The Influence of Comma- and Period-pause Duration on the Listener's Impression of Speeches Made in Mandarin Chinese

Mingji Lin, Kyushu University, Japan
Yoshitaka Nakajima, Sound Corporation, Japan
Shimeng Liu, Kyushu University, Japan
Kazuo Ueda, Kyushu University, Japan
Gerard B. Remijn, Kyushu University, Japan

The Asian Conference on Language 2021
Official Conference Proceedings

Abstract
The influence of pause duration at commas and periods on listeners' impressions of speeches made in Mandarin Chinese was investigated. Spoken excerpts of speeches from textbooks were presented to native Chinese listeners (n=20). In the first experiment, the pause durations of both commas and periods in the speeches were manipulated together, in 8 steps from 0 - 4.8 s. The listeners were asked to rate the speeches on 23 categories on a rating scale, including categories regarding the tempo, quality, and continuity of the speeches. Factor analysis (based on principal component analysis) over the rating data showed that out of four extracted factors, two factors prominently appeared. These two factors were interpreted as reflecting *speech naturalness* and *speech rate*. The speech rate impressions increased as the comma- and period-pause durations decreased. The speech naturalness was the highest when the pause duration was 0.6 s. In a following experiment, comma- and period-pause durations were manipulated separately, varying from 0.15 - 2.4 s. Original speech and speech without pauses were included as control conditions. Factor analysis over the rating data (n=20) again showed *speech naturalness* and *speech rate* as the two main factors. Taken together, both experiments convincingly indicated that speeches with a comma-pause duration of 0.6 s, along with a period-pause duration of 0.6 s or 1.2 s, are heard as having the highest *speech naturalness*, i.e., close to that of the original speech. These findings may be incorporated in artificial speech, and may be useful when practicing the delivery of speeches.

Keywords: Speech Perception, Pause Duration, Mandarin Chinese

1. Introduction

There are three types of temporal gaps in speech. These are voice onset time (VOT) within syllables, pauses or energy dips in the speech signal for word segmentation between words, and pauses for punctuation between clauses. Figure 1 shows each type of temporal gap. According to previous research, the three types of temporal gaps affect the perception of speech in different aspects. For example, VOT influences the perception of consonants in syllables, as evidenced from various studies on different languages and speakers (Lisker & Abramson, 1964; Li, 2013; Kang, 2014). Word segmentation research commonly pertains to the perception of stress patterns in speech and how this contributes to sentence parsing (Curtin, Mintz, & Christiansen, 2005). The most obvious temporal gaps in speech are pauses for punctuation.

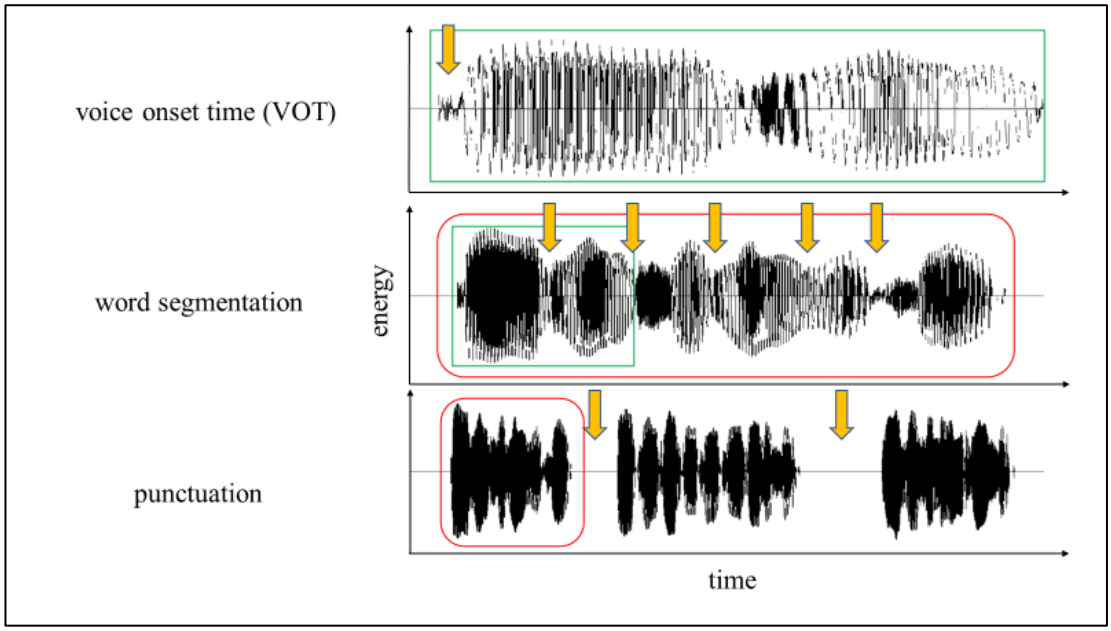

Figure 1. Examples of three types of temporal gaps in speech. Frames with the same color refer to the same part of the sentence with different resolutions, and the arrows indicate the exact places of each type of temporal gap.

Research about pause duration of punctuation marks in English speech, such as commas and periods, has been done for many years. The word *pausology* was firstly mentioned in 1965, referring to the use of pauses in speech and music (Tosi, 1965). Efforts have been made to identify hesitations and pause duration and their frequency of occurrence in speech in an automatic way (Horii, 1983). With regard to speech perception, previous research on pause duration has shown that different lengths of physical pause durations of American English and German can lead to an overestimating or underestimating of perceived pause durations (Stuckenberg & O'Connell, 1988). As for punctuations, though, little is known about which pause durations are most natural. A recent study used listeners' judgements and principal component analysis to shed light on the matter. Liu and colleagues (2019) performed a series of listening experiments in which the length of comma- and period-pause duration was varied in short English sentences. Listeners were asked to judge the perception of the sentences on 23 items or 12 items, respectively, in two experiments by means of a rating scale (Liu, Nakajima, & Elliott, 2018; Liu et al., 2019). Following Principal Component Analysis, two main factors were extracted from the rating judgements. These were "speech naturalness" and "speech rate", which we will refer to as the "Speech Rate factor" and the "Naturalness factor"

from here on. As for the Speech Rate factor, the obvious result was that if pause duration was longer, speech rate was perceived as slower. More importantly, as for the Naturalness factor, the most natural pause duration for English was the condition in which commas and periods were 0.6 s (600 milliseconds). That is, when both commas and periods were fixed at 0.6 s, speech naturalness was not significantly different from that of the original speech.

In the present study, we used the same research methods as in Liu et al. (2018; 2019) to investigate the naturalness of punctuation pause durations of commas and periods in Mandarin Chinese speech. The reason why we choose Chinese as target language was that there is no pausology research for languages that have a short history of using punctuation marks, like Mandarin Chinese. Punctuation marks in Chinese have been used for about only 100 years. Other than English, which has 13 kinds of punctuation marks, Chinese has 17 kinds of punctuation marks (Straus, Kaufman, & Stern, 2014; General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China & Standardization Administration [GAQSIQ & SA], 2011). Furthermore, Chinese is a tonal language, and English is a stress language. It is therefore likely that research findings on punctuation pauses and speech naturalness in English may not directly pertain to Chinese.

2. Methods

Two experiments were performed in order to identify the most natural comma- and period-pause duration for Mandarin Chinese. The results were compared with the previous research for English speech (Liu, Nakajima, & Elliott, 2018; Liu et al., 2019)

2.1 Participants

For each experiment there were 20 native-Mandarin-Chinese participants. They were 6 males and 14 females in Experiment 1, and 9 males and 11 females in Experiment 2. All participants were university students and had normal hearing. Before each experiment, all participants provided written, informed consent as to their participation. The experimental procedures were pre-approved by the Ethical Committee of Kyushu University, Fukuoka, Japan.

2.2 Apparatus

Experiments were completed in a soundproof booth with a background noise of 25.1 dB L_{AF} and 25.4 dB L_{AF} in Experiment 1 and Experiment 2, respectively. Speech stimuli were binaurally played to participants through a PC (Microsoft Surface 3 64GB, OS Windows 8.1), an amplifier (AT-HA40USB) and headphones (Roland RH-300).

2.3 Stimuli

The comma- and period-pause durations of Mandarin Chinese from HSK (Hanyu Shuiping Kaoshi) level 5 textbooks were varied together or separately in the two experiments. Speech samples were selected from 2 male speakers and 2 female speakers in Experiment 1, and from 1 male and 1 female speaker in Experiment 2. Comma- and period-pause duration in the sentences were presented in the following steps, as in Liu and colleagues (2018, 2019). In Experiment 1, the comma and period durations were: original speech, 0 s, 0.075 s, 0.15 s, 0.3 s, 0.6 s, 1.2 s, 2.4 s and 4.8 s. In Experiment 2, we used original speech, 0 s, 0.15 s, 0.3 s, 0.6 s, 1.2 s and 2.4 s. Since comma- and period-pause durations in Experiment 2 were varied

individually, we used fewer steps in Experiment 2, in order not to make the experiment too long.

2.4 Procedures

Participants controlled the stimulus presentation themselves through buttons on the screen. After they clicked a "Play" button, the speech stimuli were played automatically after 0.5 s. Each stimulus was presented only once, and the time for participants to rate was not limited. Participants were asked to rate the speech stimuli on 23 or 12 evaluation items, respectively, in Experiment 1 and 2. Ratings were made on a semantic differential scale from 1 to 10, with "1" meaning "not appropriate at all" and "10" meaning "very appropriate". Evaluation items used in Experiment 1 were "rushed", "natural", "rough-timbred", "skillful", "speedy", "at a suitable tempo", "well-practiced", "fast", "with appropriate pause duration", "friendly", "high-pitched", "with appropriate rhythm", "smooth", "nervous", "experienced", "shrill", "fluent", "easy to understand", "elegant", "intelligible", "polite", "dynamic", and "clear-cut". Evaluation items used in Experiment 2 were "rushed", "natural", "rough-timbred", "skillful", "speedy", "at a suitable tempo", "well-practiced", "fast", "with appropriate pause duration", "friendly", "high-pitched", and "with appropriate rhythm". All the research protocols and experimental designs were the same as used in the previous research on pause durations in English speech (Liu, Nakajima, & Elliott, 2018; Liu et al., 2019).

3. Results

The results were analyzed in the following steps. Firstly, we calculated the means of the rating scale judgments in order to see which of the evaluation items were influenced by the changes in the pause duration. Then we performed Principal Component Analysis to extract the main factors. Principal Component Analysis can "simplify" the results of the ratings by summarizing them into "factors". Four factors were extracted in Experiment 1 (Table 1). These were "speech naturalness", "speech rate", "speech friendliness" and "tone height". In Experiment 2 we obtained 3 factors. These were "speech naturalness", "speech rate" and "else". We found that two factors that were extracted in both experiments, were the same as for the previous research with English pause durations (Liu, Nakajima, & Elliott, 2018; Liu et al., 2019). These factors were the Speech Rate factor and the Naturalness factor.

Figure 2 shows the relationship between pause duration and the Naturalness factor and the Speech Rate factor obtained from the results of Experiment 1. As for the Speech Rate factor, indicated in red, when the pause duration got longer, the speed of the sentences was perceived to be globally slower, even though we just adjusted the pause duration and not the speech itself. The perceived speech rate thus can be influenced by pause duration itself. As for the Naturalness factor, indicated in green, the factor score of the original speech sentences (the green filled circle) was significantly higher ($p<0.05$) than the factor scores of all the speech sentences in which the pause duration was varied. However, the factor scores of speech sentences with a fixed comma- and

	factors			
	1	2	3	4
at a suitable tempo	**0.899**	-0.003	0.132	0.026
with appropriate pause duration	**0.897**	-0.030	0.111	0.041
with appropriate rhythm	**0.883**	0.013	0.161	0.000
natural	**0.805**	0.127	0.343	0.000
well-practiced	**0.747**	0.413	0.288	0.076
experienced	**0.737**	0.345	0.352	0.115
intelligible	**0.728**	0.025	0.132	-0.089
easy to understand	**0.720**	0.064	0.187	-0.123
skillful	**0.687**	0.409	0.372	0.088
fluent	**0.618**	0.604	0.237	-0.032
smooth	**0.615**	0.578	0.280	-0.049
fast	-0.067	**0.852**	-0.125	0.205
rushed	-0.189	**0.849**	-0.180	0.140
speedy	0.281	**0.821**	0.052	-0.062
dynamic	0.378	**0.609**	0.192	-0.086
elegant	0.483	-0.056	**0.682**	0.075
clear-cut	0.222	0.139	**0.669**	0.119
rough-timbred	0.063	-0.106	**-0.657**	0.362
polite	0.469	-0.097	**0.646**	-0.152
friendly	0.483	-0.168	**0.592**	-0.181
nervous	-0.365	0.386	**-0.438**	-0.165
high-pitched	0.037	0.044	0.026	**0.896**
shrill	-0.075	0.090	-0.124	**0.875**

Table 1. The result of the Principal Component Analysis performed over the 23 evaluation items used in Experiment 1, after varimax rotation. Factors 1-4 represent "speech naturalness", "speech rate", "speech friendliness" and "tone height", respectively, from top to bottom delimited by thick black frames. The Speech Rate factor and the Naturalness factor were also among the 3 factors extracted from the results of Experiment 2 (not shown here).

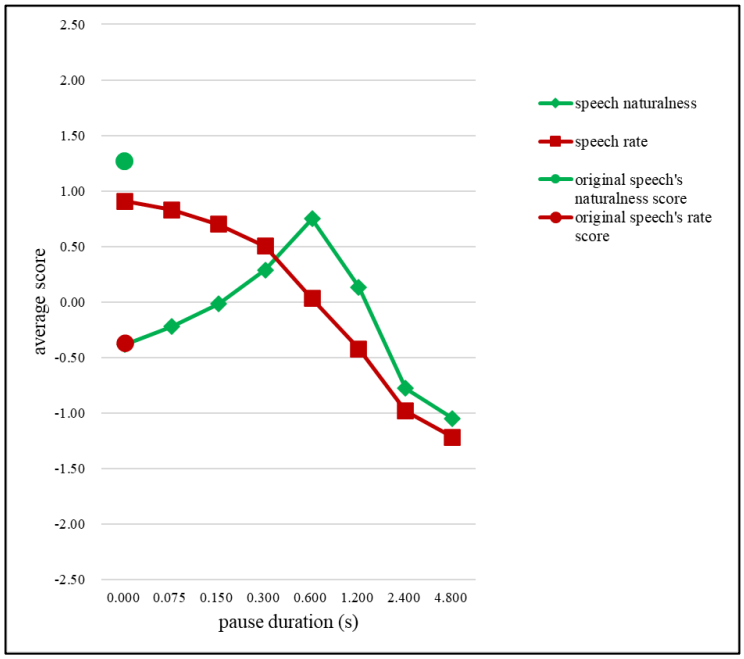

Figure 2. The relationship between pause duration and the Naturalness factor and the Speech rate factor in Experiment 1

Period-pause duration of 0.6 s were statistically significantly higher than that of any of the other manipulated speech sentences. Although significantly less natural than the original speech, Mandarin Chinese speech with a comma- and period-pause duration of 0.6 s was the most natural.

For Experiment 2, the same analysis method was used as in Experiment 1. As for the Speech Rate factor, similar to Experiment 1, when the pause duration got longer, the perceived speed of the sentences was perceived to be globally slower. Figure 3 shows the factor scores for the Naturalness factor obtained from the rating results of Experiment 2. The statistical analyses showed that the factor scores of speech sentences with a comma-pause duration of 0.6 s, and those with a period-pause duration of 0.6 s and a period-pause duration of 1.2 s were not significantly different from the original speech sentences. For the results of Experiment 2, we thus can conclude that the sentences with a comma duration of 0.6 s, and the period durations of 0.6 s and 1.2 s were perceived as having the same naturalness as the original speech, while the most natural speech was for a comma duration of 0.6 s, and a period duration of 0.6 s.

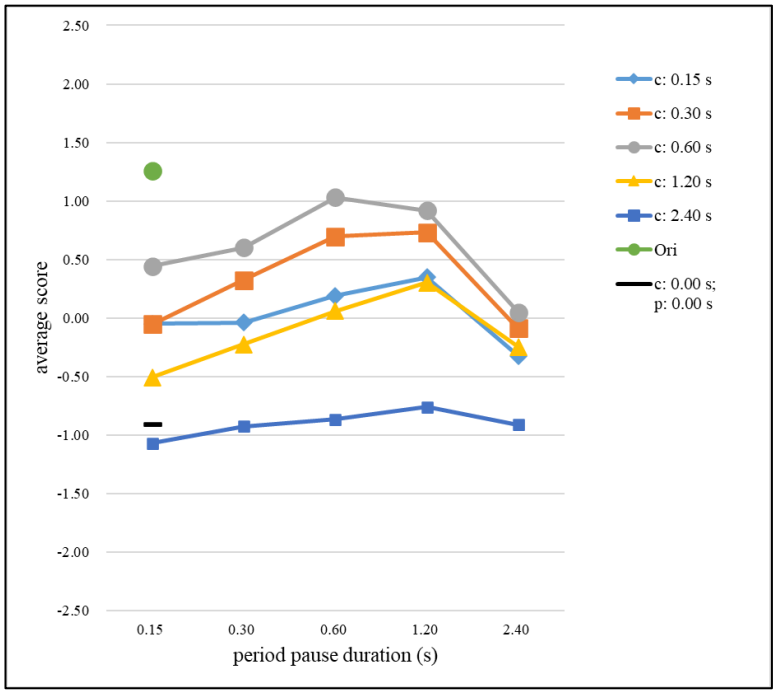

Figure 3. The relationship between pause duration and the speech Naturalness factor in Experiment 2. Letters "c" and "p" refer to the pause duration of commas and periods, respectively.

4. Conclusions

Comparing these results from Mandarin Chinese with previous research on English punctuation duration (Liu et al., 2019), we can find differences and similarities between the results obtained for the two languages. The main difference was that the average factor scores of the Naturalness factor for Chinese original sentences were higher than for English ones (0.82 in Experiment 2 for English, Liu et al., 2019; 1.27 in Experiment 1 and 1.25 in Experiment 2 in the present study with Mandarin Chinese). One reason for the higher factor scores for Chinese could be that the writing style of the speech stimuli used here was somewhat different. In the study with English sentences, public speeches were used (Liu, Nakajima, & Elliott, 2018; Liu et al., 2019), while in the present study announcements were

used. Another reason for the difference in factor scores could be that the speaking rates of the original speech stimuli used for each language were different. More experiments are required.

The similarities were very clear. Obviously, a longer pause duration resulted in a slower perceived speech rate. Most importantly, there is an optimum pause duration, which is 0.6 s, for both commas and periods, and for both English sentences (Liu, Nakajima, & Elliott, 2018; Liu et al., 2019) and in the present study with Mandarin Chinese. In the future, we wish to further study whether the speaking style (i.e., variations in speaking style to convey the purpose of the text), the participants' native language, or manipulating the duration of pauses at other punctuation marks affect the perceived speech rate and/or naturalness or not. Besides, it would be very fruitful to know whether the same results can be found for other languages as well. If so, our results can help to develop more suitable artificial speech technology (either speech generation or recognition). Using a constant pause duration may also be useful when practicing the delivery of speeches or lectures. The fact that to fix pause duration at 0.6 s makes the speech as natural as the original speech is practical for education.

Acknowledgements

This study was supported by JSPS KAKENHI Grant Number 19H00630. We are indebted to all the participants in the experiments and all the members of the Perceptual Psychology Laboratory, Kyushu University.

References

Curtin, S., Mintz, T.H., Christiansen, M.H. (2005). Stress changes the representational landscape: evidence from word segmentation. *Cognition*. Volume 96, Issue 3, July 2005. pp. 233-262.

General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China & Standardization Administration. (2011). *General rules for punctuation (GB/T 15834-2011)*. Beijing: Standard Press of China.

Horii, Y. (1983). An automatic analysis method of utterance and pause length and frequencies. *Behavior Research Methods & Instrumentation*. Volume 15, Issue 4, pp.449-452.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics.* Volume 45. July 2014. pp. 76-90.

Li, F. (2013). The effect of speakers' sex on voice onset time in Mandarin stops. *The Journal of the Acoustical Society of America.* Volume 133. EL142 (2013).

Lisker, L., & Abramson, A.S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*, 20:3. pp. 384-422.

Liu, S., Nakajima, Y., Elliott, M.A. (2018). Pause Duration Influences Impressions of Speech Style in English Public Speaking: Comparison of English Speakers and Chinese Speakers. *Japanese Society of Music Perception and Cognition.* 2018(2)-7.

Liu, S., Nakajima, Y., Elliott, M.A., Chen, L., Remijn, G.B., Arndt, S., Pang, Z. (2019). Pause Duration Influences Impressions of English Speech Style Rated by Native and Non-native Speakers. In du Bois, N., Arndt, S., Özsoy, E.V., Bayraktar, S., Gülbetekin, E., Elliot, M.A. (Ed.), *Conference Proceedings, 35th Annual Meeting of the International Society for Psychophysics, Fechner Day 2019* (pp. 14).

Straus, J., Kaufman, L., & Stern, T. (2014). *The Blue Book of Grammar and Punctuation* (11th ed.). Indianapolis, IN: Wiley.

Stuckenberg, A., & O'Connell, D.C. (1988). The Long and Short of It: Reports of Pause Occurrence and Duration in Speech. *Journal of Psycholinguistic Research.* Vol. 17, No.1.

Tosi, O.I. (1965). A Method for Acoustic Segmentation of Continuous Sound into Pauses and Signals and Measurement of Segment Durations. *Unpublished doctoral dissertation*. Ohio State University.

Contact email: Mingji Lin: vodkaddicter@gmail.com
Gerard B. Remijn: remijn@design.kyushu-u.ac.jp