whitney U-tests for *Pretest* and *Posttest* suggest that the improvement in performance is highly significant ($p < .002$ and $p < .027$ for accent and intelligibility, respectively). This suggests that the training yields improvements rather with regards to the strength of the perceived accent and less so for intelligibility.

Table 5. *Mean and standard deviation of accent and intelligibility for Pretest and Posttest in both Production 1 and Production 2.*

Scoring	*Production 1* & *Production 2*			
	Pretest		*Posttest*	
	mean	SD	mean	SD
Accent	3.32	0.66	3.56	0.53
Intelligibility	3.99	0.74	4.21	0.50

4.5. Evaluation of CALL-Mandarin System

After the *Posttest* the TS evaluated several functions of the training system on a scale from one to five, one being the best score, and made suggestions how to improve the software. The audiovisual feedback by showing the *F0* contour and energy curve in real-time, which is a distinct function of the *CALL-Mandarin System* in contrast to common CALL systems, was assessed to be beneficial to improve their pronunciation. The same applies to the recording and playback function. Furthermore, the TS suggested integrating more tonal exercises for "difficult" tones like T2 and T3; different modes to display character and Pinyin with/without tone markers; more "game-like" exercises to practice pronunciation. Most of the suggestions referred to exercise types. In general, the TS readily adapted themselves to use the *CALL-Mandarin System*, especially to receive visual feedback of *F0* contour and energy curve.

5. Conclusions

This paper reported on the benefits of a computer-aided phonetic learning system for German learners of Mandarin. The German learners of Mandarin pronounced segmental syllable components and tone more correctly after having practiced with the phonetic-training software but not tone combinations in general. Tone is pronounced less accurate when Pinyin is not available. Tone 3 is the most difficult tone to pronounce and mostly confused with tone 2 even if a tone marker is presented. Tone pair drills in the *CALL-Mandarin system* should focus on these combinations, preferably presented in Hanyu Pinyin transcription as well as existing words from the Chinese vocabulary. The TS assessed the *CALL-Mandarin System* which gives a real-time audiovisual feedback by displaying *F0* contour and energy curve as useful. More variations on exercise types to practice both pronunciation and vocabulary should be integrated. The analysis of *F0* parameters of Mandarin tones for *Posttest* shows that tone pronunciation after the training was better. The ratings of accent and intelligibility after training with the phonetic system were also better.

However, the authors of this paper are aware that the corpus used in these experiments was far too small to be significant. Additional, there should have been a control group to compare the results with in order to see whether students obtain better results when they train with the software. These limitations were due to difficulties finding testing subjects among the students at the East Asian Seminar who were willing to participate in an extensive experiment setting. Further test runs with the *CALL-Mandarin System* on a larger scale and with a control group are necessary to verify its benefits for Chinese learners.

6. Acknowledgements

This work is funded by 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.

7. References

[1] Seneff, S., "Interactive Computer Aids for Acquiring Proficiency in Mandarin", Proc. of ISCSLP, pp. 1-12, Singapore, 2006.

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

[3] Hunold, C., "Chinesische Phonetik. Konzepte, Analysen und Übungsvorschläge für den Unterricht Chinesisch als Fremdsprache", Sinica, Vol. 17, Bochum, 2005.

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

[5] Zhou, J.-L., Tian, Y., Shi, Y., Huang, C., Chang, E., "Tone Articulation Modeling for Mandarin Spontaneous Speech Recognition", Proc. of ICASSP, pp. 997-1000, 2004.

[6] Hussein, H., Mixdorff, H., Do, H. S., Mateljan, M., Gao, Q., Hu, G., Wei, S. and Chao, Z., "Comparison of Fujisaki-model parameters between German Learners and native speakers of Mandarin", Proc. of ESSV, pp. 146-153, Aachen, Germany, September 2011.

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

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

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

[10] Hussein, H., Do, H. S., Mixdorff, H., Ding, H., Gao, Q, Hu, G, Wei, S. and Chao, Z.: Mandarin Tone Perception and Production by German Learners. Proc. of SLaTE Workshop on Speech and Language Technology in Education, Venice, Italy, August 2011.