

Paralinguistic Effects on Turn-Taking Behavior in Expressive Conversation

Hiroki Mori and Hideki Kasuya

Graduate School of Engineering, Utsunomiya University, Japan

hiroki@speech-lab.org

Abstract

Speaker and paralinguistic properties of dialogue speech that affect the timing at turn changes are investigated by analyzing the UU Database for Paralinguistic Information Studies. The results showed a large variation among speakers and a strong interaction with partner in pause/overlap duration. In addition, perceived emotional states of utterances had significant effects on the pause/overlap duration.

Index Terms: turn-taking, timing, emotional states

1. Introduction

Turn-taking in everyday conversation with familiar people is quite a complex phenomenon. We often predict transition relevance places (TRPs)[1] of another speaker's turn and start to speak without waiting. Sometimes we even do not mind TRPs and interrupt others, not only for arbitrarily taking up the turn but also for collaboratively completing the partner's speech fragment. Conversely, we often hesitate to speak when we are thinking, or when we are not so sure about what we are to speak. Therefore, it is very likely that temporal structure of turn-taking in conversation reflects emotional states of speakers.

In this paper, the effects of (i) speaker, and (ii) perceived emotional states of speakers, on the speech timing at turn changes are investigated, adopting the Utsunomiya University Spoken Dialogue Database for Paralinguistic Information Studies (UUDB) [2] as material of expressive conversational speech.

2. Method

In the UUDB, an utterance boundary is defined as the union of silence (> 400 ms) and *slash unit* (related to TRP) boundary. Each utterance is assigned a six-dimension vector that represents the perceived emotional state of the speaker. (For its reliability issues, see [2]). The dimensions are pleasant-unpleasant, aroused-sleepy, dominant-submissive, credible-doubtful, interested-indifferent, and positive-negative, with a value from 1 to 7. In order to analyze the structure of turn-taking, we further annotated for each utterance its *preceding* one. Consequently, every utterance in the UUDB (of 12 females and 2 males) was classified into the categories described below, which resulted in the number shown in Fig. 1.

Hold Preceding utterance is the current speaker's one.

Change Preceding utterance is the another speaker's one, and terminates at a TRP.

Interruption Preceding utterance is the another speaker's one, and either terminates at a non-TRP or continues.

Back-channel Current utterance is a back-channel.

3. Results

Because the main interest of this paper is the timing at turn changes, we focus on the **Change** utterances in this section.

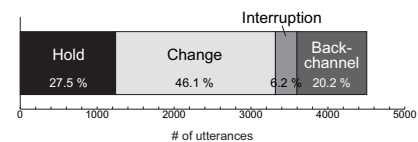


Figure 1: Categorization of utterances.

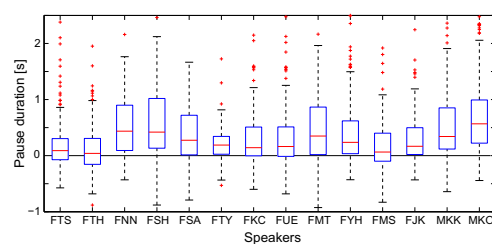


Figure 2: Pause/overlap duration distribution.

Table 1: Mean pause durations. The dataset was optimally split for four significant emotion dimensions.

Pleasant (≥ 4.33)	0.313 s	Dominant (≥ 3.33)	0.441 s	Credible (≥ 4.00)	0.354 s	Positive (≥ 4.33)	0.331 s
Unpleasant (≤ 4.00)	0.457 s	Submissive (≤ 3.00)	0.246 s	Doubtful (≤ 3.67)	0.558 s	Negative (≤ 4.00)	0.551 s

Figure 2 shows the duration distribution of pause between current and preceding utterances. Negative duration corresponds to an overlapping interval. There existed a remarkable variation among speakers. Interestingly, mean pause durations of every speaker pair were very similar ($r = 0.98$).

An ANOVA revealed significant ($p < 0.01$) effects of four emotion dimensions among six on mean pause duration, as shown in Table 1. These results can be interpreted as "pleasant, submissive, credible, or positive utterances are spoken early."

4. Conclusions

Timing properties at turn changes in expressive conversation were investigated. Although the ranges of pause duration varied from speaker to speaker, there existed a strong interaction with partner. It was also revealed that the pause duration was affected by perceived pleasantness, dominance, credibility and positivity of speakers. These observations may be applied to improving response timing control for spoken dialogue systems with capability of expressing paralinguistic information.

5. References

- [1] Sacks, H., Schegloff, E. A. and Jefferson, G., "A simplest systematics for the organization of turn-taking for conversation," *Language*, 50(4):696–735, 1974.
- [2] Mori, H., Satake, T., Nakamura, M. and Kasuya, H., "UU Database: A spoken dialogue corpus for studies on paralinguistic information in expressive conversation," TSD 2008, Brno, Czech Republic, Sept. 8–12, 2008 (to appear).