A Preliminary Study on the Inter-lexical Pauses in L2 English Read by Chinese Students

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

[1]*School of Foreign Languages, Shanghai Jiao Tong University, China*
[2]*Laboratory of Acoustics and Speech Communication, TU Dresden, Germany*
hwding@sjtu.edu.cn, ruediger.hoffmann@tu-dresden.de

Abstract: This study is concerned with the organization of inter-lexical pauses of L2 English speech produced by Chinese students. Foreign language learners (L2 learners) have often been observed to lack a natural "rhythm" when reading aloud. This investigation analyzed read speech of 5 English sentences at three different speech tempos by 18 Chinese native speakers. We compared the patterns of inter-lexical pauses of the Chinese subjects in the collected speech with those of the native English speakers reported in previous literature. It was found that the Chinese speakers could hardly read English at a very fast speech rate. Their patterns of inter-lexical pauses were more similar to those of the native speakers at a slow or normal speech rate than at a fast speech rate. However, most inter-lexical pause patterns of L2 English by the Chinese speakers deviate from those of the native speakers, which may result in foreign-accented speech with degraded intelligibility. We suggest that the importance of the organization of pauses between words should be taken into account in L2 English speech learning and teaching.

1   Introduction

In continuous speech words are usually grouped together into prosodic units by their rhythmic properties. On one hand listeners automatically process the prosodic information to understand the speech. On the other hand, speakers endeavor to employ adequate prosodic parameters to express the meaning of their utterances as accurately as possible. As it has been claimed by Frazier et al. that "the global pattern of prosodic phrasing is what matters in sentence comprehension" [4]. As one important prosodic parameter, pauses are usually inserted to signal boundaries of prosodic units. It has further been demonstrated that not only the position but also the duration of pauses are important to indicate the prosodic and syntactic structures of speech utterances [4, 12, 7, 6].

 Therefore, how to organize the pause patterns in oral speech plays a major role in the perception of naturalness of rhythm. However, different languages employ different ways to group prosodic units and to express prosodic prominence [8]. Because of the negative transfer of native language, second language learners may have some prosodic deviations in the production of oral speech, which may result in unintelligible foreign-accented speech of L2. Researches have shown that inter-lexical pauses can be predictable by algorithms [12, 7]. Pause patterns in spontaneous speech are complex due to the variation of speaking rate and normal dysfluencies. But pauses in read speech can be accurately predicted and the influencing factors can be controlled in some sense. The current investigation aims to study whether the Chinese speakers as English language learners pause between words differently than the native speakers, and to find out the correlations of pause duration patterns between the investigated Chinese speakers and the native English speakers reported by Grosjean et al. in [7].

2   Method

To investigate the pause patterns of L2 English, sentences which have rich hierarchical structure and not uniform pausing throughout are appropriate to examine the organization of pauses. Grosjean et al. provided not only well-structured sentences but also proportions of duration between the words [7]. Therefore we employed these sentences as the reading material for this investigation. Native Chinese students were then recruited as subjects to read these sentences. The recordings were annotated and duration of inter-lexical pauses were correlated with the duration values provided by Grosjean et al. [7] to show whether the Chinese speakers demonstrated similar patterns of inter-lexical pauses in reading these English sentences.

2.1   Performance Structures

Many researches have shown that the occurrences and lengths of pauses are correlated with the degree of inter-lexical cohesion [4, 7]. A performance structure is a psychological structure that can capture various degrees of cohesion between the words in a sentence and is believed to be able to predict the location and length of pauses [5]. Grosjean et al. asked subjects to read 14 sentences at five different rates. The pausing values they obtained were averaged and these mean duration values expressed as a percent of the total pause duration in each sentence can reflect the hierarchical organizations of the sentences. Following are the first 5 of the 14 sentences used by Grosjean et al. [7].

1. When 5 the 0 new 2 lawyer 20 called 3 up 5 Reynolds 30 the 3 plan 10 was 1 discussed 21 thoroughly.

2. In 2 addition 10 to 3 his 3 files 33 the 1 lawyer 8 brought 17 the 0 offices 10 best 13 adding-machine.

3. By 0 making 11 his 2 plan 5 known 38 he 0 brought 4 out 19 the 2 objections 16 of 3 everyone.

4. That 5 a 5 solution 15 couldn't 7 be 3 found 30 seemed 9 quite 6 clear 17 to 3 them.

5. Not 11 quite 10 all 17 of 4 the 3 recent 9 files 23 were 7 examined 10 that 6 day.

The values at each word boundary indicate the percentage pause duration obtained at that boundary. In the current investigation the subjects were asked to read the same sentences, and the duration values listed above were also regarded as the standard reference for detection of deviation of the L2 Chinese speakers.

2.2   Data Collection

We recruited 18 native Chinese speakers, including 9 males and 9 females, with the age range between 20-25. All of them were born and had lived in China before they came to Germany. All the subjects spoke Mandarin Chinese as their native language, they had learned English as their first foreign language for about 10 years, and they were learning German in Germany as their second foreign language. The read speech was collected one month after they had arrived in Dresden.
 The recordings were conducted in a very quiet room at TU Dresden. The subjects were asked to get familiarized with the sentences and we ensured that they rightly understood the meanings of the sentences before their readings were recorded. Following the procedure of Grosjean et al. [7], we also required our subjects to read these 14 sentences at 5 rates with normal rate for two

times: (1) normal, (2) slow, (3) very slow, (4) normal, (5) very fast. There were many hesitations in the speech from the first normal rate, but great improvements had been made at the second normal rate after two repetitions at slow and very slow rate. Since no obvious differences could be found between speeches at slow and very slow rate, recordings from steps (3), (4) and (5) were analyzed to represent speeches at slow, normal, and fast tempos.

Though 14 sentences were recorded for each speaker, the 5 sentences listed above were analyzed in this study as a preliminary investigation. The data studied consist of 270 sentences, which were resulted from 3 repetitions x 5 sentences x 18 subjects.

2.3   Data Analysis

The 270 sentences were first automatically annotated by an automatic aligner SPPAS [1]. Then the annotations were corrected manually with Praat program [2] by one phonetics student and checked by the first author. The duration values were obtained by a Praat script and analyzed using the R software [10].

Great attention was paid to the annotation of word boundaries. According to Zellner [12] pauses can be classified by two criteria: (1) physical and linguistic classification (2) psychological and psycholinguistic classification. The physical and linguistic classification differentiates between intra-segmental pauses and inter-lexical pauses. Intra-segmental pauses occur inside words and are due to the occlusion of the vocal tract in normal speech production, so the closure phase of a plosive consonant is one example of the intra-segmental pauses. While inter-lexical pauses appear between words, which are likely to facilitate the perceptual interpretation of the speech utterance. The psychological and psycholinguistic classification distinguishes silent pauses from filled pauses. Silent pauses correspond to section where no speech signal appears. While filled pauses are related to the voiced sections in the speech material such as drawls, repetitions of utterances or false starts. The current study is focused on the inter-lexical pauses. Since repetitions normally appear in spontaneous speech, they were excluded in our calculation of pause in the read speech.

3   Results

Because the duration values of pause are closely relatively to speech rate [9], in the results we present firstly the speech rates and then the correlations of the inter-lexical pause duration values between the Chinese speakers and the native speakers reported by Grosjean et al. [7].

3.1   Speech rate

The speech rates were manifested by speaking rate with pauses between words and by articulation rate without pauses between words. The speaking and articulation rates of each speaker at slow, normal and fast tempos are displayed in Figure 1, Figure 2, and Figure 3, respectively. The value of every speaker was the average across the 5 sentences.

The means, standard deviations (sd) and ranges of speech rate (the lowest rate to the highest rate of the speakers) at three different tempos can be observed in Table 1.

Referring to the inform in the table and figures, it is clear that the speech rate varies from speaker to speaker. There are some overlaps of the speech tempos. For instance, a slow tempo for one speaker could be a normal tempo for another. However, the tempo differences for any individual speaker could still be identified.
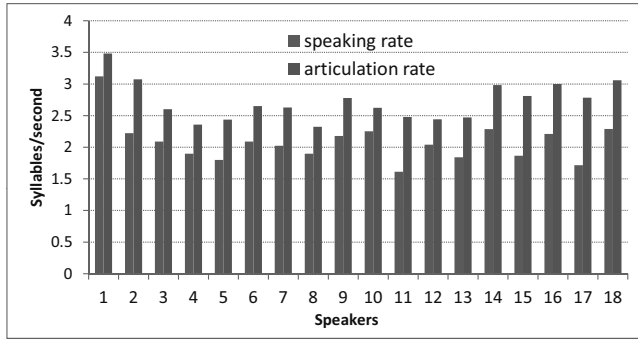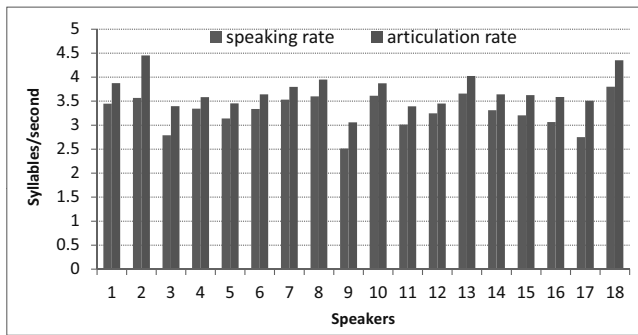
Figure 1 - Speech rates at slow tempo.

Figure 2 - Speech rates at normal tempo.
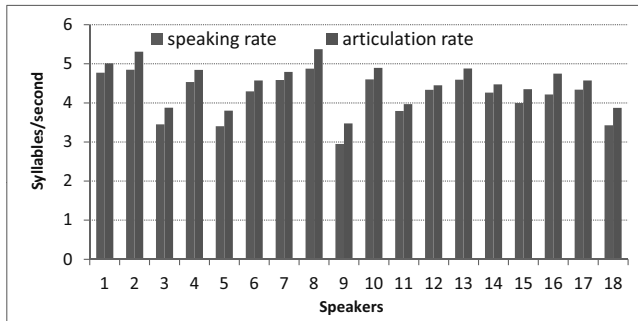
Figure 3 - Speech rates at fast tempo.

Table 1 - Overview of the mean, sd and range of speech rate at different tempos

Tempo	Slow		Normal		Fast	
Rate	speaking	articulation	speaking	articulation	speaking	articulation
Means (syl/s)	2.08	2.72	3.27	3.70	4.18	4.52
SD(syl/s)	0.33	0.31	0.35	0.35	0.56	0.54
Range(syl/s)	1.61-3.12	2.32-3.48	2.51-3.80	3.06-4.45	2.95-4.87	3.48-5.28

3.2 Correlation

It is generally accepted that a pause with a duration around 200 - 250 ms is more easily perceived [12]. According to the criterion, most inter-lexical pauses could be perceived at a slow speech rate, some at a normal rate, and few at a fast rate, which meets the requirements of different speech tempos. However, whether the patterns of the pauses are similar to those of the performance structures remains to be examined.

We correlated the duration values of the Chinese speakers with those predicted by Grosjean et al. [7]. If the correlation was significant, that means the speaker demonstrated a similar inter-lexical duration proportion to that of the native speakers. If the correlation was not significant or negative, that implies the pause pattern deviated from that of the native speakers.

The correlation coefficients of the speakers for Sentence 1-5 are displayed in Table 2-6 respectively. The values printed in bold indicate the correlations are significant at the 0.05 level.

Table 2 - Correlation coefficients of Sentence 1

speakers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
slow	0.52	**0.73**	-0.06	**0.73**	0.52	0.58	0.17	0.26	**0.73**	0.05	0.24	0.55	0.59	0.48	0.51	**0.75**	**0.64**	0.48
normal	0.10	0.36	**0.78**	**0.67**	-0.07	0.48	0.57	0.21	-0.01	**0.74**	0.11	-0.40	-0.12	0.02	**0.69**	0.06	0.01	-0.21
fast	0.48	-0.15	0.22	-0.17	-0.21	0.45	-0.02	0.02	0.00	0.02	-0.25	0.11	-0.05	0.08	-0.08	0.23	-0.11	0.24

Table 3 - Correlation coefficients of Sentence 2

speakers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
slow	0.54	**0.74**	0.23	**0.65**	**0.74**	0.57	**0.70**	**0.86**	0.77	**0.83**	**0.83**	0.30	**0.88**	**0.73**	**0.66**	**0.63**	**0.81**	0.54
normal	0.56	0.50	**0.66**	0.21	-0.10	**0.70**	**0.85**	**0.94**	0.36	**0.92**	**0.85**	-0.13	0.49	0.22	**0.92**	0.38	**0.58**	**0.64**
fast	0.55	0.51	0.26	-0.11	-0.14	0.48	0.04	0.57	-0.04	0.41	0.14	0.04	0.27	0.18	0.07	0.17	0.04	0.23

Table 4 - Correlation coefficients of Sentence 3

speakers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
slow	0.12	-0.05	-0.27	0.23	**0.78**	**0.72**	-0.41	0.18	-0.27	**0.77**	-0.19	0.44	0.07	-0.14	**0.78**	-0.22	**0.62**	-0.06
normal	-0.27	-0.07	0.21	**0.63**	0.19	0.10	-0.39	-0.11	-0.27	**0.66**	-0.10	-0.41	-0.29	-0.01	**0.75**	-0.51	**0.94**	-0.17
fast	-0.34	0.42	-0.28	-0.42	-0.53	-0.24	-0.16	-0.39	-0.49	-0.33	-0.18	-0.34	-0.31	-0.18	-0.43	-0.10	-0.31	0.57

Table 5 - Correlation coefficients of Sentence 4

speakers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
slow	**0.75**	0.34	0.11	0.53	**0.88**	**0.80**	0.53	0.32	0.27	**0.63**	0.37	0.60	0.51	**0.72**	0.38	-0.09	**0.76**	**0.74**
normal	0.20	**0.81**	-0.02	0.04	-0.03	-0.14	-0.02	0.06	0.33	0.44	**0.93**	0.24	0.08	0.17	0.15	**0.71**	**0.75**	-0.23
fast	0.21	-0.04	-0.24	-0.13	-0.27	-0.06	-0.04	-0.44	-0.14	-0.14	0.02	-0.17	-0.02	0.05	-0.18	-0.20	-0.08	0.62

Table 6 - Correlation coefficients of Sentence 5

speakers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
slow	0.40	0.11	0.13	0.51	0.61	0.42	0.15	0.60	0.09	-0.03	0.10	0.24	0.50	-0.27	-0.27	**0.79**	**0.68**	0.49
normal	-0.06	-0.47	-0.15	0.40	-0.04	0.27	**0.75**	-0.26	-0.03	0.05	0.59	0.05	0.03	0.01	**0.72**	0.50	**0.72**	-0.24
fast	-0.39	**0.75**	-0.13	-0.04	-0.20	0.04	-0.16	0.13	-0.31	0.21	-0.08	-0.15	-0.12	-0.23	**0.74**	0.22	-0.13	0.30

It can be observed that the correlation coefficients vary from sentence to sentence, but most of them are not significant. It is clear that duration values in L2 speech at slow and normal tempos are more correlated than that at fast tempo.

4 Discussion

Based on the results, several conclusions can be drawn:

- Speech rates in L2 English of the Chinese speakers are lower compared with those reported in the literature [11]. The Chinese learners can hardly read English at a very fast rate due to lack of coarticulation and non-reduction of syllables [3].

- Few Chinese students can demonstrate similar pause patterns in L2 English to those of English native speakers, which is due to the different organizations of inter-lexical pause patterns from the native speakers.

- In slow and normal L2 English speech, pause patterns display better correlation than those in fast speech. One reason is that in fast speech pauses occur only at the boundaries of prosodic phrases; therefore, a wrong location of pause can produce a greater deviation.

Some improvements can still be made in this investigation:

- Results were only compared with the studies reported in previous literature. The results would have been more reliable if native subjects had also been recruited as references.

- If all 14 sentences had been analyzed, more comprehensive results would have been obtained.

However, in the investigation of inter-lexical pauses several problems remain to be solved:

- Silent or filled pauses between words are not the unique criterion for grouping words. It was observed in this investigation that many Chinese students inserted pauses within the word, which disturbed the rhythms seriously.

- Acoustic pauses and pauses in perception should both be investigated. Pauses perceived sometimes do not correspond to the silent or filled pauses. For example, pitch and intensity variations can also indicate the syntactic or prosodic boundaries.

These affecting factors work together with the acoustic pause duration to express the prosodic information of oral speech, which should be taken into consideration in the inter-lexical pause investigations in the future.

5 Conclusion

This study has displayed that the inter-lexical duration patterns in L2 English read by Chinese students deviate from those of the native speakers, which may lead to less intelligible L2 English speech with Chinese accent.

6 Acknowledgements

References

[1] BIGI, B. and D. HIRST: *SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody*. In *Proceedings of Speech Prosody 2012*, 2012.

[2] BOERSMA, P. and D. WEENINK: *Praat: doing phonetics by computer [Computer program]*. http://www.praat.org, 2015.

[3] DING, H., R. JÄCKEL and R. HOFFMANN: *A Preliminary Investigation of German Rhythm by Chinese Learners*. In WAGNER, P. (ed.): *Tagungsband der 24. Konferenz Elektronische Sprachsignalverarbeitung*, vol. 65, pp. 79–85. TUDpress, 2013.

[4] FRAZIER, L., K. CARLSON and C. CLIFTON: *Prosodic Phrasing is Central to Language Comprehension*. TRENDS in Cognitive Sciences, 10(6):244–249, 2006.

[5] GEE, J. P. and F. GROSJEAN: *Performance Structures: A Psycholinguistic and Linguistic Appraisal*. Cognitive Psychology, 15:411–458, 1983.

[6] GOLDMAN-EISLER, F.: *The Distribution of Pause Durations in Speech*. Language and Speech, 4(4):232–237, 1961.

[7] GROSJEAN, F., L. GROSJEAN and H. LANE: *The patterns of silence: Performance structures in sentence production*. Cognitive Psychology, 11:58–81, 1979.

[8] JUN, S.-A.: *Prosodic Typology*. In JUN, S.-A. (ed.): *Prosodic Typology: The Phonology of Intonation and Phrasing*, pp. 430–458. Oxford University Press, 2005.

[9] KENDALL, T.: *Speech Rate, Pause and Sociolinguistic Variation*. New York : Palgrave Macmillan, 2013.

[10] TEAM, R. D. C.: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

[11] TROUVAIN, J.: *Tempo variation in speech production: Implications for Speech Synthesis*. PhD thesis, Saarland University, 2003.

[12] ZELLNER, B.: *Pauses and the Temporal Structure of Speech*. In KELLER, E. (ed.): *Fundamentals of speech synthesis and speech recognition*, pp. 41–62. Chichester: John Wiley, 1994.