# 1B1-4 合意形成過程における会話の意味的側面とダイナミクスの関係

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

○学 吉田 誠 (東京工業大学) 正 三宅美博 (東京工業大学)

Makoto Yoshida, Tokyo Institute of Technology, Midori ,Yokohama, Kanagawa 226-8502, Japan
Yoshihiro Miyake, Tokyo Institute of Technology,

**Abstract**: We studied process of information integration thorough observing dynamics of turn-taking in dialogue for building consensus between two persons. We analyzed this phenomenon based on kinematical viewpoint and linguistical viewpoint. From kinematical viewpoint, the temporal development of response time of utterance was analyzed. On the other hand, from linguistical viewpoint, temporal development of pragmatics regarding their consensus was analyzed. As a result, their typical developments were clarified; Response time is oscillating temporally. And synchronization of response time between two persons was observed. Besides timing of this synchronization corresponds to timing of getting to consensus. This result indicates that rhythmic phenomenon in dialogue has an important roll in this kind of information integration.

**Key Words**: information integration, conversation, dialogue, consensus, dynamics, utterance, response time, pragmatics, semantics

## 1. Introduction

Dynamics, especially rhythm and synchronization, is one of the important viewpoints regarding to information integration. Principle of information integration had been studied in the brain science by Singer[1] and Varela[2]. Singer had clarified that oscillation in the visual cortex of cat establishes recognition of visual pattern. Besides, Varela had suggested that a kind of desynchronization of neural ensemble is necessary to change of cognitive state. These studies suggested rhythm and synchronization contribute to integrate information.

We picked up communication among human as one of such information integration, and we studied conversational dynamics. To date, there are some studies in linguistical field and kinematical field.

In the field of linguistic dialogue analysis, research about the pragmatics of a dialogue between persons who meet for the first time has been done by Usami[3]. Besides, speech level shift had been studied by Ikuta[4]. 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[5] [6]. 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[7] and Sato[8]. Ohsuga investigated the relationship between turn-taking and prosodic features. Sato made use of turn-taking to make a dialogue system between robots and human. Besides it is clarified by Hirschberg[9] 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[10].

However the research which analyzes the relationship of the linguistic pragmatics of dialogue and kinematic dynamics of it has not been conducted yet. From this background, in this research, we aimed to clarify principles of information integration through connecting pragmatics-side and kinematics-side in dialogue-communication of two persons. We analyzed temporal development of response time as kinematical side, at the same time, temporal development of pragmatics relating to consensus was analyzed as linguistical side.

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 timing 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. In this report, this task was to speculate the correct rent of a room of an apartment through dialogue. Handed material was about detail of another room in the same kind of apartment. Different materials were handed to two persons to speculate respectively.

## 2.2 Procedure

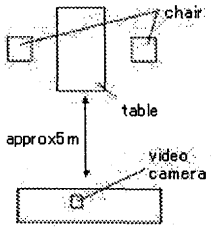
The subjects started this procedure of experiments only after having confirmed that they wouldn't have any trouble or external pressures that would affect them on the time. Once in the room, the procedure and rules of the task were explained to each subject. Just after this step, subjects had a chance 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.



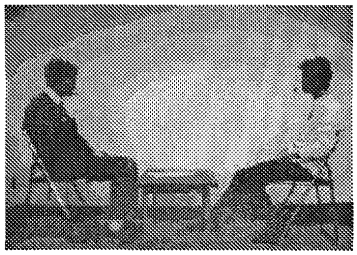



**Fig. 1 Experimental layout Fig. 2 Image of experiment**

## 2.5 Analysis method

In this experiment, a 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.

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 utterance and length of response time to each person referring to raw data of amplitude or intensity of utterance as Fin.4 and 5 indicate. And graphs about response time were described using an application entitled Excel 2003 (Ver11.5612.5606, Microsoft). These calculating processes can be done automatically using Excel to some extent, so we can obtain result roughly. But data in this report was done by manual work for the accuracy. To obtain auto correlation and cross correlation from utterance length data, MATLAB (Ver.7.04.365, The Math works) was used. Sound rate had always been 32 kHz and 16 bit in all the experiments.

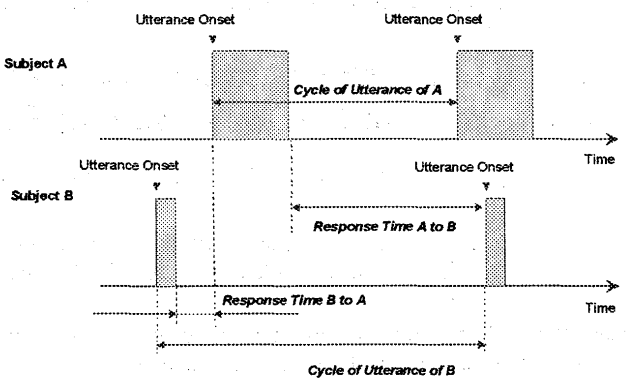



**Fig. 3. Definition of response time**

## 3. RESULT

This series of results was chosen as typical set of result in the price speculating task.

Figure 4 indicates raw-sound data of amplitude recorded over 7 minute in the consensus building task. Fig.5 indicates 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.

Figure 6 indicates response time. Horizontal axis shows temporal development. And vertical axis shows response time. Solid line indicates response time of subject A to B. Dotted line indicates that of subject B to A.

Figure 7 indicates response time averaged every 10 s. The way to look markers and axis is parallel with those of Fig.6.

Figure 8 indicates auto correlation of averaged response time subject A to B based on the values indicated in Fig.7. Horizontal axis shows lag time. And vertical axis shows correlation value. This auto correlation was taken every 10 s as a lag. Fig.9 also indicates auto correlation of averaged response time of subject B to A.

Figure 10 indicates cross correlation of averaged response time between subject A to B and B to A based on the values indicated in Fig.7. The way to look markers and axis is parallel with those of Fig.8 to 9. Fig.11 also indicates cross correlation of averaged response time between subject B to A and A to B.

Figure 12 indicates 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. 432 s was chosen by subject A. And 323 s, 695 s and 428 s were chosen by subject B.

Figure 13 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 indicated on Fig.14. Pragmatics in Fig.14 was expressed based on semantics, for example, as follows between 270 and 280 s, between 295 and 310 s, and between 420 and 435 s described in Fig.15, 16 and 17 separately.

Figure 15 indicates semantics when they are at a loss to proceed in their conversation. Fig.16 indicates semantics around near consensus in one part of conversation. Besides, Fig.17 indicates the semantics near a final consensus.

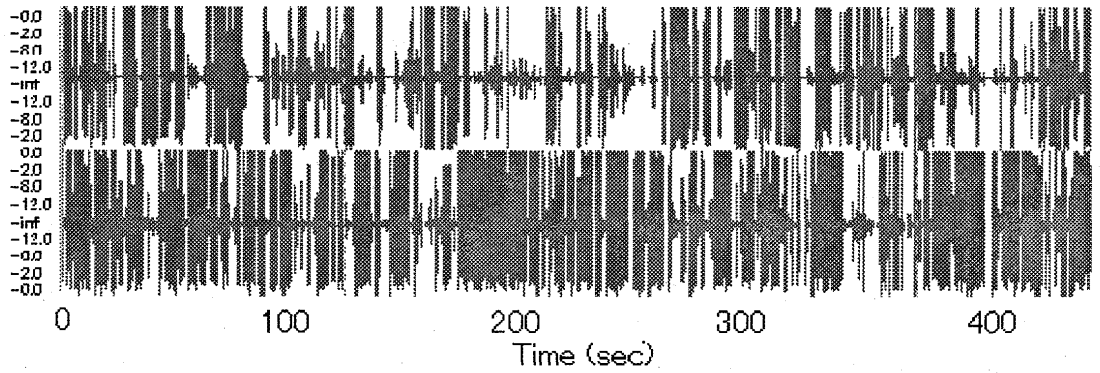



Fig.4 Raw data of amplitude (Upper tier: Subject A, Lower tier: Subject B)

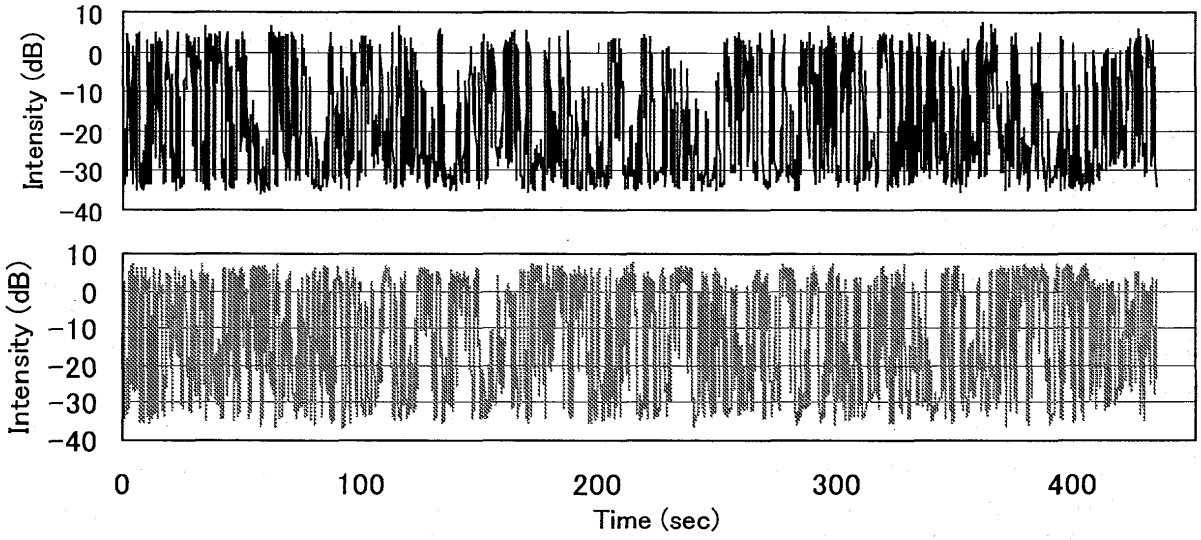



Fig.5 Intensity of utterance (Upper tier: Subject A, Lower tier: Subject B)

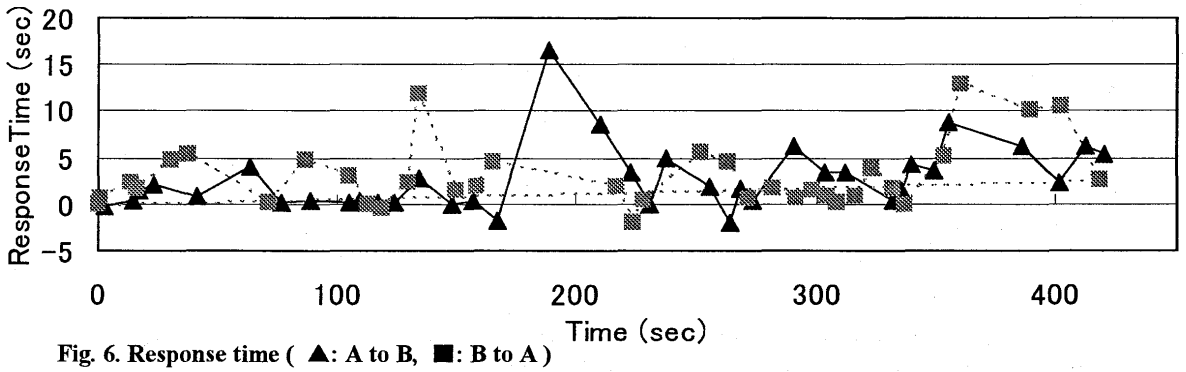



Fig. 6. Response time ( ▲: A to B, ■: B to A )

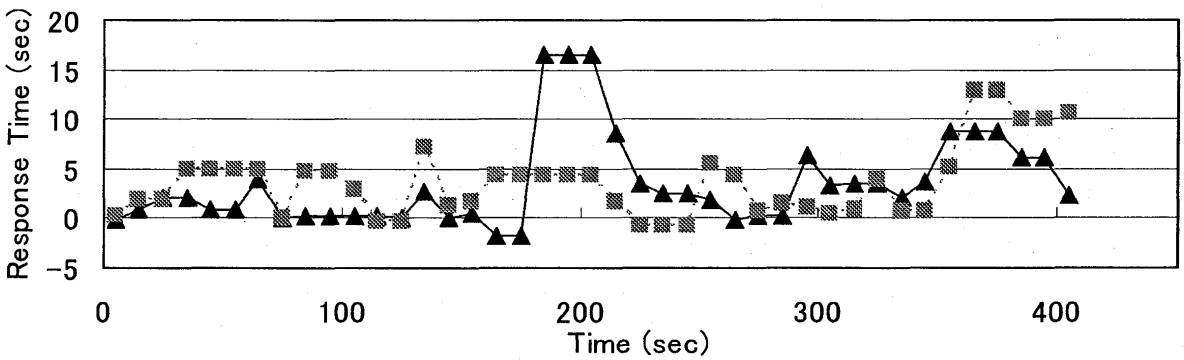



Fig. 7. Response time averaged every 10 s ( ▲: A to B, ■: B to A )

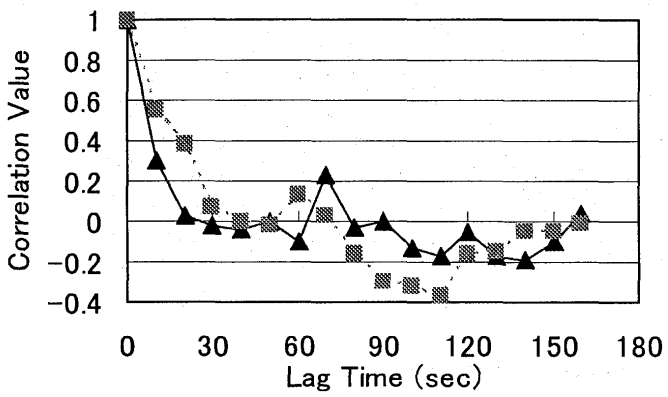



Fig.8 Auto Correlation of Response Time A
( ▲: Beginning, ■: Ending )

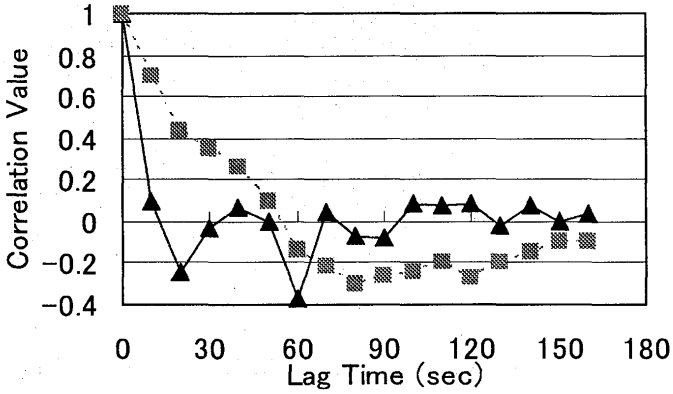



Fig.9 Auto Correlation of Response Time B
( ▲: Beginning, ■: Ending )

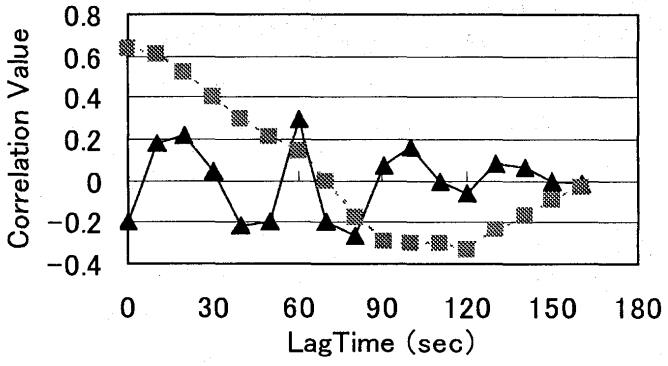



Fig.10 Cross Correlation of Response Time
A to B and B to A ( ▲: Beginning, ■: Ending )

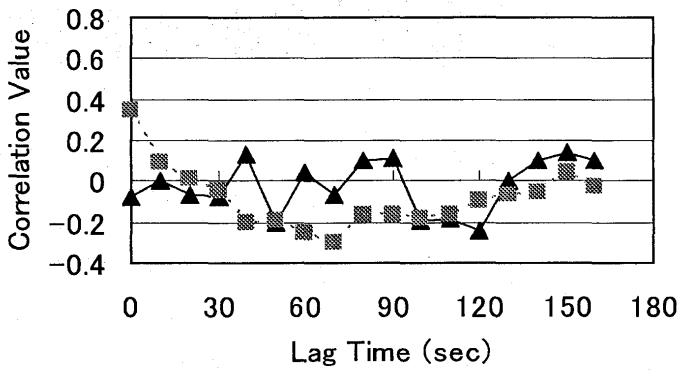



Fig.11 Cross Correlation of Response Time
B to A and A to B ( ▲: Beginning, ■: Ending )

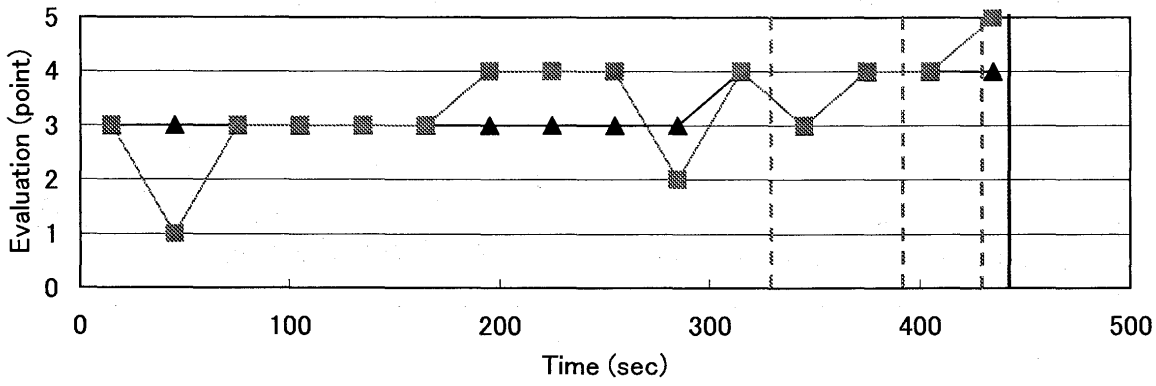



Fig.12 Subjective evaluations of extent of consensus by subjects own( ▲: Subject A, ■: Subject B )
Vertical lines show subjective consensus time (Solid line: Subject A, Dotted line: Subject B)

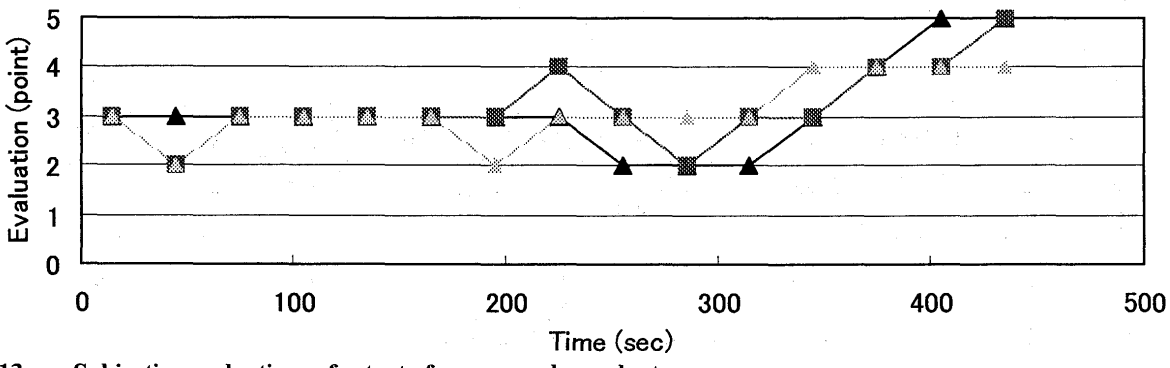



Fig.13 Subjective evaluations of extent of consensus by evaluators
(Each marker represents an evaluator separately)

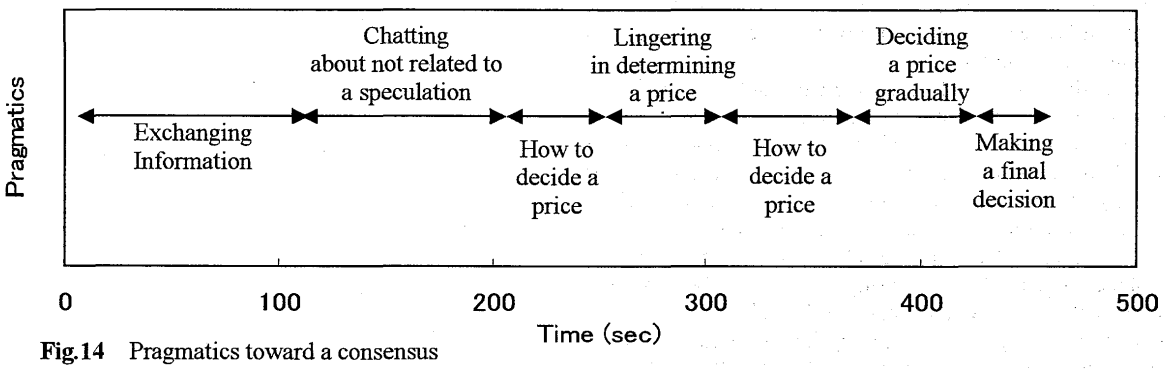



**Fig.14** Pragmatics toward a consensus

**A26**
Doh shiyoukanah.
(How can we decide? )

**B26**
Sono mensekino sa-...
(Well, that amount of surface, well...)

**A27**
Ah, ima itteta nedannyori takai hazu nanda.
(Ah, it must be more expensive than that price, right?)

**B 27**
Ettu cyottomatte, ima...
(Oh, wait wait, now...)

**Fig.15** Semantics between 270 and 280 s (Translation in English)

**B30**
Honto ha tabun motto sa ga
(In fact, maybe bigger difference is .)

**A31**
Sa ga. Aruhazu
(There must be difference.)

**B31**
Sa ga aruhazu nandayone...
Nandayone...
(There is difference, right.)

**A32**
Hazu nandayone...
(Must be, right.)_

**Fig.16** Semantics between 295 and 310 s (Translation in English)

**B70**
Jyuuniman kyuusen nanahyaku en.
(129700 yen)

**A71**
Jyuusannmann kirukannji?
(Less than 130000 yen?)

**B71**
Iikannjidayo.
(Sounds good.)

**A72**
Aridayone
(It is very possible.)

**B72**
Jya jyuuniman kyu-sen nanahyakuen de.
(So decided on 129700 yen.)

**A73**
De ikou.
(It's done.)

**Fig.17** Semantics between 420 and 435 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 oscillated variations. For example, response time of subject B to A in Fig.7 seems to be oscillated repeatedly, like it increased around 50 s, around 90 s, and 140 s. About response time of subject B to A, it also repeatedly oscillated like it increased around 200 s, around 300 s and around 370 s. We will confirm this oscillation statistically later, using auto correlation and cross correlation.

In addition, there are trends to be synchronized each other, for example, response time of both subject A to B and B to A in Fig.7 seems to be synchronized each other after around 280 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.14, stages they said they lingered to speak and relatively positive stages were shown alternately. Including considering about subjective evaluations in Fig.12, 13, subjective consensus timings indicated by vertical lines are in relatively positive evaluation in Fig.12, 13. They correspond to relatively positive stages in which positive exchange had conducted in Fig.14.

And vertical lines around 450 s in Fig.12 show they had reached their subjective consensus. The timing of consensus can be also confirmed in semantics described in Fig.17. Besides, Fig.13 shows these evaluators also gave high point around that timing. Through these subjective evaluations mentioned above, pragmatic process toward consensus was confirmed.

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.

In order to indicate this, we picked up two temporal parts of range of averaged response time. And dynamics of the two part was analyzed using auto correlation and cross correlation.

The two parts are the first 180 s and the last 180 s in the dialogue. These part were pragmatically different significantly in evaluations used the subjective evaluation values in Fig.12 and Fig.13 (T-test, n=6, $\alpha$ =0.05). About the last 180 s, we can say the period is at the state of around consensus pragmatically, it has

statistically high evaluation and different from beginning 180 s semantically as I mentioned above. So we tried to find also the kinematical difference or same properties between the two periods using correlation analysis.

About the ending part, Fig.8 and 9 indicates auto correlation of response time of the two subjects separately. Auto correlation represents extent of having cycle of variation of response time of each subject. In these figures, both beginning and ending part have some significant values around 0.3 at absolute value. So it indicates there were some oscillations in these dynamics.

Fig.10 and 11 indicates cross correlation of response time of the two subjects separately. Comparing between the begging part and ending part, there are recognized relatively high value of ending part in short lag time. In Fig.10, the correlation value on lag 0 is over 0.6 and in Fig.11, it is around 0.4. But those of in the beginning part in Fig.10 to 11, they are relatively low value rather than that of ending part. This means response time had become to be synchronized each other at the end part of dialogue.

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.

These results indicate that oscillating and synchronizing phenomena have important rolls in this kind of information integration. This principle can be applied to various fields.

## References


[1] Gray,CM, Konig,P, Engel,AK, Singer,W, . (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which resflects global stimulus properties. Nature338, pp334-337.

[2] Rodriguez,E, George,N, Lachaux,JP, Martinerie,J, Renaut,B, Varela,FJ. (1999). Perception's shadow: long-distance synchronization of human brain activity. Nature397, pp430-433.

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

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

[5] 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.

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

[7] 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.

[8] 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.

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

[10] 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.