A Preliminary Investigation of German Rhythm by Chinese Learners

Hongwei Ding[1], *Rainer Jäckel*[2], *Rüdiger Hoffmann*[2]

[1]*School of Foreign Languages, Tongji University, China*
[2]*Laboratory of Acoustics and Speech Communication, TU Dresden, Germany*
hongwei.ding@tongji.edu.cn, {rainer.jaeckel;ruediger.hoffmann}@tu-dresden.de

Abstract: This study focuses on the temporal and metrical features of the German speech produced by Chinese speakers. It is well known that the values of proportion of vocalic intervals (*%V*) and the standard deviation of the consonant intervals (Δ*C*) within the sentence can classify the languages into stress-timed and syllable-timed categories. German is described to be a stress-timed language, while standard Chinese is regarded as a syllable-timed language. It has been suggested that the rhythm of the target language can be influenced by the learner's native language. In this study we conduct an investigation with 18 Chinese students of the same proficiency level in comparison with 6 native German speakers. Ten sentences from the whole recordings are selected for analysis, we come up with the following results: a) Chinese learners have higher values of *%V* than those of the German speakers; b) most Chinese learners have higher Δ*C* values than those of German speakers; c) Chinese learners speak much slower than German speakers. It is argued that these Chinese learners try to add additional vowels after syllable final consonants, they can hardly reduce vowels, but tend to delete certain consonants in consonant clusters due to the influence of their native CV structured syllable-stressed language.

1 Introduction

It is well known that Pike [9] and Abercrombie [1] proposed that the languages of the world can be classified into two types of rhythm patterns: a) stress-timed rhythm, and b) syllable-timed rhythm. Ramus [10] showed that stress-timed languages have a higher standard deviation of consonantal intervals Δ*C* and relatively lower proportion of the vocalic intervals *%V*; while syllable-timed languages have a lower Δ*C* and a higher *%V*. Grabe and Low [5] found that stress-timed languages have a higher variation in vowel durations, whereas syllable-timed languages show a lower variation in vowel length. Barry et al. [2] and Dellwo & Wagner [3] found that Δ*C* correlates negatively with speech rate in stress-timed languages.
German is usually described as a stress-timed language, one difference between prominent and non-prominent syllable is the duration [7]. One obvious example is that many word-final syllables in infinite verbs with *-en* are usually pronounced with syllabic consonants. For example, *laufen (to walk)* is usually pronounced as /laUf=n/ rather than /laUf@n/, in which the vowel /@/ is reduced and only the syllabic consonant /=n/ is pronounced. The syllable structure of German is also very complicated, which can be represented by (CC)V(CCCC) [7]. German can allow 2 consonants in the onset and up to 4 consonants in the coda in one syllable. Both the onset and coda consonants may be empty, forming a vowel-only syllable, or the nucleus can be occupied by a syllabic consonant as mentioned in the example /f=n/. While the syllable structure of standard Chinese is very simply, it is mainly formed by vowels with one onset consonant (C)V. Chinese does not allow consonant codas except for *-n* and *-ng*. The syllable onset can

be empty, but the vowel can hardly be reduced. The phonotactic differences between German and Chinese will affect the German phonology acquisition by Chinese learners, which may be reflected in the aspect of temporal organization of vocalic and consonantal intervals.
It has been suggested that the rhythm of the target language can be influenced by the learner's native language. Gut [6] described German L2 is influenced by L1 of Chinese, English, French, Italian and Romanian in terms of ΔC, $\%V$ etc. In her investigation, only four Chinese speakers were included, we are not sure whether they are representative. In our previous prosodic study [4], few sentences were employed, and no clear statistics was carried out in relation to rhythms. In this study we conduct an investigation with 18 Chinese students in comparison with 6 native German speakers.

2 Method

This study employed the same method which was described by Ramus [10] to investigate temporal and metrical features of 10 sentences read by 18 Chinese learners and 6 native German speakers. We managed to find some Chinese students of the same proficiency level, who arrived in Germany shortly before the recording. Perception of their German teacher confirmed that these students spoke with a typical syllable-timed melody, which is quite different from German stress-timed rhythm. In order to ensure comparability, the annotation technique used by Ramus [10] was adopted.
After the wave files had been automatically labelled with a German aligner developed at TU Dresden, annotation was conducted in two steps on Praat (http://www.fon.hum.uva.nl/praat/):

1. *phonetic segmentation* of the sentence into German phonemes; and
2. *classification* of separate phonemes into vowels and consonants.

In the first step, following the standard of phonetic criteria, the authors corrected automatic annotation manually as accurately as possible by referring to both visual and audio cues. The changes of spectrogram, waveform and formants (especially the first formant) served as the visual cues for setting the boundary of segmentation. Stops, affricates and nasals will be further separated into closure part (if exist) and burst part in the phoneme level. One example is shown in Figure 1 with the references of formants, in which the phoneme label is in the second tier.
In the second step, phonemes were then classified as vowels or consonants. In order to ensure comparability, the annotation technique of consonant and vowel intervals used by Ramus [10] has been adopted: pre- and inter-vocalic glides were treated as consonants, whereas post-vocalic glides were treated as vowels. Thus checked (free) vowels, free (long) vowels, unstressed schwa /@/, glottal stop /?/ before syllable initial vowels, and the vowel realisation of *r* /6/ were coded as V (vowel). Plosives, affricates, fricatives, sonorants (nasal and liquids) were coded as C (consonants). The classification can be observed in the first tier of annotation in Figure 1.
Therefore we measured the duration values of V and C, which refer to:

- *vocalic intervals*: the duration of sequences of consecutive vowels;
- *consonantal intervals*: the duration of sequences of consecutive consonants.

From the measurements we calculated two relevant variables of every sentence of each speaker:

- $\%V$: the proportion of vocalic intervals in the sentence; and
- ΔC: the standard deviation of consonantal intervals within the sentence.

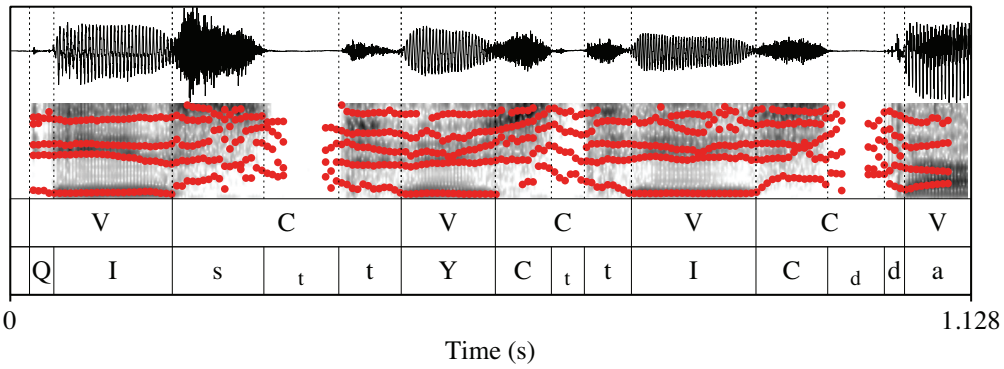

Figure 1 - Segmentation of consonant and vowel intervals

The phonetic segmentation was actually straightforward, especially of native speakers. The problem of labeling was the pause, especially of Chinese learners. Short pauses before the burst of stops and nasals were labeled as closure part of the corresponding phoneme. If there were some pauses and hesitations, which could not be identified as part of a sound, these breath parts were then marked as "_". Any two consonant intervals split by "_" (pauses or hesitations) were combined into the same consonantal interval in calculation by subtracting the duration of pause or hesitation. The same approach was used for vowel intervals as well.

2.1 Subjects

We recruited 18 native Chinese speakers, including 10 men and 8 women, who come from different parts of China, but all of them speak standard Chinese. At the time of the data collection, their lengths of residence were all 2 months, and they were having an intensive German course at Dresden University. Their ages ranged from 22 to 28. All of them had learned German for one to one and half year, and the length of formal German instructions had been around 1,200 hours. These Chinese participants formed a homogeneous group in terms of age, L1 background, motivation, proficiency of German language, length of residence in Germany, and so on. Among the 6 German native speakers, 1 was male and 5 were female speakers. They were middle-aged and represented the average German native speakers.

2.2 Data

The recording consisted two parts. First, every one was asked to read out 65 sentences, which are phonetic and phonological rich texts. Second, a short interview (approximately 5 minutes) was conducted, in which various questions about their language learning histories were asked. All recordings were carried out in the recording room at TU Dresden.
For the current study only ten read sentences were selected and analyzed. Because they include different sentence types, and the vowel and consonant percentage vary from sentence to sentence. It is better to concentrate on a small amount of data, because the accuracy of annotation is essential for the measurement, which requires much carefulness and patience. The interviews were only fragments of speech, which are too difficult for the statistical analysis of temporal metrics. Before the recording, the subjects were given as much time as they needed to read the text to become familiarized with it. Then each subject was individually recorded with 16 bit and 44.1k Hz by a German phonetics expert, who controlled the quality of their production and provided necessary assistance when they came across problems with the pronunciation.

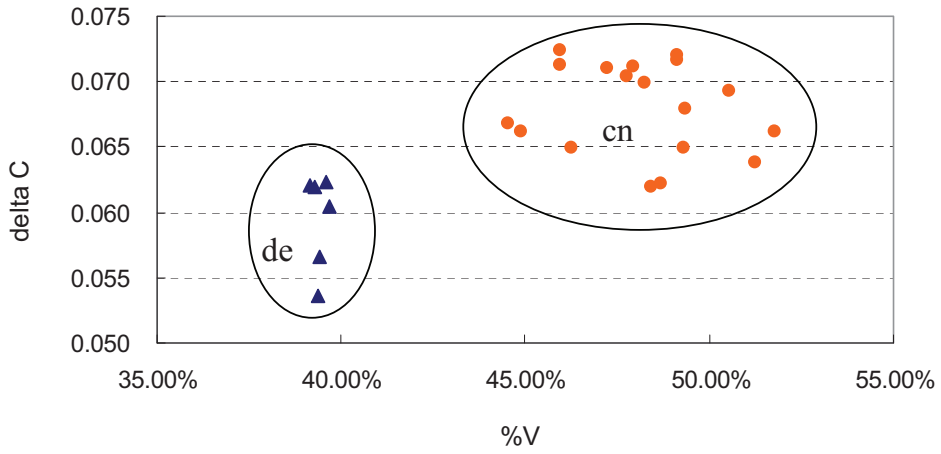

Figure 2 - Values of vocalic proportion (%V) and standard deviation of consonantal intervals (delta C) for all Chinese speakers (cn) and the German native speakers (de)

3 Results

3.1 %V and ΔC

The values of *%V* and Δ*C* are the averages of 10 sentences of each speaker, which is illustrated in Figure 2. Two results can be obviously derived from the figure:

- the values of *%V* of all the Chinese learners (ranging from 44.52% to 51.79%) are higher than those of the German native speakers (ranging from 39.14% to 39.67%).
- the values of Δ*C* of the Chinese learners (ranging from 0.062 to 0.072) are also slightly higher, but with a little overlap to those of the German native speakers (ranging from 0.054 to 0.062).

3.2 Duration and ΔC

It is also natural that these Chinese learners spoke much slower and made more pauses than the German native speakers. The average duration values of the 10 sentences for all the speakers are presented in Figure 3. The average duration is divided into two parts: (1) *durations without pauses* include vocalic and consonantal intervals, which were employed for the calculation of *%V* and Δ*C*; (2) *pauses* are the silent or breath periods which were included in the utterances of the sentence, but were excluded from the calculation. However these pauses can represent the rate of speech on one hand and also the fluency of the learners on the other hand. Three German speakers had no pauses at all in reading all the sentences, other 3 speakers made one or two pauses after the comma. All Chinese learners made more or less pauses, some at appropriate places between prosodic words, some at unappropriate places within the prosodic word.
We further calculated the correlation of the average duration with *%V* and Δ*C*. The speakers are

Table 1 - Correlation of duration with *%V* and Δ*C*

	German speakers		**Chinese learners**	
	%V	Δ*C*	%V	Δ*C*
duration without pauses	0.127	**0.931**	**0.508**	0.138
total duration	0.027	**0.949**	0.301	0.212

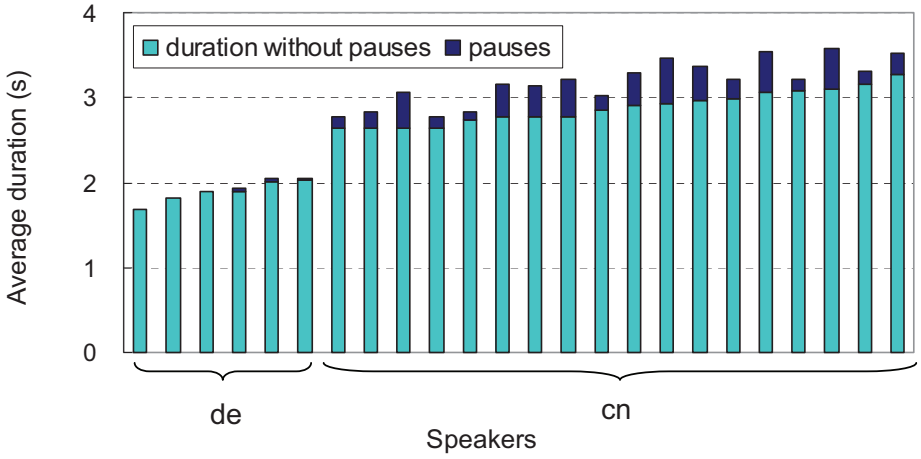

Figure 3 - Average duration value of all speakers

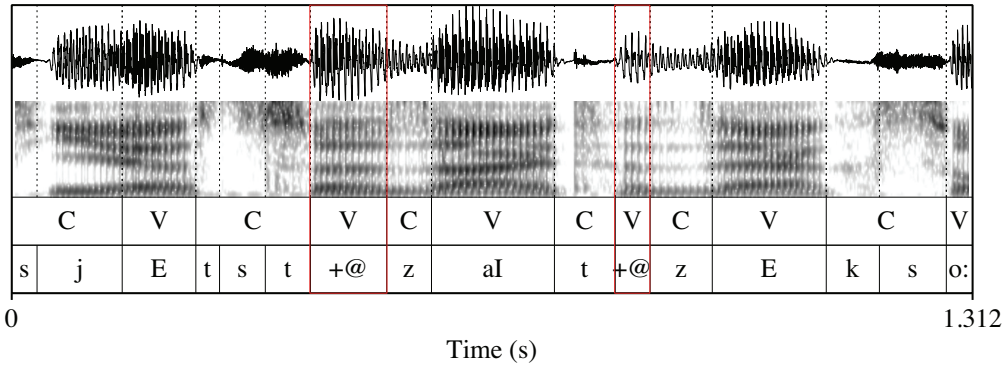

Figure 4 - A piece of speech with two additional schwas (marked as +@) by a Chinese learner

divided into two groups: German speakers and Chinese learners. And the duration values are also divided into two types: duration without pauses and total duration. For German speakers ΔC has a significant correlation with the total duration of utterances and with the duration (without pause) at the 0.01 level with r=0.949 and r=0.939, respectively. For Chinese speakers $\%V$ has a significant correlation with the duration (without pause) at the 0.05 level (r=0.508). No other correlations have been found between duration and $\%V$ for German speakers, or between duration and ΔC for Chinese speakers.

3.3 Epenthesis of Chinese learners

It is a common phenomenon that Chinese learners add vowels, especially schwa @ after consonant finals, which creates additional syllables for them to produce. One example is shown in Figure 4. This piece of speech should be *...jetzt seit sechs..*, most Chinese speakers added an @ after *jetzt* and *seit* as the speaker shown in the figure.

4 Discussion

Comparing the results presented by our investigations with those already published in the literature, we try to give some explanations:

- The vocalic proportion ($\%V$) of German native speakers is slightly lower than those reported in previous investigations in [10, 3, 6]. This can be resulted from the differences

of reading material and pronunciation characteristics. Most of our German speakers reduced vowel combination *-en*, *-el*, etc into syllabic consonants, in which no vowel part was annotated, and resulted in reduced $\%V$.

- $\%V$ values of our Chinese speakers are widely spread and higher than those reported in [6]. The reason for that might because that our participants had a shorter residence in Germany, and still put many additional schwas (@) after consonant finals just like in Chinese. In standard Chinese $\%V$ is reported to be over 56% [8]. The phenomenon of epenthesis is usually found at the beginning phase of Chinese learners of German [4], which will be improved with the advance of proficiency level.

- The standard deviation of consonantal intervals (ΔC) of our German speakers are between the values of the previously reported results [10, 3, 6]. The differences are quite acceptable and can be attributed to the differences of speech rate [3].

- ΔC values of our Chinese speakers are similar to some of the Chinese speakers reported in [6], they do not have very high values over 0.080. Perhaps they usually added a schwa between consonant clusters, or just deleted one or two consonants in the consonant codas. In this way, very long consonantal intervals were avoided, and ΔC would not be very high. In standard Chinese, ΔC is reported to be around 0.045 [8].

- The high correlation between ΔC and duration is found for German speakers, which indirectly supports the findings by Dellwo & Wagner [3]. Because all the speakers read the same syllables, the longer the duration, the slower the rate, which means that ΔC correlates negatively with the speech rate.

- The correlation between the duration and $\%V$ for the Chinese speakers means that some Chinese learners did not reduce the vowels when unstressed, or added schwas, which prolonged their duration.

- Though Chinese learners read the same sentences as the German speakers, because of epenthesis, they produced much more syllables than the German speakers. No correlation can be found between the duration and ΔC for Chinese speakers. Perhaps we have to calculate the number of syllables per second in reading, and correlate the rate of speech with ΔC, to see whether there is any correlation for Chinese speakers.

Because it was very difficult for the Chinese learners to read longer and complex German sentences, the sentences we used are shorter than those reported in previous investigations. As the Chinese learners make progress in their German proficiency, we will also try to use longer sentences, or sentences in conversation.
Since the quality of L2 phonology also depends on factors such as length and quality of exposure to the target language. Some studies show that negative transfer from L1 is more likely to be found in the beginning period. We will conduct a similar investigation after the Chinese subjects have finished their intensive German course and a series of systematic phonetic training three months later. In the future we will present some results in longitudinal investigations.

5 Conclusion

As reported in the previous literature [10, 3, 8] that stress-timed languages like German have lower $\%V$ and higher ΔC, while syllable-stressed languages like Chinese have higher $\%V$ and lower ΔC. When Chinese learners read German sentences, because of the organization of German text, they showed much lower $\%V$ than that when they read Chinese, but they still produced

higher percentage of the vocalic intervals. One reason is that they did not reduce vowels as the natives do, another reason is that they added an extra schwa to get a CV syllable structure out of German. And also because of the nature of German text, Chinese learners produced much higher ΔC than that in their Chinese speech, and even higher than those of German speakers. The explanation for that might be due to the slower rate of speech than that of the German natives.

6 Acknowledgements

The first author is sponsored by Shanghai Pujiang Program (11PJC099) and by the fund of the Tongji University for Promoting Academic Exchange (daz1201ky05) with Germany for this research work.

References

[1] ABERCROMBIE, D.: *Elements of general phonetics*. Aldine: Chicago, 1967.

[2] BARRY, W. J., B. ANDREEVA, M. RUSSO, S. DIMITROVA and T. KOSTADINOVA: *Do rhythm measures tell us anything about language type?*. In *Proceedings of the 15th international congress of phonetics sciences*, pp. 2693–2696, Barcelona, 2003.

[3] DELLWO, V. and P. WAGNER: *Relations between language rhythm and speech rate*. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 471–474, Barcelona, 2003.

[4] DING, H., O. JOKISCH and R. HOFFMANN: *A Phonetic Investigation of Intonational Foreign Accent in Mandarin Chinese Learners of German*. In MA, Q., H. DING and D. HIRST (eds.): *Proc. 6th International Conference on Speech Prosody*, vol. 1, pp. 374–377, Shanghai, China, 2012.

[5] GRABE, E. and E. L. LOW: *Durational variability in speech and the rhythm class hypothesis*. In GUSSENHOVEN, C. and N. WARNER (eds.): *Laboratory Phonology 7*, pp. 515–546. Berlin: Mouton, 2002.

[6] GUT, U.: *Non-native Speech: A Corpus-based Analysis of Phonological and Phonetic Properties*, vol. 9 of *English Corpus Linguistics*. Peter Lang GmbH, 2009.

[7] KOHLER, K.: *Einführung in die Phonetik des Deutschen*. Schmidt, Berlin, 1977.

[8] LIN, H. and Q. WANG: *Mandarin Rhythm: An Acoustic Study*. Journal of Chinese Language and Computing, 17(3):127–140, 2007.

[9] PIKE, K. L.: *The intonation of American English*. University Press: Michigan, 1945.

[10] RAMUS, F., M. NESPOR and J. MEHLER: *Correlates of linguistic rhythm in the speech signal*. Cognition, 72:1–28, 1999.