

Variance and Invariance in Speech Rate as a Reflection of Conceptual Planning

Janice Fon

The Ohio State University

ABSTRACT

This study investigates the variance and invariance in speech rate as a reflection of conceptual planning and cognitive rhythm. A four-frame comic strip was used to elicit speech and low-pass smoothing was done afterwards to filter out high frequency noise. Results showed that subjects invariably planned narration in terms of story plots. Variance lies in the way subjects synchronize planning and execution stages. Some tended to start the execution stage before the planning stage ends while others were inclined to speak only after macroplanning was done. Lexical retrieval failure is one of the main causes for disruptive story-plot-based temporal cycle.

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 different functions with speech rate. Some have claimed that speech rate is an indicator of memory spans [1–4]. Children and elderly people tend to have slower speech rate than 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 [5–8] or even prosodic structures [9–10]. It is found that 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 [11–13]. 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 [15–17], 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, whereas in the fluent phase, speakers execute their plans made in the hesitant phase and pausing is due primarily to microplanning, or lexical item selection. In general, more conceptual coherence exists within a cycle than across different cycles. When speakers shift to a new focus, 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 alternations, such as speech rate and vocal intensity. According to this principle, resources are the lowest in supply during the conceptual planning stage. As a result, any other level of planning should be thwarted to a certain extent during this stage. On the other hand, in the execution stage when only low level motor-phonetic planning is undergoing, cognitive resources should be more exuberant.

In this study, we would like to look at 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 hierarchical principle claims [11, 13], then it should reflect how planning is laid out during speech. In other words, speech should be slower when the talker is forming concepts but faster when concepts are being verbalized and executed.

2. METHOD

2.1. Subjects

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

2.2 Materials and Procedure

In order to elicit speech, a comic strip of four frames with no dialogues was chosen from *Shuangxiangpao*, a very famous comic series in Taiwan [IMAGE 0842.JPG]. Subjects were seated in a sound-treated room and were told to study the comic strip and retell the story afterwards. Recordings were made individually with SONY TCM-5000EV recorder and SONY ECM-G3M super-directional microphone. Transcriptions were done afterwards in terms of intonation units (IU) following the discourse analysis tradition [18].

2.3 Measurement

Two main problems exist concerning rate calculation. One is how to deal with pause and nonlinguistic utterances. The other is the unit of calculation. In this study, pause and other nonlinguistic utterances are treated as segments and are included in rate calculation since they are indicators of conceptual planning and their existence contributes to rate perception. The unit of calculation is the syllable. Duration was determined by both the waveform and the spectrogram of the segment on CSL KAY4300. Afterwards, the inverse of syllable duration was taken to calculate the speech rate (syll/sec). Since what is of interest here is the general trend of the speech rate instead of the idiosyncratic fluctuation contributed by individual syllables, the lowpass-smoothing filter from SigmaPlot was used to cut off high frequencies. This algorithm takes the rawdata and performs fast Fourier transform in order to filter out the designated percentage of high frequency noises.

The narratives were grossly divided according to frames. Inferences made before Frame 1 (F_1) will be labeled as the introduction (I) and those after Frame 4 (F_4) as the conclusion (C). Those made between frames were considered as transitions ($T_{\#}$) and were labeled numerically according to the preceding frame.

3. RESULTS

Although there are only four frames in the comic strip,

interspeaker variability is still high. Table 1 shows the basic statistics of the four subjects, including the number of syllables uttered, total narration duration, average speech rate, and narration parts. Narration styles also vary. All subjects used some kind of introduction and conclusion. LZY and LJY used no between-frame transitions whereas CCF and LZE used one (T_3).

ID	#Syll	Duration (sec)	Rate (syll/sec)	Narration
LJY	136	26.76	5.08	IF ₁ F ₂ F ₃ F ₄ C
LZY	285	67.61	4.22	
CCF	121	22.93	5.28	IF ₁ F ₂ F ₃ T ₃ F ₄ C
LZE	167	31.97	5.22	

Table 1: Basic statistics of the four subjects

3.1. LJY & LZY: IF₁F₂F₃F₄C

Both LZY and LJY used the same narration strategy. They provided an introduction and a conclusion but left the description of the four frames transition-free. In total, there were 6 parts in their narration.

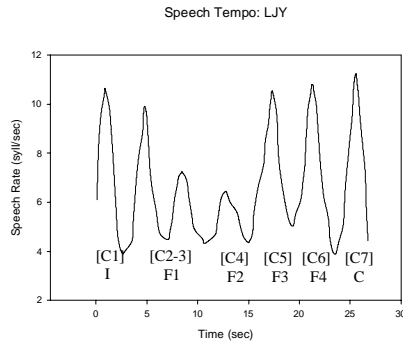


Figure 1: Speech rate patterns of LJY.

88% of the high frequency was removed for LJY. As shown in Fig. 1, the speech tempo pattern is an alternation of peaks and valleys. If cycles are defined as having a peak and two flanking valleys, LJY had 7 temporal cycles. All of the cycles except the second and the third (C_{2-3}) coincide with story part boundaries (denoted by square brackets in Fig. 1) in a one-to-one relationship. As shown in (1), C_2 , a partial description of F_1 , ends with the retrieval of *mama* ‘mom’. LJY had difficulties deciding what the figure in the comic was, as was evidenced by both the long successive pause breaks (P) and the phrase *buzhidao* ‘don’t know’. Generally speaking, the planning strategy of LJY is story-plot-based.

(1) LJY: [F₁, C₂]

- 4 *er tamende* P
and their .33
...and their...
- 5 *buzhidao na yige ren de* P
don’t know which one person Gen. .36
...don’t know which one’s...
- 6 *mama P*_{C₂}
mom .60
...mom...

Although the narration of LZY was the longest of all four subjects, his narration was full of filled pauses, speech error, and repetition, which created great noise for speech rate calculation. As a result, the filtering threshold was set at 92%. As shown in Fig. 2, LZY had 8 temporal cycles.

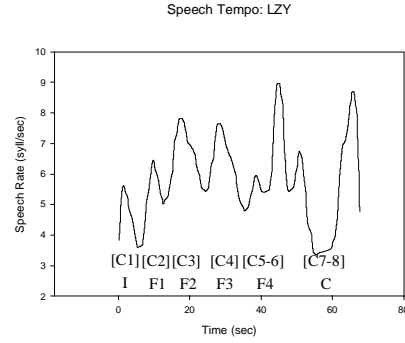


Figure 2: Speech rate patterns of LZY.

Unlike those of LJY, his alternating cycles are more variable in terms of peak speaking rate. However, his C_{1-4} have a one-to-one relationship with $I-F_3$. There are two cycles contributing to F_5 and the conclusion, respectively, both of which are due to an overlap in speech planning and execution. As shown in (2), the hesitation and filled pauses show that LZY was planning and speaking at the same time. It is interesting to find that such an overlap occurs toward the end of the narration. This is understandable since narration at this point is near the transition from story description to conclusion, where subjects have to plan for themselves instead of having a visual aid (the comic strip) to help them. However, generally speaking, LZY also used a plot-based planning strategy in his narration.

(2) LZY: [C, C₇₋₈]

- 53 ... *jiehuen* P
... marry .10
...to marry...
- 54 *e shi yige* P
er... BE one .57
er...(marriage) is one...
- 55 *m...* P
hm... 3.12
hm...
- 56 *e* P
er .77
er...
- 56 *e* P
er .32
er...

3.2. CCF & LZE: IF₁F₂F₃T₃F₄C

CCF and LZE both used a transition between the first three frames and the final one (T_3). Therefore, in total, there were 7 parts in their narration.

CCF was a fluent speaker and his narration was almost error-free. Therefore, the filtering threshold was set at 85%. As shown in Fig. 3, he had 7 alternating cycles. Except for the introduction and F_4 , all of the other cycles show a one-to-one relationship

with story parts.

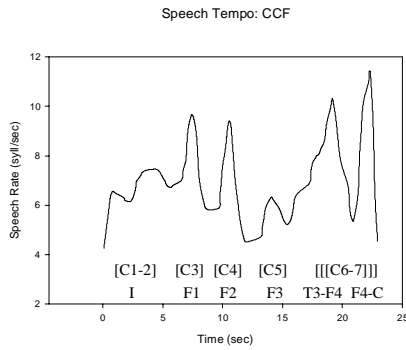


Figure 3: Speech rate patterns of CCF.

C_{1-2} contribute to the introduction. The relatively slow speaking rate in C_1 is due to an overlap between speech planning and execution. As shown in (3), C_1 is full of successive silent pauses. A stretch of words without pause in C_1 is in average 2.5 syllables long. Words such as *wo xiang* ‘I think’ and *haoxiang* ‘like’ are also telling of the uncertainty in the planning process. However, as narration went on, speech execution ran more smoothly.

(3) CCF: [I, C_{1-2}]

- | | | | | |
|---|--------------------|---------------|----------------------|----------------|
| 1 | <i>wo</i> | <i>xiang</i> | <i>P</i> | |
| | I | think | .19 | |
| | I think,... | | | |
| 2 | <i>tamen</i> | <i>jiushi</i> | <i>P</i> | |
| | they | just | .55 | |
| | They just... | | | |
| 3 | <i>haoxiang</i> | <i>C1</i> | <i>P</i> | |
| | like | .17 | | |
| | (they) are like... | | | |
| 4 | <i>zai</i> | <i>zou</i> | <i>nage</i> | <i>hongtan</i> |
| | Asp. | walk | that | red carpet |
| | (they) | are walking | on the red carpet... | |

C_{6-7} contribute to three story parts— T_3 , F_4 , and the conclusion. As shown in (4), C_6 ends at the phrase *laibuji* ‘late’ in IU_{17} . This is because this phrase was already used in IU_{15} , and CCF had intended to substitute the phrase with synonyms but failed. Another piece of evidence is the long pause at the end of IU_{16} , indicating that CCF was trying to search for an appropriate synonym. In other words, if it were not for this lexical retrieval failure, the planning strategy for CCF would be to group these three narration parts together. That is, transition is used only as a “filled” planning process. It does not require much cognitive resources itself. Generally speaking, CCF’s planning is also story-plot-based.

(4) CCF: [T_3 —C, C_{6-7}]

- | | | | | |
|----|-----------------------------|-----------------|---------------|------------------|
| 14 | ... | <i>haoxiang</i> | <i>yijing</i> | <i>P</i> |
| | ... | as if | already | .07 |
| | ...as if (it is) already... | | | |
| 15 | <i>kuai</i> | <i>laibuji</i> | <i>le</i> | <i>zheyangzi</i> |
| | almost | late | Asp. | this way |
| | ...almost too late... | | | |
| 16 | <i>ranhou</i> | <i>dao</i> | <i>disige</i> | <i>P</i> |
| | then | to | fourth | .53 |

Then to the fourth (picture)...

- | | | | | | | | |
|----|----------------|------------|-----------|-------------|-----------|------------------|----------|
| 17 | <i>zhongyu</i> | <i>lai</i> | <i>C6</i> | <i>buji</i> | <i>le</i> | <i>zheyangzi</i> | <i>P</i> |
| | finally | late | Asp. | this way | .43 | | |

Finally, it is too late...

- | | | | | | |
|----|----------------|---------------|--------------|-----------|-----------|
| 18 | <i>tamen</i> | <i>zoudao</i> | <i>zheli</i> | <i>le</i> | <i>C7</i> |
| | they | walk to | here | Asp. | |
| | They got here. | | | | |

LZE had a speech style of frequent repetition. To eliminate the noise created by such, the filter threshold was set at 91%. As shown in Fig. 4, there are in total 7 alternating cycles.

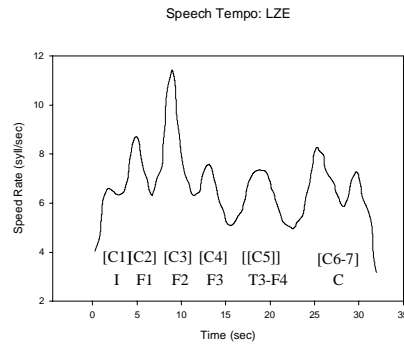


Figure 4: Speech rate patterns of LZE.

Fig. 4 shows the temporal cycles of LZE. Like CCF, the majority of the cycles demonstrate a one-to-one relationship with story parts. Exceptions lie in T_3 , F_4 , and the conclusion. T_3 and F_4 form only one cycle, C_5 , indicating that they are of one unit in terms of planning. C_6 and C_7 contribute to the conclusion. The existence of C_6 is due to overlapping in speech planning and execution, as shown in (5). LZE used both silent and filled pauses while doing speech planning. The recurring adverbs *dagai* ‘probably’ and *haoxiang* ‘as if’ are also telling cues showing that the speaker was not so sure of what he was talking about. However, generally speaking, LZE also planned his speech in terms of story plots.

(5) LZE: [C, C_{6-7}]

- | | | | |
|----|-----------------|-------------|------------|
| 22 | ... | <i>m</i> | |
| | ... | <i>m</i> | |
| | ...hm... | | |
| 23 | <i>dagai</i> | <i>P</i> | |
| | probably | .29 | |
| | probably... | | |
| 24 | <i>dagai</i> | <i>P</i> | |
| | probably | .06 | |
| | probably... | | |
| 25 | <i>dagai</i> | <i>yisi</i> | <i>shi</i> |
| | probably | meaning | BE |
| | Probably | it means | that... |
| 26 | <i>zhege</i> | <i>P</i> | |
| | this | .58 | |
| | well... | | |
| 27 | <i>haoxiang</i> | <i>P</i> | |
| | as if | .18 | |
| | as if... | | |
| 28 | <i>ni</i> | <i>dao</i> | <i>C6</i> |
| | you | go | priest |
| | You | go to the | priest... |

4. DISCUSSION

The same story told by different people can have very different interpretations. In this study, two types of narration styles are examined. One is $IF_1F_2F_3F_4C$ and the other is $IF_1F_2F_3T_3F_4C$. It is interesting to find that aside from the main story parts F_{1-4} , all four of the subjects added an introduction and a conclusion to their narration, even though it is only a simple description of a short comic strip. This implies that no matter how short or simple a story may be, introduction and conclusion are two essential parts of story grammar, even if when information concerning these two parts is not given. Another interesting finding is that when only one transition is added, as in the narration of CCF and LZE, an inclination of inserting a T_3 is strong. One possible explanation might be the comic strip chosen. Since the first three frames are almost of identical theme, it is likely that speakers tend to group them together and plan them as a unit. Another possibility might be that T_3 is a convenient way to plan the final conclusion. More data would be needed in order to determine the essentiality of T_3 .

Although the narration style may be different from person to person, speech rate fluctuation consistently correlates with story parts. If speech rate is an accurate reflection of conceptual planning and cognitive rhythm, this would indicate that people tend to plan speech in terms of plots when narrating. This correlation may be disrupted during lexical retrieval failure (LJY), overlapping of speech planning and execution (LZY, LZE), and transition (CCF, LZE). The latter is of the most interest. Although introduction, transition, and conclusion are all additional information provided by the speakers, transition seems to take up the least conceptual planning resources. Instead, transition acts like a filled planning phase and seemed to ease away the planning load. Further studies would be needed concerning the difference between transition on the one hand and introduction and conclusion on the other.

Speakers also have different priorities concerning speech planning. Most of the subjects finished preliminary planning before entering into speech execution stage. This is evidenced by a high correlation between temporal cycles and narration parts from the start (LJY, LZY, and LZE). However, the preliminary planning process does not seem to involve detailed lexical selection, which is why lexical retrieval can be a cause for disruptive temporal cycles (LJY). CCF was the only person who planned and spoke at the same time in the beginning of the narration. Interestingly, his narration was also free of repetition, filled pauses, and lexical retrieval failure which are common in extemporaneous speech. The fact that planning while speaking tends to happen towards the end of the narration (LZY, LZE) implies that macroplanning is not homogeneous. Instead, it is more detailed in places nearer to the time reference point but more general when farther apart.

5. CONCLUSION

In this study, the fluctuating pattern of speech rate in story narration is examined. It is found that people tend to plan narration in terms of story parts. This cognitive rhythm can be broken when lexical retrieval failure or synchronization of planning and execution occur. Interspeaker variations lie in the synchronization pattern of the two stages. Some tend to plan speech before execution while others prefer to process both at

the same time.

Generally speaking, invariance does not lie in the absolute speed, or the 'optimal' speed of each speaker. Instead, invariance of speed lies in its fluctuating patterns and correlation with conceptual planning. If speakers show a high correlation between rate and planning due to limited cognitive resources, it would be very likely for listeners to use such a cue in processing incoming signals. Further studies are needed in order to examine the relationship between temporal cycles and perception.

6. REFERENCES

1. Haselager, G. J. T., Slis, I. H., and Rietveld, A. C. M., *CL&P* 5(1), 53–63, 1991.
2. Wingfield, A., Wayland, S. C., and Stine, E. A. L., *J. Gerontology: Psychological Sciences* 47(5), 350–356, 1992.
3. Henry, L. A., *IJBD* 17(1), 37–56, 1994.
4. Katz, W. F., Beach, C. M., Jenouri, K., and Verma, S., *J. Acoust. Soc. Am.* 99(5), 3179–3191, 1996.
5. Brubaker, R. S., *J. Psycholinguist. Res.* 1(2), 141–147, 1972.
6. O'Connell, D. C., and Kowal, S., *J. Psycholinguist. Res.* 1(2), 155–164, 1972.
7. Ruder, K. F. and Jensen, P. J., *JSHR* 15, 49–60, 1972.
8. Max, L. and Caruso, A. J., *JSLHR* 40, 1097–1110, 1997.
9. Ferreira, F., *Psychol. Rev.* 100(2), 233–253, 1993.
10. Dankovičová, J., *J. Phon.* 25, 287–312, 1997.
11. Berger, C. and diBattista, P., *Communication Monographs* 60, 220–238, 1993.
12. Berger, C. R., Knowlton, S. W., and Abrahams, M. F., *CT* 6(2), 111–142, 1996.
13. Mitchell, H. L., Hoit, J. D., and Watson, P. J., *JSHR* 39, 93–104, 1996.
14. Henderson, A., Goldman-Eisler, F., and Skarbek, A., *Lang. Speech* 9, 207–216, 1966.
15. Beattie, G., In H. W. Dechert and M. Raupach, eds., *Temporal Variables in Speech*, 131–143, The Hague, Mouton, 1980.
16. Beattie, G., *Talk: An Analysis of Speech and Non-verbal Behavior in Conversation*. Milton Keynes: Open University Press, 1983.
17. Beattie, G., *Lang. Speech* 27, 193–195, 1984.
18. Du Bois, J. W., *Pragmatics: Quarterly Publication of the International Pragmatics Association* 1(1), 71–106, 1991.