

INTER-SPEAKER INTERACTION OF THE DURATION OF SENTENCES AND INTERSENTENCE INTERVALS

Kuniko Kakita

Department of Liberal Arts and Sciences
Toyama Prefectural University

ABSTRACT - The aim of the present study is to examine how one speaker's speech affects another speaker's speech. The speech parameters investigated are the sentence duration and the intersentence interval duration. The results of the study indicated that when the subjects 'took over' from the preceding speaker and read the remaining part of the speech material, both the sentence duration and the interval duration deviated from one's 'preferred' duration obtained in the single speaker readings. The deviation was mostly assimilative. The results of further analysis indicated that the sentence duration and the interval duration differed characteristically in the way they were affected by another speaker's speech.

INTRODUCTION

Speaker interaction has been studied mainly in the field of speech pathology, language development and sociolinguistics and psycholinguistics (Guitar et al., 1992; Jasnow & Feldstein, 1986; Matarazzo et al., 1963). For example, the studies focusing on the temporal patterns of conversational exchange have revealed that there are inter-speaker accommodations over the course of conversation, and that they are associated with aspects of speakers' personalities and social attitudes (Feldstein & Welkowitz, 1978; Jaffe & Feldstein, 1970). However, little is known about the actual manifestation of interspeaker interaction on the speech output or its underlying mechanism. The present study aims to examine the effect of speaker interaction on the temporal aspects of speech production. The specific questions asked are as follows: (i) Is one speaker's speech characteristics, viz., sentence duration and intersentence interval duration, affected by those of the preceding speaker? (ii) If so, how? (iii) Do sentences and intersentence intervals differ in the way they are affected by the preceding speaker's speech? (iv) What governs the degree of inter-speaker interaction in sentences and intersentence intervals?

METHOD

Materials

The utterance texts used in the present study were made up of sequences of five sentences. The five sentences had the following basic structure in common: 'X sanwa Y gakkadesu.', i.e. 'X is in the Y Department.'. 'X' was one of the following five Japanese last names of persons: Ohama, Okuma, Onuma, Otani, or Otomo. 'Y' was one of the following five Department names in Japanese: Gengo (Linguistics), Kango (Nursing), Rinri (Ethics), Ronri (Logic), and Shinri (Psychology). The persons' last names and department names were randomized independently and were combined to form the five sentences constituting an utterance text. Ten such utterance texts were prepared to be read by the subjects. All the sentences consisted of 10 syllables, or 15 moras. This meant that one could obtain the mean syllable duration and the mean mora duration by dividing the sentence duration by 10 and 15, respectively.

Experimental Procedure

Recording for each subject was carried out in two sessions, A and B, as shown in Figure 1. In Session A, the five consecutive sentences were read through by one and the same subject. In Session B, the first three sentences, prerecorded and stored in the computer, were played back to the subject through headphones. The subject took over from the preceding speaker (the prerecorded voice) after the

Experimental Sessions

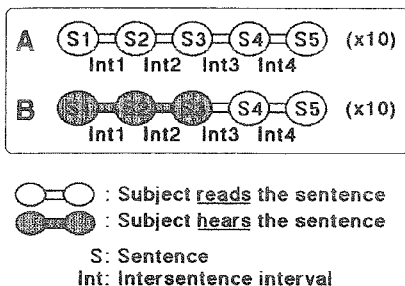


Figure 1. Experimental sessions

third sentence and read the remaining two sentences. The same set of ten utterance texts were used in the two sessions, so that in the later stage of analysis, comparisons could be made of the corresponding pairs of sentences/intervals in the two sessions.

The subjects were instructed 'not to insert a pause between the noun phrase ('X sanwa') and the predicate phrase ('Y gakkadesu') - the only probable intrasentence interval location - when reading the sentences. Aside from this instruction, the subjects were simply asked 'to take over from the preceding speaker and read the remaining sentences in a manner most natural to them'. There were five subjects. All of them read the utterance texts in Standard Japanese.

Recording

The utterances were recorded by use of a stereo digital audio cassette deck (Sony: TCD-D10) and an electret condenser microphone (Sony: ECM-67) in a sound-proof booth (Rion: AT-80) at Human Intelligence Laboratory, Kanazawa Institute of Technology, Kanazawa, Japan. The recording was carried out with the aid of a self-navigating experiment software custom-made to run on the NeXT Computer. The utterance texts to be read by the subjects were displayed on the computer screen. The subjects were free to 'redo' the reading any number of times.

Measurements

The speech signal was sampled at a rate of 22kHz, digitized with a 16 bit Analog-to-Digital converter, and was stored in the NeXT Computer by use of Digital Ears (Metaresearch, Inc., Portland, OR, U. S. A.). Data was inspected and measurements were made by using a customized speech editing software. Both the 'all passed' and '6-9 kHz band passed' speech waves were consulted in the measurement, the latter being used for a detailed observation of the [s] sound at the end of each sentence. The points measured were the beginning and the end of each sentence, i.e., the start of voicing for [o] in /ooCVCVsanwa/ and the end of friction noise for [s] in /CvNcVgakkadesu/, where the last vowel /u/ was devoiced and therefore the sentence ended in [s]. The duration of sentences and intersentence intervals, derived from the measurements, served as the basis of present analysis.

RESULTS AND DISCUSSION

In the discussion to follow, the five consecutive sentences constituting an utterance text will be called S1, S2, S3, S4, and S5. Similarly, the four intersentence intervals (henceforth simply intervals) will be called Int1, Int2, Int3, and Int4.

'Inherent' duration vs. 'Post-takeover' duration

In Session A, one and the same subject read the sentences through without any takeover. The sentence and interval durations obtained in this session may be considered to represent each subject's 'preferred' durations. In the present paper, I will call them the '(subject-)inherent' durations. In Session B, the subjects took over from the preceding speaker after S3. The durations of S4, S5, Int3 and Int4 in this session will be referred to as the 'post-takeover' durations. Also, the median sentence/interval duration of the preceding speaker derived from S1-S3 and Int1-Int2 that were played back to the subjects in Session B will be referred to as the preceding speaker's 'inherent' sentence/interval duration.

Fig. 2 compares the five subjects' 'post-takeover' durations (filled data points) with their 'inherent' durations (unfilled data points). The data points for both the 'inherent' and 'post-takeover' durations represent the median duration values derived from 20 sentences (S4 x 10 and S5 x 10) and 20 intervals (Int3 x 10 and

Int4 x 10). The cross in the figure indicates the 'inherent' sentence/interval duration of the preceding speaker, and is derived from the 30 sentence samples (S1 x 10, S2 x 10 and S3 x 10) and 20 interval samples (Int1 x 10 and Int2 x 10) obtained in Session B.

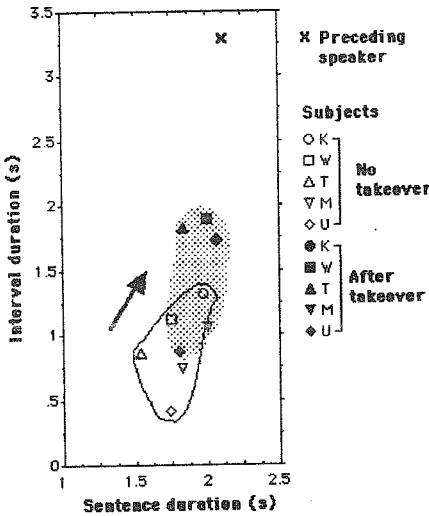


Figure 2. The duration of sentences and intervals in the two experimental sessions [Session A (no takeover): unfilled data points; Session B (takeover after S3): filled data points]

The subjects' 'inherent' sentence duration ranged from 1.53 s to 1.97 s, while 'inherent' interval duration ranged from 0.45 s to 1.32 s. The preceding speaker's 'inherent' durations for sentence and interval were 2.12 s and 3.29 s, respectively. This means that all the subjects took over from a speaker with a slower speech production rate.

The subjects' 'post-takeover' values ranged from 1.80 s to 2.06 s for sentences, while they ranged from 0.87 s to 1.90 s for intervals. For all five subjects, the median 'post-takeover' durations of the sentences and intervals were longer than the corresponding 'inherent' durations. This indicates that the sentence and interval durations of all the subjects were affected by the preceding speaker's speech. A closer comparison of the corresponding pairs of sentences and intervals in the two sessions, i.e., S4 in Session A vs. S4 in Session B and so forth, revealed that 195 out of 200 cases were 'assimilative', the durations becoming more like the preceding speaker's.

Figure 3 presents a typical example of sentence and interval durations obtained in the two experimental sessions. Each box-plot is based on ten utterance samples. The sentence and interval duration of the 'preceding speaker' are shown in square brackets. For both sentences and intervals, the 'post-takeover' duration (boxplots for Session B) settles somewhere between the subject's 'inherent' duration (boxplots for Session A) and the preceding speaker's 'inherent' duration (boxplots in square brackets). If the subjects' speech had not been affected by the preceding speaker's speech, the sentence and interval duration obtained in Session B would have been similar to what was obtained in Session A. There were other possibilities, too, such as a complete assimilation to the preceding speaker's durational values, or an overshoot in the subjects' post-takeover duration as a result of over-assimilation, or even dissimilation as a result of hyper-correction in the adjustment of duration. It is extremely interesting that all the subjects' post-takeover duration became closer to the preceding speaker's values but not quite to the extent of completely assimilating to the preceding speaker. This fact leads one to hypothesize that when the subjects were exposed to the preceding speaker's speech, they reorganized their speech according to two competing criteria, their own temporal criterion and the preceding speaker's temporal criterion. Inter-speaker interaction, then,

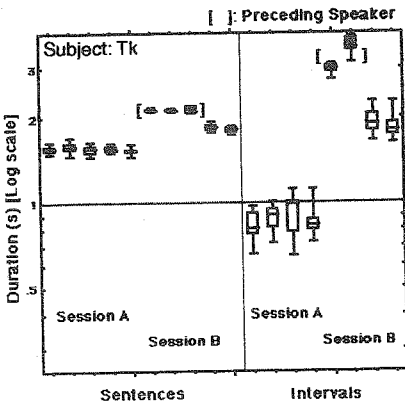


Figure 3. A typical example of sentence and interval durations obtained in the two experimental sessions. Each box-plot is based on ten utterance samples.

may tentatively be defined as a reorganization of one's speech with reference to one's own speech production framework, on one hand, and to the speech production framework of another speaker, on the other.

'Approximation' vs. 'Deviation'

Based on the above discussion, we can look at the results of interaction from two different viewpoints. One is to see how close the subjects' durational values came to the preceding speaker's value. I will call this 'approximation', and will define it as the difference between the preceding speaker's 'inherent' duration and the subjects' 'post-takeover' duration. The other way of looking at the result of interaction is to see how much one's 'post-takeover' value deviated from one's 'inherent' value. I will call this 'deviation', and will define it as the difference between the subject's 'inherent' duration and the subjects' 'post-takeover' duration.

In Figures 4(a) and 4(b), 'approximation' and 'deviation' are plotted against 'inherent difference', the difference between the 'inherent' durations of the preceding speaker and the subject. The filled data points pertain to sentences, while the unfilled data points pertain to intervals. The dotted line in Figure 4(a) denotes 'full approximation', while the dotted line in Figure 4(b) denotes 'no deviation'.

In Figure 4(a), the plots for both sentences and intervals exhibit fairly clear positive correlation. That is, the smaller the inherent difference between the preceding speaker and the subject, the better the approximation, and vice versa. The correlation was high for both sentences and intervals, the correlation coefficient value (r) being 0.699 for sentences and 0.702 for intervals. [Both significant at the 0.01 level] In Figure 4(b), the plots for sentences show positive correlation (r : 0.593, significant at the 0.01 level) but the plots for intervals reveal a very poor correlation (r : 0.044). In other words, for sentences, the greater the 'inherent difference', the greater the deviation from one's own durational value, and vice versa. However, for intervals, the degree of 'inherent difference' was in no way correlated to the degree of deviation.

We can reexpress the above observations in the following manner: Sentences are characterized by fairly

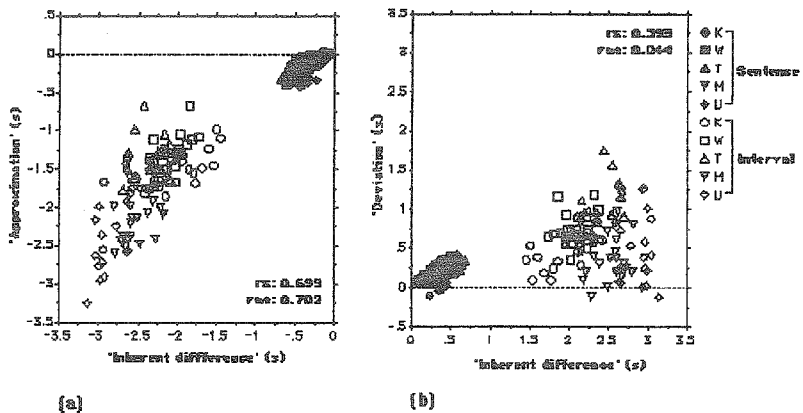


Figure 4. (a) 'Approximation' (the difference between the preceding speaker's 'inherent' duration and the subject's 'post-takeover' duration (s)) plotted against 'inherent difference' (the difference between the 'inherent' durations of the preceding speaker and the subject (s))

(b) 'Deviation' (the difference between the subject's 'post-takeover' duration and the 'inherent' duration (s)) plotted against 'inherent difference' (the difference between the 'inherent' durations of the preceding speaker and the subject (s)) [The 'inherent' duration denotes the duration of Int3, S4, Int4, or S5 in Session A, while the 'post-takeover' duration denotes the duration of Int3, S4, Int4, or S5 in Session B. See Fig. 1.]

good correlation for both 'approximation' and 'deviation', whereas intervals are characterized by fairly good correlation for 'approximation' but distinctly poor correlation for 'deviation'. If we assume 'approximation' and 'deviation' to be the two principal factors governing the degree of speaker interaction in speech production, the results can be interpreted to suggest that, for sentences, both factors exert about the same amount of influence in determining the degree of speaker interaction, while, for intervals, 'approximation' plays a more dominant role compared with 'deviation'.

In the previous section, 'approximation' was described as 'how close the subjects' durational values came to the preceding speaker's value', while 'deviation' was described as 'how much one's 'post-takeover' value deviated from one's 'inherent' value. This implies that, in regard to 'approximation', the criterion for the reorganization of speech is 'the other(preceding) speaker's temporal framework', while in regard to 'deviation', it is 'one's own temporal framework'. In other words, the production of sentences are governed both by one's own temporal criterion and the preceding speaker's criterion, whereas the production of intervals are governed primarily by the preceding speaker's criterion, and that one's own criterion does not play a governing role in determining the degree of interaction in interval production.

SUMMARY AND CONCLUSIONS

The findings of the present study can be summarized as follows:

Both the sentence duration and the interval duration deviated from one's 'inherent' duration when exposed to another speaker's speech. The deviation was mostly assimilative, or the subjects' sentence and interval duration became more like the preceding speaker's sentence and interval duration.

The comparison of the correlations between the 'inherent distance' (the difference between the inherent durations of the preceding speaker and the subject), on one hand, and 'approximation' (the difference between the subject's inherent and post-takeover durations) and 'deviation' (the difference between the subject's post-takeover duration and inherent duration), on the other, suggested that, with regard to sentences, both 'approximation' and 'deviation' have about the same amount of influence in determining the degree of speaker interaction, while with regard to intervals, 'approximation' seems to be the stronger determinant of speaker interaction compared with 'deviation'. In other words, the results suggested that the production of sentences are governed both by 'one's own temporal criterion' and 'the preceding speaker's criterion', whereas the production of intervals are governed primarily by 'the preceding speaker's criterion'.

ACKNOWLEDGMENTS

This work was partly supported by the Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science and Culture, Japan (No. 06232213).

REFERENCES

- Feldstein, S & Welkowitz, J. (1978) *A chronography of conversation: in defense of an objective approach*, Nonverbal behavior and communication, edited by Siegman, A. W. & Feldstein, S., (Erlbaum: Hillsdale, NJ) 329-378.
- Guitar, B. et. al. (1992) *Parent Verbal Interactions and Speech Rate: A Case Study in Stuttering*, Journal of Speech and Hearing Research, 35, 742-754.
- Jaffe, J & Feldstein, S (1970) *Rhythms of dialogue*, (Academic Press: New York).
- Jasnow, M. & Feldstein, S. (1986) *Adult-like Temporal Characteristics of Mother-Infant Vocal Interactions*, Child Development, 57, 754-761.
- Matarazzo, J. D. et. al., (1963) *Interviewer Influence on Durations of Interviewee Speech*, Journal of Verbal Learning and Verbal Behavior, 1, 451-458.