The 6th International Workshop on Emergent Synthesis-IWES '06
August 18-19, 2006, The University of Tokyo, Japan

Relationship between Utterance Dynamics and Subjective Pragmatics in the Conversation of Consensus Building Process

Makoto Yoshida and Yoshihiro Miyake

Interdisciplinary Graduate School of Science and Engineering
Tokyo Institute of Technology
Yokohama, Kanagawa 226-8502, Japan
yoshida@myk.dis.titech.ac.jp

ABSTRACT

We measured the turn-taking process in a dialogue for building consensus between two subjects. Additionally, the temporal development of cycle and response time of utterance was analyzed and we investigated their conversational dynamics. As a result, temporal development of turn-taking and its typical development in the consensus building process were clarified.

Key words: conversation, dialogue, consensus, dynamics, utterance, response time, pragmatics, semantics

1. INTRODUCTION

Communication among human is co-creative and we can sympathize together. In order to support co-creative communication, technology in this field is becoming to be needed to establish the flexible connection between robots and human, especially in this society, where telecommunication is commonly used. But co-creative principles have not been clarified enough. And thus, we aim to further study in this field. To make such a kind of robot-human system, principles of co-creative communication in human-human system should be investigated.

Temporal development of a dialogue between two persons is a phenomenon of co-creative communication of human-human system. Former studies of human communication have explored conversational differences due to social relationships, intimateness and physical position of two speakers from the viewpoint of linguistics. Studies have also been conducted on how temporal dynamics differ from the viewpoint of kinematics.

However the research which analyzes the relationship of the two, that is to say, the relationship of the linguistic pragmatics of the dialogue and kinematic dynamics of it has not been conducted yet. From this background, this research is an experiment connecting pragmatics-side and kinematics-side in dialogue-communication of two persons.

To date, research between two persons in the field of linguistic dialogue analysis, including research about the pragmatics of a dialogue between persons who meet for the first time has been done by Usami. Besides, speech level shift had been studied by Ikuta. In this field, discoveries have been made relating to how social relationships are reflected to structures of a dialogue; such findings have included the differences in the temporal change of politeness and frequency of inserting topics, due to the relative power of the two persons socially.

On the other hand, in the field of kinematic dialogue analysis, there are researches about dialogue dynamics by Ikegami. In these studies, it is clarified that transition of dynamics of a dialogue is shown for each topic. Researches related to turn-taking have been conducted by Ohsuga and Sato. Ohsuga investigated the relationship between turn-taking and prosodic features. While Sato made use of turn-taking to make a dialogue system between robots and human. Besides it is clarified by Hirschberg that variation and length of pitch in a phrase are related to termination and continuation of utterance. In addition, methods to detect utterance are showed by Takeuchi.

As mentioned in the above, most these researches until now have focused on the transition of pragmatics of a dialogue or on the kinematic point of view. So it is considered that studies, which connect these two fields of research, have not been conducted as yet. Thus in this research, we aimed to clarify the relationship between these two sides, using the turn-taking between two subjects. Specifically, the purpose of this research is to clarify the relationship between utterance dynamics and pragmatics

calculating length of utterance and response time between utterances using the intensity of utterance between two subjects under consensus building process.

In this report, methods of experiments are explained in chapter 2, the results are illustrated in chapter 3, and we discuss the results in chapter 4.

2. METHODS

2.1 Consensus building task

An experiment named 'The Consensus Building Task' was as follows. We asked two subjects to make one speculation about something unknown for them using given data in several minutes. Only two subjects were in a room. The same material concerning to the speculation was handed to subjects respectively. In addition different material was also handed respectively. They could tell everything by conversation. But it was restricted to show another person their own data and to use stuff that was not given like a pen, a notebook, a calculator and so on. They were asked to speak as much as they could about what they were considering.

There had to be correct answer for a topic. The topic and its material had been given to the subjects just before an experiment.

After the conversation, they told an experimenter one speculated answer. The amount of fee had been varied to their answer. After that they answered a questionnaire. In the questionnaire, they answered two questions. One was subjective evaluation of level of consensus distinguished 5 levels between two persons every 30 s. Another was the subjective time of consensus between them. This evaluation had been conducted in a room alone. When everything had been finished, fee was paid for their participation.

In this report, a task named 'Price Speculating Task' was used as a Consensus Building Task. In this task, subjects presented one price of something whose property was shown in given materials.

2.2 Procedure

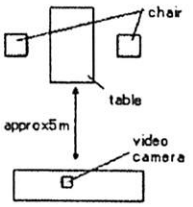
The subjects started this procedure of experiments only after having confirmed that they wouldn't have any trouble or external pressures that would affect them on the time. Once in the room, the procedure and rules of the task were explained to each subject. Just after this step, subjects had a chance to ask questions on the series of procedures. Moreover, they were asked to inactivate their mobile phones. During the experiments, drinking and eating were forbidden. We recorded their behavior as well as their voice as stereo sound. This procedure was adapted to the whole experiment.

2.3 Subject

Two subjects in the consensus building task were regulated as follows. They already knew each other, same academic grade if they are students, same sex, same nationality, and could speak naturally about a range of issues without hesitation.

2.4 Experimental system

We had the subjects sit down in chairs with no arm rests, face-to-face, and with a 50 cm high table between them. It was silent, under normal illumination and comfortable temperature with the experiments. We used a digital video camera (DCR- PC300K, SONY) and a headset-type oriented microphone (MS=HS67BK, ELECOM) to record the behavior and voice of the subjects. These microphones had monaural output. So the voice of each subject was recorded into a video camera as one part of stereo sound. The location of a table, chairs and a video camera is as Fig.1 and 2.



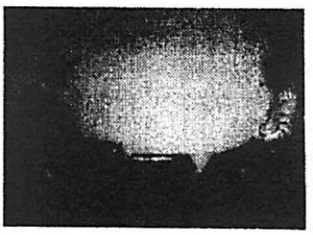


Fig. 1. Experimental layout **Fig. 2. Image of experiment**

2.5 Analysis method

In this experiment, a cycle of utterance was defined as time length between onset time of utterance of that person and next onset time of that person. Besides, a response time was defined as time length between utterances; offset time of utterance of one person and onset time of another person as Fig.3. Over 1.5 s pause was regarded as a response time.

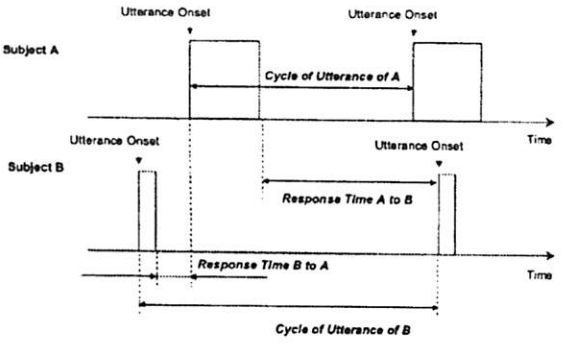


Fig. 3. Definition of cycle of utterance and response time

Audio and Visual data was inputted to PC as AVI files using an application entitled Movie Maker (Ver2.1.4026.0, Microsoft) from a video camera. After that, only sound data was picked up as WAVE files using an application entitled Virtual Dub (Ver1.6.11). In addition, we picked up data of each subject as a monaural sound from the stereo sound, using SP4WIN Custom (Ver.2.2b, NTT-AT). Additionally, we got temporal data of the intensity of each voice using honing window from raw data of the sound. We got 10 flames of temporal intensity data per second. After that, we obtained the length of cycle and response time of utterance to each person. And their graphs were described using an application entitled Excel 2003(Ver11.5612.5606, Microsoft). Sound rate had always been 32 kHz and 16 bit in all the experiments.

3. RESULT

This result was chosen as one of the typical result in Price Speculating task. This task was to speculate the correct rent of a room of an apartment. Material was about detail of another room in the apartment. Different materials were handed to two persons to speculate.

Fig.4 is raw-sound data recorded over 9.5 minute in the consensus building task. Fig.5 is intensity of sound data based on that raw-sound data. These both tiers illustrate comparisons between subject A and subject B. Fig.4 and 5 show that two subjects kept talking during the range of time. Horizontal axis shows temporal development. Figures 4 to 10 are adjusted temporally.

Fig.6 is moving average of cycle of utterance. This moving average was taken using 5 utterances. And onset times of 5 utterances were also averaged. Fig.7 is moving average of response time. This moving average was taken using 5 response times. And onset times of the 5 utterances were also averaged.

Fig.8 shows evaluations which had done by two subjects subjectively after the experiment. Subjects were asked to answer the extent of consensus between them using recorded video. Besides, vertical lines show the time of subjective consensus of both subjects. 263 and 563 s were chosen by subject A. And 546 s was chosen by subject B.

Fig.9 shows subjective evaluations which had done by 5 people who are not subjects. The method was parallel with that of evaluation by subjects.

Pragmatics over whole conversation is shown on Fig.10 Besides, semantics between 270 and 290 s, between 400 and 420 s, and between 540 and 570 s are shown in Fig.11, 12 and 13 separately.

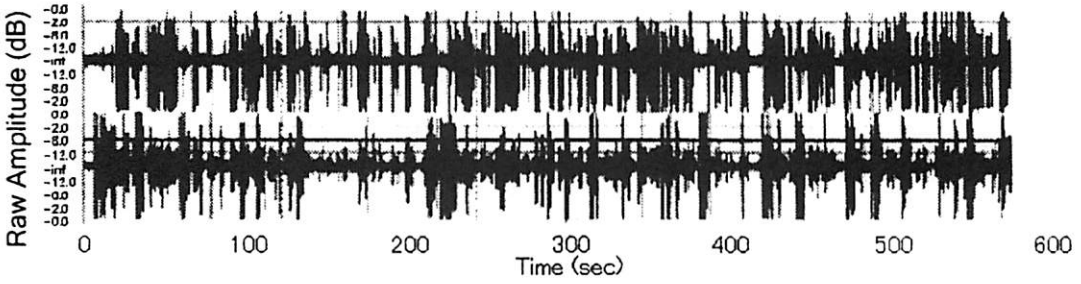


Fig.4 Raw data of amplitude (Upper tier: Subject A, Lower tier: Subject B)

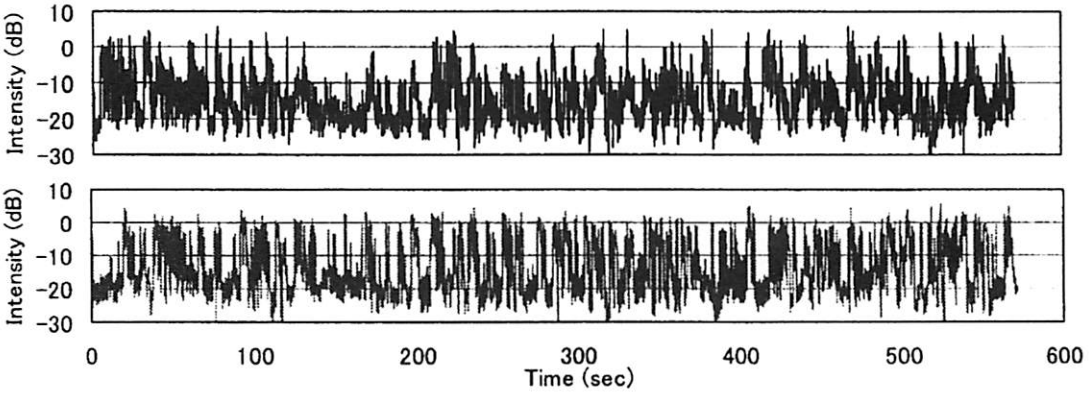


Fig.5 Intensity of utterance of subject (Upper tier: Subject A, Lower tier: Subject B)

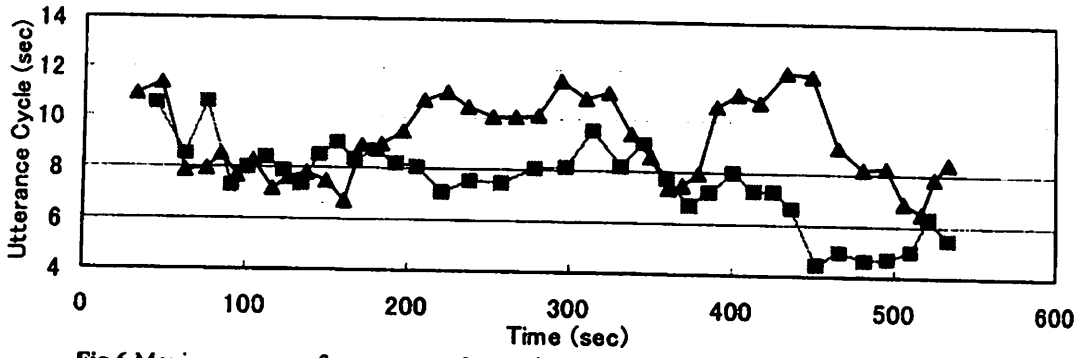


Fig.6 Moving average of utterance cycle (▲: Subject A, ■: Subject B)

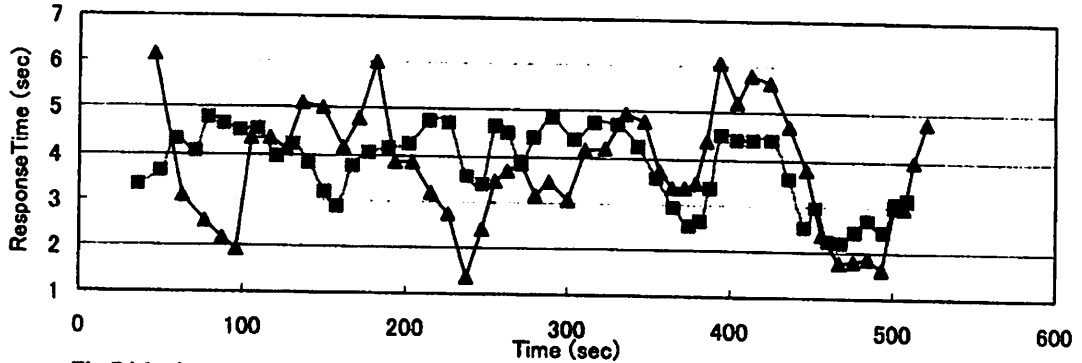


Fig.7 Moving average of response time (▲: Subject A, ■: Subject B)

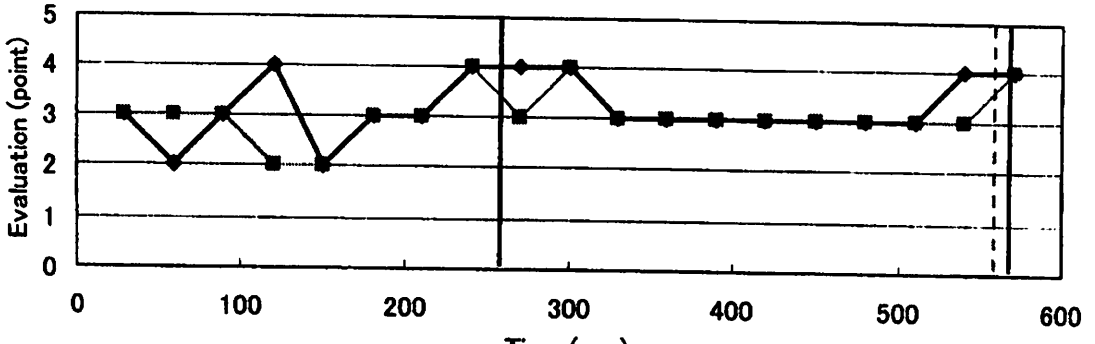


Fig.8 Subjective evaluations of extent of consensus by subjects own(▲: Subject A, ■: Subject B)
Vertical lines show subjective consensus time (Solid line: Subject A, Dotted line: Subject B)

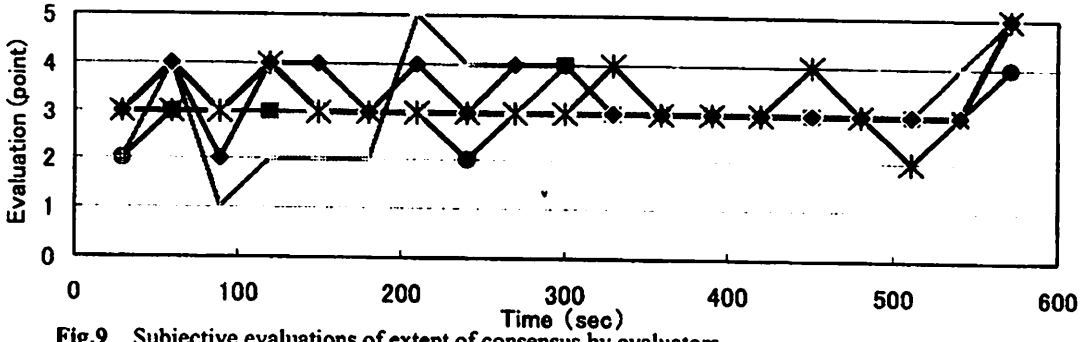


Fig.9 Subjective evaluations of extent of consensus by evaluators
(Each maker represents an evaluator separately)

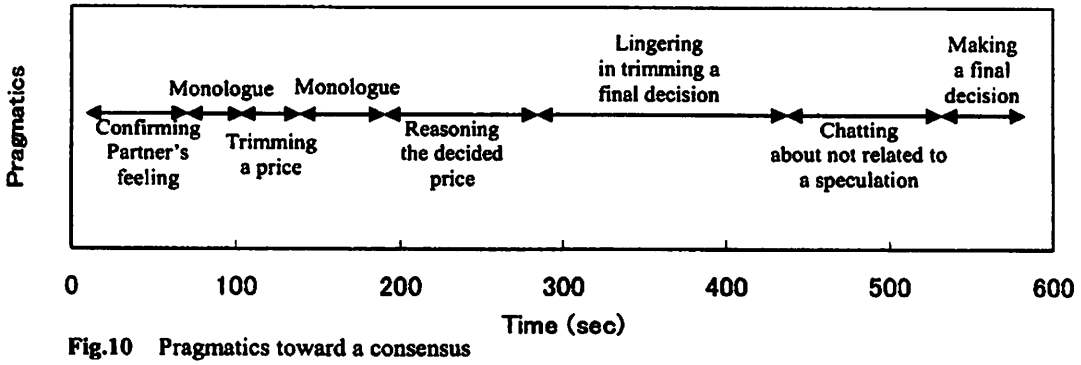


Fig.10 Pragmatics toward a consensus

A17
Sorekuraikana, nijyu san man kurai kana cyokkan wo shinjite hajimeno.
(Around that, around 230000yen, isn't that. Trusting an inspiration of beginning.)
B17
Hyakuen tan-i no bicyousei mo.
(Fine adjustment by 100 yen, too.)
A18
Hyakuen tan-i no bicyousei ka, fufufu, zenzen wakaranwa. Dotti no kanousei ga takaito-omou? Takaika,hikuika, Nijyu san man yori.
(Fine adjustment by 100yen, hum, I can't suppose it at all. Do you think which possibility is higher? Higher or lower than 230000 yen.)
B18
Ah, nanka agarisouna.
(Well, I think it become to be higher.)

Fig.11 Semantics between 270 and 290 s (Translation in English)

A26
Fu-n, doushimasuka? Saigono, nijyu san man.
(Let's me see, how can we decide?)
B26
Uh-n, muzukashi.
(Oh, this is difficult.)
A27
Sotti ikurayattakke?
(How much price have you said?)
B 27
Jyukyu man Yon sen roppyaku en.
(194600 yen.)

Fig.12 Semantics between 400 and 420 s (Translation in English)

A32
Nijyuu san man ni kyoumiga wakuna. Ohra kun ha don-na kanji?
(I'm attracted to 230000 yen. How do you feel now ?)
B32
Boku ha nijyusan man issen gohyakuen kanato. Konkyomonaku.
(I suppose it's 232500 yen. Without a reason.)
A33
Ah,haha, sorekuraide ochituitannyattara Sorekurainishitemiyouka? Yoshi,soredekou. Fufu
(Well, being concluded around that price, how about deciding that price? OK,done. Ha-ha.)
B33
Fufu
(Ha-ha.)

Fig.13 Semantics between 540 and 570 s (Translation in English)

4. DISCUSSION

In this experiment, we measured dynamics of utterance and pragmatics of utterance simultaneously.

About temporal development of dynamics of utterance, it seems that there are cyclic variations and trends to be synchronized each other. For example, utterance cycle of subject A in Fig.6 seems to be cyclic repeatedly continued like it decreased until around 150 s, increased until around 300 s, decreased until 360 s. About utterance cycle of subject B, cycle also repeatedly continued like it decreased until around 220 s, increased until around 320 s and decreased until 450 s. Response time of subject A in Fig.7 is also seems to be cyclic repeatedly continued like it decreased until around 100 s, increased until around 180 s, decreased until 240 s. About response time of subject B, it became to be cyclic gradually especially after 320 s. In addition, response time of both subject A and B in Fig.7 seems to be synchronized each other after around 330 s.

About temporal development of pragmatics of utterance, it seems that stages of exchanging opinions positively and stages of considering matters were repeated. In Fig.10, stages they said monologues or they lingered to speak and relatively positive stages were shown alternately. Including considering about subjective evaluations in Fig.8, subjective consensus time points are in relatively positive stage in Fig.8. They correspond to relatively positive stages in Fig.10.

And vertical lines around 560 s in Fig.8 show they had reached their subjective consensus. The timing of consensus can be also confirmed in Fig.13. Besides, Fig.9 shows that evaluators also gave high point around that time.

It also seems to be there are relationships between temporal development of dynamics and that of pragmatics. It means there are relationships between sympathy in pragmatics and synchronization in dynamics toward a final consensus.

For example, around 280 s, response times between two people become to be similar in dynamics as Fig.7 shows. While they began to trim the price based on round consensus that they had made in semantics as Fig.11 shows. Besides, subjective evaluation was relatively high as Fig.8 and 9 show.

As another representative example, around 400 s, both utterance cycle and response time were increased in dynamics as Fig.6 and 7 show. While in pragmatics, this was a situation that two subjects were at a loss to proceed to make a speculation as Fig.12 shows. Besides, subjective evaluation was not relatively high as Fig.8 and 9 show.

Around 550 s, response time became to be increased in Fig.7. And the final consensus was recognized in the temporal subjective evaluation and subjective consensus time points in Fig.8.

Considering these correspondences between dynamics and pragmatics toward a final speculation, it seems that there are some correspondent relationships between them. We have found these properties in other examples of same kind experiments.

We will try to find relationships between pragmatics and dynamics in detail using other ways to measure them and classifying types of conversations. These principles in this report will help systems to support technology of robot-human communication.

REFERENCES

Hayashi, M. (2001). Postposition-initiated utterances in Japanese conversation. *Studies in Interactional Linguistics*, John Benjamins, Amsterdam and Philadelphia, pp.317-343.

Hayashi, M. (2003). *Joint utterance construction in Japanese conversation*, John Bejamins, Amsterdam and Philadelphia.

Hayashi, M. (2004). Discourse within a sentence: An exploration of postpositions in Japanese as an interactional resource. *Language in Society*, Vol.33, pp.343-376.

Ikuta, S. (1983). Speech level shift and conversation strategy in Japanese discourse. *Language Science*, Vol.5, No.1, pp.37-53.

Iizuka, H., Ikegami T. (2003). Adaptive Coupling and Intersubjectivity in Simulated Turn-Taking Behaviours. *Proceedings of ECAL 03*, Advances in Artificial Life, Lecture Notes in Artificial Intelligence, pp.336-345.

Iizuka, H., Ikegami T. (2004). Adaptability and diversity in simulated turn-taking behavior. *Artificial Life*, Vol. 10, Issue 4, pp.361-378.

Hirschberg, J. (1983). Communication and prosody: functional aspects of prosody. *Speech Communication*, Vol.36, pp.31-43.

Ohsuga, T., Nishida, M., Horiuchi, Y., and Ichikawa, A. (2005). Investigation of the Relationship between Turn-taking and Prosodic Features in Spontaneous Dialogue. *Proceedings of the 9th European Conference on Speech Communication and Technology (EUROSPEECH)*, pp.33-36.

Sato, R., Higashinaka, R., Tamoto, M., Nakano, M., Aikawa, K. (2002). Learning decision trees to determine turn-taking by spoken dialogue systems. *Proceedings of ICSLP-98*, pp.487-490.

Takeuchi, M., Kitaoka, N. and Nakagawa S. (2004). Timing Detection for Realtime Dialog Systems Using Prosodic and Linguistic Information. *International Conference: Speech Prosody 2004*, pp.529-532.

Usami, M. (2002). Discourse Politeness in Japanese Conversation: Some Implications for a Universal Theory of Politeness, Hituzi Syobo, Tokyo.