SPEECH RATE AS A REFLECTION OF VARIANCE AND INVARIANCE IN CONCEPTUAL PLANNING IN STORYTELLING

Janice Fon
The Ohio State University, USA

ABSTRACT

This study investigates conceptual planning and cognitive rhythm as they may be reflected in speech rate fluctuation. According to the hierarchy principle, speech rate is slower when higher level planning is at work and faster when the planning is complete. Two four-frame comic strips were used to elicit speech. Speech rate was calculated in terms of syllables per second and a low-pass smoothing filter was used to reduce the impact of idiosyncratic fluctuation contributed by individual syllables. Results showed that most speakers were sensitive to story structures, although their narration strategies varied. Speakers aimed for planning units that accommodated story part boundaries. A solely clause-based planning strategy is not preferred.

1. INTRODUCTION

Speech rate differs not only from person to person, but also from time to time within the same person's speech. Previous studies have related speech rate with different functions. Some have claimed that speech rate is an indicator of memory span [1, 2, 3, 4]. Children and elderly people tend to have slower speech rates than younger adults because they have shorter memory spans. Higher speech rates are also more detrimental to perception for them for the same reason. Others have linked speech rate with syntactic [5, 6, 7, 8] or prosodic structures [9, 10]. Speech rate becomes progressively slower at syntactic/prosodic boundaries and hesitation pause is sensitive to syntactic complexity. Another viewpoint is to link speech rate with communication and conceptual planning [11, 12, 13]. Planned speeches are more fluent and pause-free than extemporaneous ones. When misunderstanding occurs, speakers can consciously monitor and alter speech rate in order to aid communication.

According to Henderson *et al*. [14] and Beattie [15, 16, 17], speech comes in alternating cycles of hesitant and fluent phases. In the hesitant state, speakers are preoccupied with goal elaboration and information retrieval, or macroplanning, and pause time is much longer than phonation time. On the other hand, during the fluent phase, speakers execute their plans made in the hesitant phase and pausing is due primarily to microplanning, or lexical item selection. As a result, phonation time is much longer than pause time. A temporal cycle consists of a hesitant (planning) phase and a fluent (executing) phase. In general, more conceptual coherence exists within a cycle than across different cycles. When speakers shift to a new topic, another elaboration cycle will start.

The hierarchy principle proposed by Berger and diBattista [11] and Berger *et al.* [12] claims that higher level planning such as the structure and sequencing of message content has higher demands and priorities on cognitive resources than lower level activities, such as the demands of speech motor control and coordination. According to this principle, resources are less available for speech motor planning during the conceptual planning stage since the latter takes up a large amount of cognitive resources. Any other level of planning is thus thwarted to a certain extent during this stage. On the other hand, in the execution stage, more cognitive resources are available for motor-phonetic planning.

In this study, we focus on speech rate and its link to conceptual planning. If speech rate is a reflection of memory span and processing load on cognitive capacity [3, 13], and if speech rate has lower priority in sharing cognitive resources as the hierarchy principle claims [11, 12], then it should reflect how planning is laid out during speech. In other words, the prediction is that speech is slower when the talker is forming concepts but faster when concepts are being verbalized.

2. METHOD

1.1. Subjects

Five female and five male subjects participated in this study, all of whom were students from National Taiwan University. Their ages ranged from 20 to 25 years old at the time of recording.

1.2. Materials

In order to elicit speech, two four-frame comic strips of different structures with no printed dialogues were chosen from *Shuangxiangpao*, a very famous comic series in Taiwan. Comic 1 is of an AAAB structure and there is a high correlation between frame boundaries and the progression of story plots [IMAGE 0430.JPG]. Comic 2 is more of an ABCD structure and the relationship between frame boundaries and story plots is not as clear-cut.

1.3 Procedure and Measurement

Subjects were seated in a quiet room and were given the comic strips. They were told to study the comic strip and retell the story afterwards. Recordings were made individually with a SONY TCM-5000EV recorder and a SONY ECM-G3M super-directional microphone. Transcriptions were done afterwards

The unit of rate calculation was set at the syllable level rather than the conventional sentential level since real-time rate monitoring is desired. Duration of each syllable was determined by both the waveform and the spectrogram on CSL KAY4300. Afterwards, the inverse of the duration of each syllable was taken to calculate the speech rate (syll/sec). Pauses and nonlinguistic utterances were treated as individual units and were included in rate calculation since they might be indicators of conceptual planning and their existence might also contribute to rate perception. Therefore, their durations were also measured and their inverses were taken in order to calculate their contribution to rate.

By taking the inverse of each linguistic and nonlinguistic unit, we have included a tremendous amount of idiosyncratic fluctuation contributed by individual syllables. Since what we are interested in is the general trend of the speech rate, a low-pass smoothing was done on the syllable-by-syllable speech rate measure. The assumption is that "intrinsic" segmental effects on the syllable rate will fluctuate more rapidly than effects of alternation between the slower speech rate during intervals of high cognitive load during conceptual planning and the faster speech rate during intervals of low cognitive load during speech execution. Smoothing was done by an FFT inverse filter for each set of speech rate data. A cutoff frequency of .3 Hz was chosen to filter out high frequency fluctuations.

3. RESULT

3.1. Summary Measures

Table 1 shows the summary measures of both narrations. For all three measures, the result of three 2 (comic) × 2 (order) mixed factorial designs showed no significant main effect of comic type. The order of the comic strip given is not significant, either. However, there is a high correlation between average rates of the two comic strips ($r = .88$, $p < .05$), indicating that the overall speaking rate of the speakers are fairly stable across different trials.

	syllable (no.)		duration (sec.)		avg. rate (syll/sec)	
	I	II	I	II	I	II
mean	183.3	148.2	38.32	32.43	4.91	4.86

Table 1: Summary measures.

3.2. Fluency Measures

Fluency was measured in terms of pauses. Two kinds of pauses were distinguished—unfilled (silent) pause and filled pause. Four kinds of filled pauses were counted—*en*, *e*, *ei*, and *m*. Table 2 shows the fluency measures in terms of percentages. A 2 (comic) × 2 (pause type) × 2 (order) mixed factorial design showed that there was a significant main effect of pause type on the percentage of time spent [$F(1, 8) = 60.35$, $p < .05$, $\hat{\eta}^2 = .88$]. However, comic type does not impose any significant effect, indicating that fluency level was constant across trials. Order does not have any significant effect.

	unfilled pause (%)		filled pause (%)	
	I	II	I	II
mean	29.97	26.27	1.61	1.78

Table 2: Fluency measures.

3.3. Rate Cycles

Based on the hierarchy principle, planning can be determined by speaking rate alternations. Figure 1 is an example of this for LZH's narration for Comic 2.

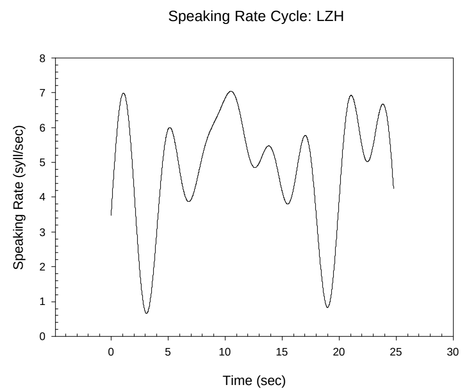

Figure 1: Planning cycle alternation in terms of speaking rate alternation after smoothing for subject LZH in Comic 2

Table 3 shows the summary of the number of rate cycles on average and the mean duration of the cycles. A 2 (comic) × 2 (order) mixed factorial design showed that no significant main effect of comic type and order exist. Both comic strips elicited about the same number of rate cycles. A paired correlation showed that there is a significant medium correlation between the number of cycles and total percentage of pause in Comic 2, but not in Comic 1 ($r = .67$, $p < .05$). In other words, the more hesitant one is in narration, the more cycles one has to make.

	cycles (no.)		mean duration (sec.)	
	I	II	I	II
mean	9.60	8.50	3.99	3.78

Table 3: Rate cycles.

3.4. Narration Strategy

3.4.1. Comic 1. Since different people have different ways of describing the comic strip, only the story parts that were mentioned by all the subjects were considered indispensable and analyzed accordingly. Table 4 shows what the story parts (P) are. Notice that both P_1 and P_2 pertain to the first frame. All the other descriptions have a one-to-one correlation with the frames—P_3 with Frame 2, P_4 with Frame 3, and P_5 with Frame 4. Most of the subjects used a narration strategy that is more or less chronologically ordered, except for LJS and ZZH. They used a narration strategy that combined the first three frames together instead of describing them in a chronological order.

		mean duration (sec.)
P_1	couple getting married	5.02
P_2	first sign holder	5.88
P_3	second sign holder	3.63
P_4	third sign holder	4.04
P_5	pastor's word	4.86

Table 4: Description in Comic 1 and its mean duration.

Table 4 shows on average how much time subjects spent on each of the five essential parts of the narration. The statistics of LJS and ZZH were not included in the calculation of mean due to their apparent different narration strategies. Since P_2 to P_4

constituted a repeating theme of people showing the couple signs, asking them to reconsider the marriage, a decreasing amount of time allotment is expected. A within-subject single-factor design showed that there was a near significant main effect of repeating story parts on the time spent [F (1.77, 12.42) = 3.62, p = .06]. The effect size was large, $\hat{\omega}^2$ = .23 [19]. This suggests that a main effect of repeating story parts is probably present, but our experiment here did not have significant power to detect it. A post hoc pairwise comparison after Bonferroni adjustment showed that there was a near significant difference between P_2 and P_3 (p = .05).

3.4.2. Comic 2. Table 5 shows the essential story parts of Comic 2. All of the subjects mentioned the four parts. It seems that the four parts corresponded neatly to the setting-problem-solution-result sequence. P_1 and P_2 refer to the first frame. P_3 refers to both the second and the third whereas P_4 refers to the last frame. Unlike Comic 1, there is no clear one-to-one correlation between frames and story parts.

		mean duration (sec.)
P_1	the clinic	7.00
P_2	no business	2.25
P_3	change sign	4.84
P_4	husband pulling wives/good business	7.80

Table 5: Description in Comic 2 and its mean duration.

Table 5 shows on average how much time subjects spent on each of the essential parts. The general trend seems to be that P_4 tended to be the longest where as P_2 the shortest. A single factor within-subject design shows that there is a significant main effect of story parts on duration [F (2.33, 20.93) = 7.31. p < .05, $\hat{\omega}^2$ = .19]. Post hoc comparison after Bonferroni adjustment also showed that subjects allotted more time to P_4 (result) than P_2 (problem) and P_3 (solution). More time was given to the setting (P_1) than the problem (P_2) (p < .05).

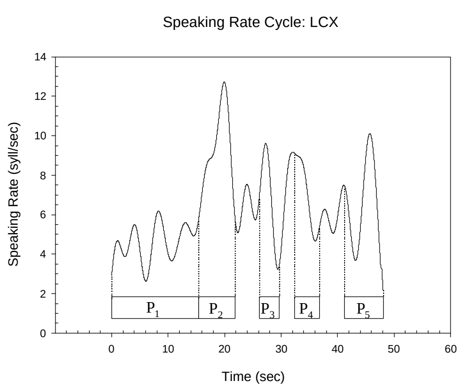

Figure 2: An illustration of the relationship between rate cycles and story parts, from LCX, Comic 1.

3.5. Narration Strategy *vs.* Rate Cycles

There are two ways a story part can be correlated with a rate cycle. It can either be subsumed within a rate cycle, or it can straddle between two or more cycles. Figure 2 demonstrates how the rate cycles and the story parts can correlate with each other. This is from subject LCX in her narration for Comic 1. Notice that P_2, P_3, and P_4 are each within one speaking rate alternation cycle, whereas P_1 and P_5 straddle between more than one cycles. LCX used 4 cycles to finished P_1 and 1.5 cycles for P_5.

Many factors can affect the correlation between story parts and rate cycles. Table 6 shows the six potential non-grammatical causes for a cycle to end. Among the six types, hesitation is the most frequent. Also, there were more instances of hesitation in Comic 2 than in Comic 1. On average, the indispensable story parts in Comic 1 took 1.48 cycles to finish and those in Comic 2 took 1.90. A paired t-test showed that the effect of comic strip type was near significant [t (9) = –2.10, p = .065].

	hesitation		repetition		speech error	
	I	II	I	II	I	II
no.	10	27	0	3	2	3
	self-correction		lexical selection		laughter	
	I	II	I	II	I	II
no.	2	0	5	0	0	3

Table 6: Summary of the instances caused by non-grammatical factors for ending a cycle.

Aside from the cycles influenced by non-linguistic factors, there is a one-to-one correspondence between story parts and rate cycles. In fact, 46 out of 51 story part instances showed this correspondence.

4. DISCUSSION

Speech rate has been considered as a reflection of memory span [1, 2, 3, 4]. The larger the memory span, the faster the speech rate. This is evidenced by the fact that the correlation between the average speech rate of the two narratives within the same subject is as high as r = .88, as shown in Table 1. In other words, the average speech rate is fairly constant for each speaker across tasks. This would be true if speech rate is a reflection of memory span since one would not expect the memory span of a subject to show much variation. The average speech rate for a narrative is 4.84 syll/sec in this study, and most subjects were within the range of 4 – 5 syll/sec. Therefore, we would conclude that such a rate is representative of narration speech rate in Mandarin. Speakers with rate as fast as 7 syll/sec are possible (as in the case of XYL), but rare.

The fluency measure also seems to be held constant across different tasks. Most speakers have a certain constant proportion of speech that is dedicated to pauses, filled or unfilled. On average, 30% of the total utterance time spent in pauses is representative of this type of narration.

The assumption that speech rate reflects the size of the memory span could not explain well why rate cycles exist. If speech rate is only a reflection of the memory span, then fluctuations of speech rate should mean changes in the memory span, which is very unlikely, since most narratives elicited were only about 30 – 40 sec., a very short time period for memory spans to fluctuate in such a great degree.

However, if speech rate is a reflection of both memory span and processing load on cognitive capacity [3, 6, 13], then such a fluctuation would be indicative of how people plan. The heavier the processing load, the less likely it is for speakers to talk at a fast speed [11, 12]. This assumption could explain why there is a medium correlation of r = .67 between fluency measures and number of rate cycles. In other words, given the same story, the more uncertain one is about what he speaks, the more processing cycles one has to make.

Most subjects showed an influence of the story structure in their narration. As indicated before, Comic 1 is of an AAAB structure whereas Comic 2 is ABCD. If speakers were sensitive to the story structure, one would expect that certain verbal or non-verbal indications of the repeating scenes could be found in narration of Comic 1, but not in Comic 2. LJS and ZZH made the most overt indications. They combined the first three frames together, and described them as a group. The others also allotted a decreasing amount of time in various degrees to the first three frames of Comic 1, indicating that not as much planning time was needed for a repeating theme.

Comic 1 had a clearer one-to-one relationship between frames and story parts whereas in Comic 2, the relationship is not as straightforward. If frame-story part correspondence could aid speakers in planning their story, then one would expect subjects to show more uncertainty in Comic 2. Although there did not seem to be any difference in fluency measures (Table 2), more instances of hesitations were detected in Comic 2 (Table 6)

The correspondence between a story part and a rate cycle is intriguing. If speech rate is a reflection of cognitive load, then fluctuation of speech rate can act as an approximate index showing how the cognitive load varies and thus how people plan. After filtering out all the non-linguistic factors, we found that in most cases, subjects tended to have a one-to-one correspondence between story parts and rate cycles. It seems that speakers aim for planning their narration with regards to complete discourse units.

5. CONCLUSION

In this study, the fluctuating pattern of speech rate in story narration is examined using different comic strips. Subjects were sensitive to different comic types. This is reflected through various ways, overt or covert. Generally speaking, invariance does not lie in the absolute speed, or the 'optimal' speed of each speaker. Instead, invariance of speed lies in the fluctuating patterns and its correlation with story parts. Most speakers used a strategy so that there is a one-to-one relationship between rate cycles and story parts, although exceptions do occur. Therefore, it is suggested that speakers would aim for planning their speech to accommodate the story parts. However, when higher level planning is not feasible, they would fall back on lower level units such as clauses. Hesitation and speech errors would occur when even this level of planning is corrupted. If speakers show a high correlation between rate cycles and story parts, it would be very likely for listeners to also use such a cue in processing incoming signals. Further studies are needed in order to examine the relationship between temporal cycles and perception.

ACKNOWLEDGMENTS

Many thanks go to Dr. Keith Johnson, Dr. Mary Beckman and all of the colleagues in the Phonetics Lab in the Ohio State University. Special thanks go to Ms. Laura Hsiu-min Liu for kindly sharing her recording data with me. I would also like to thank Mr. Pang-Hung Hsu for technical support.

REFERENCES

[1] Haselager, G. J. T., Slis, I. H., and Rietveld, A. C. M. 1991. An alternative method of studying the development of speech rate. *Clinical Linguistics & Phonetics*, 5 (1) 53–63.

[2] Wingfield, A., Wayland, S. C., and Stine, E. A. L. 1992. Adult age differences in the use of prosody for syntactic parsing and recall of spoken sentences. *Journal of Gerontology: Psychological Sciences*, 47 (5), 350–356.

[3] Henry, L. A. 1994. The relationship between speech rate and memory span in children. *International Journal of Behavioral Development*, 17 (1), 37–56.

[4] Katz, W. F., Beach, C. M., Jenouri, K., and Verma, S. 1996. Duration and fundamental frequency correlates of phrase boundaries in productions by children and adults. *Journal of the Acoustical Society of America*, 99 (5), 3179–3191.

[5] Brubaker, R. S. 1972. Rate and pause characteristics of oral reading. *Journal of Psycholinguistic Research*, 1 (2), 141–147.

[6] O'Connell, D. C., and Kowal, S. 1972. Cross-linguistic pause and rate phenomena in adults and adolescents. *Journal of Psycholinguistic Research*, 1 (2), 155–164.

[7] Ruder, K. F.and Jensen, P. J. 1972. Fluent and hesitation pauses as a function of syntactic complexity. *Journal of Speech and Hearing Research*, 15, 49–60.

[8] Max, L. and Caruso, A. J. 1997. Acoustic measures of temporal intervals across speaking rates: variability of syllable- and phrase-level relative timing. *Journal of Speech, Language, and Hearing Research*, 40, 1097–1110.

[9] Ferreira, F. 1993. Creation of prosody during sentence production. *Psychological Review*, 100 (2), 233–253.

[10] Dankovičová, J. 1997. The domain of articulation rate variation in Czech. *Journal of Phonetics*, 25, 287–312.

[11] Berger, C., and diBattista, P. 1993. Communication failure and plan adaptation: if at first you don't succeed, say it louder and slower. *Communication Monographs*, 60, 220–238.

[12] Berger, C. R., Knowlton, S. W., and Abrahams, M. F. 1996. The hierarchy principle in strategic communication. *Communication Theory*, 6 (2), 111–142.

[13] Mitchell, H. L., Hoit, J. D., and Watson, P. J. 1996. Cognitive-linguistic demands and speech breathing. *Journal of Speech and Hearing Research*, 39, 93–104.

[14] Henderson, A., Goldman-Eisler, F., and Skarbek, A. 1966. Sequential temporal patterns in spontaneous speech. *Language and Speech*, 9, 207–216.

[15] Beattie, G. 1980. Encoding units in spontaneous speech: some implications for the dynamics of conversation. In H. W. Dechert & M. Raupach (Eds.), *Temporal Variables in Speech* (pp. 131–143). The Hague: Mouton.

[16] Beattie, G. 1983. *Talk: An Analysis of Speech and Non–verbal Behavior in Conversation*. Milton Keyes: Open University Press.

[17] Beattie, G. 1984. "Are there cognitive rhythms in speech?"—a reply to Power, 1983. *Language and Speech*, 27, 193–195.

[18] Du Bois, J. W. 1991. Transcription design principles for spoken language research. *Pragmatics*, 71–106.

[19] Cohen, J. 1977. *Statistical Power Analysis for the Behavioral Sciences* (rev. ed.). New York: Academic Press.

[20] Maclay, H., and Osgood, C. E. 1959. Hesitation phenomena in spontaneous English speech. *Word*, 15, 19–44.