

# What does the Face-Turning Action Imply in Consensus Building Communication?

Tetsuro Onishi, Takatsugu Hirayama, and Takashi Matsuyama

Graduate School of Informatics, Kyoto University  
Yoshida-honmachi, Sakyo-ku, Kyoto, Japan  
t-onishi,hirayama@vision.kuee.kyoto-u.ac.jp, tm@i.kyoto-u.ac.jp

**Abstract.** When talking with someone, we convey intention to each other by verbal and non-verbal behaviors. In consensus building dialogue, the participants need to understand whether they agree or disagree. They reiterate confirmation of partner’s internal state (agreement/disagreement) and reaction for it. In this study, we considered that the timing of listener’s reaction for the confirmation by a speaker reflects listener’s internal state, therefore analyzed the multimodal timing structures between the confirmation and the reaction by utterance and body motion. Especially, we focused on an action that the speaker turns his face toward the listener as the confirming action and analyzed how it influences the timing structures. As the results, we confirmed that the timing structures relate to the internal state and the relations are controlled by *face-turning* action.

## 1 Introduction

In recent years, many researchers have developed user support systems based on the human-machine interaction. We are currently designing an interactive system that supports the human-human task-oriented dialogue making a travel plan or selecting a gift. The system needs to estimate user’s internal state, such as intention, through his behaviors in order to serve sensible information or recommend desired goods. We human basically interact using speech in the dialogue. Speech is the revealed information, and it is consciously controlled. Therefore, speech can express intention falsely. It is hard for us and the system to estimate real internal state from speech accurately. In this study, we focus on non-verbal behaviors. Humans cannot control to express their internal state perfectly, and real internal state shows through non-verbal behaviors unconsciously [1].

Body motion, such as head rotation, facial expression, and gesture, and prosody of speech are non-verbal behaviors. We focus on a behavior that a speaker turns his face toward a listener (*face-turning* action) to confirm listener’s internal state in the task-oriented dialogue. What does the *face-turning* imply in the dialogue? Humans tend to turn their gaze on the dialogists to observe reaction [2, 3]. The listener’s reaction with body motion would reflect the internal state because the gazing by the speaker captures the listener’s body action. We also consider that the *face-turning* often induces the reaction (sometimes

Speaker A : I want to go to the Netherlands. How about you? : *Confirmation*  
 Listener B : Nice! I'm interested in *Eredivisie*. : *Reaction*

**Fig. 1.** An example of the exchange.

requests the reaction). It would be a powerful signal to elicit the internal state. We focus on a specific internal state, which is listener's agreement/disagreement with an intention expressed by the speaker. The state is the most primordial one built up in the task-oriented dialogue. The participants of the dialogue need to understand whether they agree or disagree, and then reach a consensus. They reiterate confirmation of agreement/disagreement and reaction for it, with occasional *face-turning*.

Temporal feature is also non-verbal media, which is a feature in the orthogonal dimension to the non-verbal behaviors. Some researchers have argued that the internal state is expressed in the timing of speech response more than in the prosody of speech [4, 5]. They, however, analyzed only speech and ignored influences of body motion which has important function in face-to-face communication, on the timing of response. We analyze the relations between the internal state and the timing structures based on both utterance and body motion, i.e. multimodal timing structures. The listener may unconsciously reflect his internal state to the timing of the reaction by utterance and body motion for the speaker's confirmation with the *face-turning*.

In this paper, we make some hypotheses about how the multimodal timing structures relate to respondent's internal state (agreement/disagreement), and test the hypotheses by analyzing video data and audio signals recorded in some consensus building dialogues. Especially, we focus on how the *face-turning* influences the timing structures, and discuss the meaning of the *face-turning* to confirm partner's intention in the consensus building communication.

## 2 Relations between the Timing of the Reaction for the Confirmation and the Respondent's Internal State

### 2.1 Dialogue Components of Consensus Building

Consensus building dialogue is a kind of task-oriented dialogue. The dialogue usually consists of some exchanges. In this study, we address an exchange composed of *Confirmation* and *Reaction* by speech and body action, which is a primordial framework to convey agreement/disagreement. *Confirmation* is an action that starts a new exchange, represents intention, and confirms agreement/disagreement of the partner. *Reaction* is an action that reacts to *Confirmation*. Fig.1 shows an example of the exchange.

Various body motions occur with utterance of *Confirmation* and *Reaction*. These are non-verbal behaviors that emphasize and supplement verbal behaviors. We focus on some body motions based on head rotation. *Face-turning* is one of

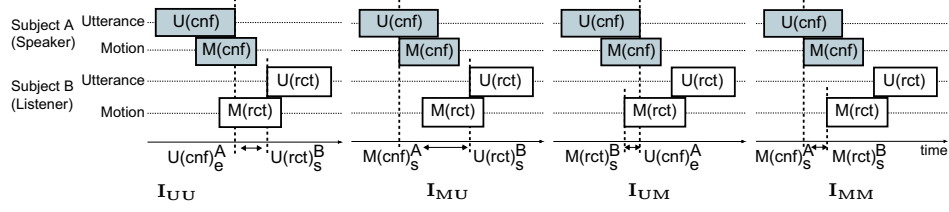


Fig. 2. Timing structures.

the typical body motions denoting *Confirmation* because its actor can observe partner's behaviors reflecting the internal state, and it often induces *Reaction*. Nodding, head-inclining, head-shaking, and *face-turning* are the typical body motions denoting *Reaction*. They can express agreement/disagreement.

## 2.2 Timing Structures between the Confirmation and the Reaction by Utterance and Body Motion

Given the situation where speaker  $A$  confirms agreement/disagreement of listener  $B$  and  $B$  reacts to the confirmation, we have four timing structures,  $\mathbf{I}_{UU}$ ,  $\mathbf{I}_{MU}$ ,  $\mathbf{I}_{UM}$ , and  $\mathbf{I}_{MM}$  expressing the interval between utterance  $\mathbf{U}$  and body motion  $\mathbf{M}$  of the two dialogists (see Fig.2). The details on the definition of the timing structures are described below.  $T(\mathbf{E}_t^X)$  denotes the time when an event  $\mathbf{E}_t$  caused by dialogist  $X$  occurs.  $\mathbf{E}_s$  and  $\mathbf{E}_e$  represent the start and end point of  $\mathbf{E}_t$ , respectively.  $\mathbf{U}(cnf)$ ,  $\mathbf{U}(rct)$ ,  $\mathbf{M}(cnf)$ , and  $\mathbf{M}(rct)$  denote utterance of *Confirmation*, utterance of *Reaction*, *face-turning* of *Confirmation*, body motion of *Reaction*, respectively.

The timing structure  $\mathbf{I}_{UU}$ , which is the interval between the end time of utterance of *Confirmation* by speaker  $A$  and the start time of utterance of *Reaction* by listener  $B$ , is described as

$$\mathbf{I}_{UU} = T(\mathbf{U}(rct)_s^B) - T(\mathbf{U}(cnf)_e^A). \quad (1)$$

Intention of speech is often revealed at the end time of its utterance. We consider  $T(\mathbf{U}(cnf)_e^A)$  as the base time of  $\mathbf{I}_{UU}$ .

The timing structure  $\mathbf{I}_{MU}$ , which is the interval between the start time of *face-turning* of *Confirmation* by speaker  $A$  and the start time of utterance of *Reaction* by listener  $B$ , is described as

$$\mathbf{I}_{MU} = T(\mathbf{U}(rct)_s^B) - T(\mathbf{M}(cnf)_s^A). \quad (2)$$

Listener  $B$  can observe *face-turning* by speaker  $A$  at the start time of it. We consider  $T(\mathbf{M}(cnf)_s^A)$  as the base time of  $\mathbf{I}_{MU}$ . Here, *face-turning* of *Confirmation* is defined below. When *face-turning*,  $\mathbf{M}^*$ , satisfies all of the following equations:

$$\begin{cases} T(\mathbf{M}_s^*) \geq T(\mathbf{U}(cnf)_s) \end{cases} \quad (3a)$$

$$\begin{cases} T(\mathbf{M}_s^*) \leq T(\mathbf{U}(rct)_s) \end{cases} \quad (3b)$$

$$\begin{cases} T(\mathbf{M}_s^*) \leq T(\mathbf{M}(rct)_s), \end{cases} \quad (3c)$$

the  $\mathbf{M}^*$  is regarded as  $\mathbf{M}(cnf)$ . If some *face-turnings* satisfy these conditional equations, the latest one is regarded as  $\mathbf{M}(cnf)$ . Note that  $\mathbf{I}_{\mathbf{MU}}$  is always positive value because of satisfying eq.(3b).

The timing structure  $\mathbf{I}_{\mathbf{UM}}$ , which is the interval between the end time of utterance of *Confirmation* by speaker *A* and the start time of body motion of *Reaction* by listener *B*, is described as

$$\mathbf{I}_{\mathbf{UM}} = T\left(\mathbf{M}(rct)_s^B\right) - T\left(\mathbf{U}(cnf)_e^A\right). \quad (4)$$

The body motion of *Reaction* is defined below. When *face-turning*, nodding, head-inclining, or head-shaking,  $\mathbf{M}^*$ , satisfies the following conditional equation:

$$T(\mathbf{U}(cnf)_s) \leq T(\mathbf{M}_s^*) \leq T(\mathbf{U}(rct)_e), \quad (5)$$

and also satisfies either one of the following conditional equation:

$$\left\{ \begin{array}{l} T(\mathbf{U}(cnf)_e) - T(\mathbf{M}_s^*) \leq 500\text{msec} \end{array} \right. \quad (6a)$$

$$\left\{ \begin{array}{l} T(\mathbf{U}(rct)_s) - T(\mathbf{M}_s^*) \leq 500\text{msec}, \end{array} \right. \quad (6b)$$

the  $\mathbf{M}^*$  is regarded as  $\mathbf{M}(rct)$ . If some body motions satisfy the conditional equations, the earliest one is regarded as  $\mathbf{M}(rct)$ .

The timing structure  $\mathbf{I}_{\mathbf{MM}}$ , which is the interval between the start time of *face-turning* of *Confirmation* by speaker *A* and the start time of body motion of *Reaction* by listener *B*, is described as

$$\mathbf{I}_{\mathbf{MM}} = T\left(\mathbf{M}(rct)_s^B\right) - T\left(\mathbf{M}(cnf)_s^A\right). \quad (7)$$

Note that  $\mathbf{I}_{\mathbf{MM}}$  is always positive value because of satisfying eq.(3c).

### 2.3 Relations between the Timing Structures and the Respondent's Internal State

Some researchers have argued that the respondent unconsciously controls the timing structure  $\mathbf{I}_{\mathbf{UU}}$  as the negative response timing is later than the positive one in the task-oriented dialogue [4]. The respondent would also control the timing of utterance,  $\mathbf{I}_{\mathbf{UU}}$ , according to agreement/disagreement with the partner. We make the first hypothesis as the following:

#### Hypothesis 1: Relation between $\mathbf{I}_{\mathbf{UU}}$ and the Internal State

$\mathbf{I}_{\mathbf{UU}}$  based on *Reaction* with disagreement is later than agreement.

Although the dialogists basically use speech to convey information to each other, they emphasize and supplement it by body motion except in some situations<sup>1</sup>. The body motion has often unignorable effects on the dialogue. Especially, *face-turning* of *Confirmation* denotes the action to observe some behaviors by the respondent, and would also suggest the action to induce some reactions reflecting his internal state. The respondent may control  $\mathbf{I}_{\mathbf{UU}}$  more explicitly than **Hypothesis1**. We make the second hypothesis as the following:

<sup>1</sup> Telephone is one of exceptive situations.

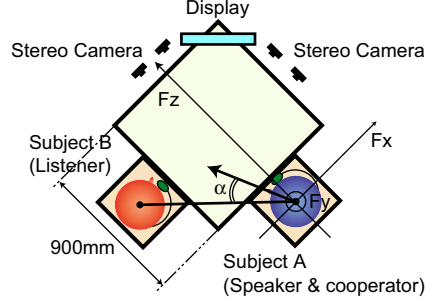


Fig. 3. Dialogue environment.



Fig. 4. Captured images.

### Hypothesis 2: Variations of $I_{UU}$ by Face-Turning

*Face-turning of Confirmation* increases the difference between  $I_{UU}$  of agreement and disagreement supporting **Hypothesis1**.

The timing structures based on the respondent's body motion,  $I_{UM}$  and  $I_{MM}$  would be controlled as well as  $I_{UU}$ . We make the third hypothesis below:

### Hypothesis 3: Relations between $I_{UM}$ , $I_{MM}$ and the Internal State

$I_{UM}$  and  $I_{MM}$  based on *Reaction* with disagreement are later than agreement. *Face-turning of Confirmation* increases the differences between  $I_{UM}$ , and  $I_{MM}$  of agreement and disagreement.

In order to test these hypotheses, we analyze video data and audio signals recorded in some consensus building dialogues, and clarify how the respondent controls  $I_{UU}$ ,  $I_{UM}$ ,  $I_{MU}$ , and  $I_{MM}$  in the following situations:

- (s1) agreement-*Reaction* for *Confirmation* by only utterance,
- (s2) disagreement-*Reaction* for *Confirmation* by only utterance,
- (s3) agreement-*Reaction* for *Confirmation* by utterance and *face-turning*,
- (s4) disagreement-*Reaction* for *Confirmation* by utterance and *face-turning*.

## 3 Construction of Dialogue Corpus

### 3.1 Recording the Consensus Building Dialogue

We constructed a dialogue corpus to analyze the four situations described in section 2.3. The dialogue task was either-or quiz. Fig.3 shows the overview of dialogue environment. A display which shows the task and two alternatives was set at the corner of the square desk, and two subjects interacted side by side in order to make them produce *face-turning* obviously.

To evenly analyze the four situations, we regulated experimental conditions by introducing an experimental cooperater to either subject. Just before they started the dialogue, the experimenter presented the quiz to each subject, and

queried about the answer to only a real subject as preliminary survey. Then, the experimenter instructed the cooperator what to do according to the answer, e.g. *Confirmation* by only utterance expressing agreement with the answer. The real subject could not sense the instruction. The experimenter also made the cooperator say “If I choose among them, I like XXX” in the dialogue, to regulate speech of *Confirmation* under constant speech rate and prosody. XXX is either alternative depending on the answer. In addition, it is important to keep a regular timing,  $\mathbf{I}_{\text{cnf}} (= T(\mathbf{M}(\text{cnf})_s) - T(\mathbf{U}(\text{cnf})_e))$ , between utterance and body motion of *Confirmation*. We investigated the distribution of  $\mathbf{I}_{\text{cnf}}$  by preliminary experiments.  $\mathbf{I}_{\text{cnf}}$  was mostly distributed between  $-750$  and  $0$ msec. We trained the cooperator to do *face-turning* according to the timing and also analyzed only experimental data which satisfy the following conditional equation:  $-750\text{msec} \leq \mathbf{I}_{\text{cnf}} \leq 0\text{msec}$ .

It is not natural to make the cooperator produce the regulated behavior frequently in a dialogue. We made a rule that the experimenter chooses a preceding speaker of the dialogue and the cooperator produces the behavior once in the first exchange only when he is chosen. We analyzed only this first exchange.

We had 13 pairs of dialogists of the same sex with friendships<sup>2</sup>. We conducted each experiment based on situation (s1)  $\sim$  (s4) twice every pair, i.e. we recorded 104 dialogues<sup>3</sup>.

### 3.2 Extraction of Utterance and Body Motion Events

Each subject wore a directional headset microphone. Stereo cameras were set on the opposite side of each subject as shown in Fig.3. The resolution of the video data was  $1024 \times 768$ pixel and the frame rate was 30fps. The sampling rate of the audio signals was 44.1kHz.

#### Extraction of Utterance Events

We detected the start and end of utterances deleting non-verbal sounds, such as laugh and cough, by using the sound processing tool “Wavesurfer”<sup>4</sup>. If there was the silent interval longer than 400msec, we regarded it as a pause between utterances. We labeled each utterance *Confirmation*, *Reaction*, and *Others*.

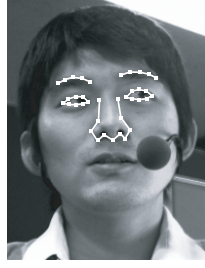
#### Extraction of Body Motion Events

*Face-turning*, nodding, head-inclining, and head-shaking were extracted by measuring the amount of head rotation and head translation. We defined the subject coordinate system whose origin is the centroid of subject’s head positions shown in Fig.3.  $F_z$ -axis of the system is a horizontal line from the origin to the center of baseline of the stereo camera,  $F_y$ -axis is a vertical line for  $F_z$  from the origin, and  $F_x$ -axis is an orthogonal line to both  $F_y$ -axis and  $F_z$ -axis. We measured the amount on the system. Let  $\alpha$  be an angle between the subject’s

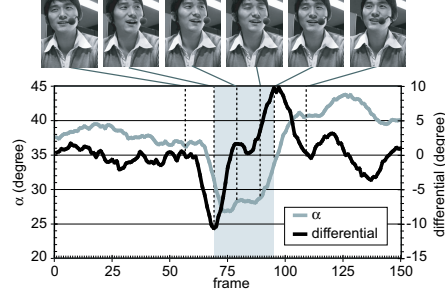
<sup>2</sup> 13 pairs are the subset combination of 4 cooperators and 13 real subjects.

<sup>3</sup> 13 pairs  $\times$  4 situations  $\times$  2

<sup>4</sup> <http://www.speech.kth.se/wavesurfer/>



**Fig. 5.** Facial feature points extracted by AAM.



**Fig. 6.** Extracted interval of *face-turning*.

head rotation vector around  $F_y$ -axis and the vector from the subject to another. We defined the time when the temporal differentiation of  $\alpha$  goes under/over a certain threshold as the start/end time of *face-turning*. We also extracted nodding, head-inclining, and head-shaking based on the amount of head rotation around  $F_x$ -,  $F_z$ -, and  $F_y$ -axis, respectively. We defined the time when the absolute amount of the temporal differentiation of head rotation around each axis goes over a certain threshold as the start/end time of each body motion.

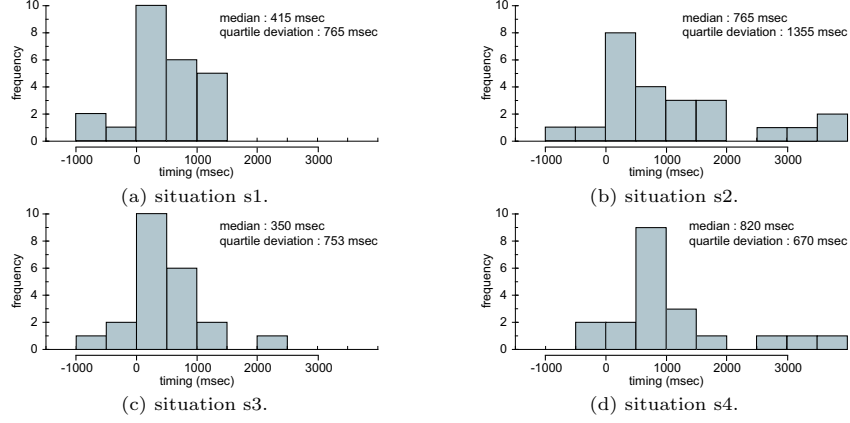
The amount of head rotation and translation were estimated as below. Firstly, we applied AAM (Active Appearance Model [6]) to captured face images (Fig.4) in order to obtain 2-D positions of facial feature points as shown in Fig.5. Secondly, we applied the stereo measurement method [7] to 2-D positions of the points in order to obtain 3-D positions of them. Thirdly, we obtained the amount of head rotation and translation by solving the minimization problem,

$$\min_{\mathbf{R}, \mathbf{T}} \sum_{i=1}^n \| \mathbf{Q}' - (\mathbf{R}\mathbf{Q} + \mathbf{T}) \|^2, \quad (8)$$

where  $n$  is the number of the facial feature points,  $\mathbf{Q}'$  denotes a  $3 \times 1$  vector which has 3-D position of capturing facial feature point  $i$ ,  $\mathbf{Q}$  denotes that of the face model which was constructed from the first frame of the captured image sequences,  $\mathbf{R}$  denotes a  $3 \times 3$  rotation matrix, and  $\mathbf{T}$  denotes a  $3 \times 1$  translation vector. We used the algorithm based on singular value decomposition method (SVD) to solve eq.(8) [8]. The algorithm applies SVD to  $\mathbf{Q}\mathbf{Q}'^T (= \mathbf{U}\mathbf{S}\mathbf{V}^T)$ . The head rotation,  $\mathbf{R}$ , was calculated as  $\mathbf{R} = \mathbf{V}\mathbf{U}^T$ . And then the head translation,  $\mathbf{T}$ , was measured by the difference between centroid of the 3-D points of capturing face and the face model. Fig.6 shows an example of extracted *face-turning*.

### Manual Compensation of Extracted Events

We used the annotation tool “Anvil” [9] and compensated some start/end times of utterance and body motion.

Fig. 7. Histograms of  $\mathbf{I}_{UU}$ .

## 4 Analysis of Relations between the Timing Structures and the Internal State and Test of the Hypotheses

### 4.1 Analysis of the Timing Structures on the Reaction by Utterance

At first, we analyze the timing structures  $\mathbf{I}_{UU}$  and  $\mathbf{I}_{MU}$  on utterance of *Reaction*. The utterance occurred for each utterance of *Confirmation* in the whole of the experiments. The number of samples for situation (s1), (s2), (s3), and (s4) were 24, 24, 22, and 20, respectively.

#### Distributions of $\mathbf{I}_{UU}$

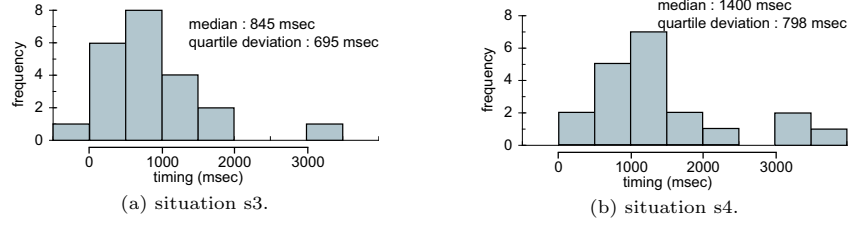
The distributions of  $\mathbf{I}_{UU}$  resulted from the dialogue corpus are shown in Fig.7. We focus on the median and the quartile deviation extracted from the distributions to eliminate outliers from sample statistics. The medians of timings of agreement- and disagreement-*Reaction* for *Confirmation* without *face-turning* were 415msec and 765msec, respectively. The quartile deviations of their timings were 765msec and 1355msec, respectively. Each peak of the distributions was between 0 and 500msec.

And the medians of timings of agreement- and disagreement-*Reaction* for *Confirmation* with *face-turning* were 350msec and 820msec, respectively. The quartile deviations of their timings were 753msec and 670msec, respectively. The peak timing of the distribution of disagreement-*Reaction* (500 ~ 1000msec) was later than that of agreement-*Reaction* (0 ~ 500msec).

#### Distributions of $\mathbf{I}_{MU}$

The distributions of  $\mathbf{I}_{MU}$  are shown in Fig.8.  $\mathbf{I}_{MU}$  has distributions and statistics similar to  $\mathbf{I}_{UU}$  because we regulated  $\mathbf{I}_{cnf}$  within a defined span as described in section 3.1.



Fig. 8. Histograms of  $I_{MU}$ .Table 1. Occurrence frequency of  $M(rct)$ .

situation	frequency of $M(rct)$ / frequency of $U(cnf)$
s1	14 / 24 (58.3%)
s2	20 / 24 (83.3%)
s3	18 / 22 (81.8%)
s4	9 / 20 (45.0%)

### Test of Hypothesis 1

The timing of disagreement-*Reaction* for *Confirmation* without *face-turning* was later than that of agreement. This result supports **Hypothesis 1**. However, the result of the median test does not show significant difference between medians of their timings (significant probability  $p = 0.56$ ).

### Test of Hypothesis 2

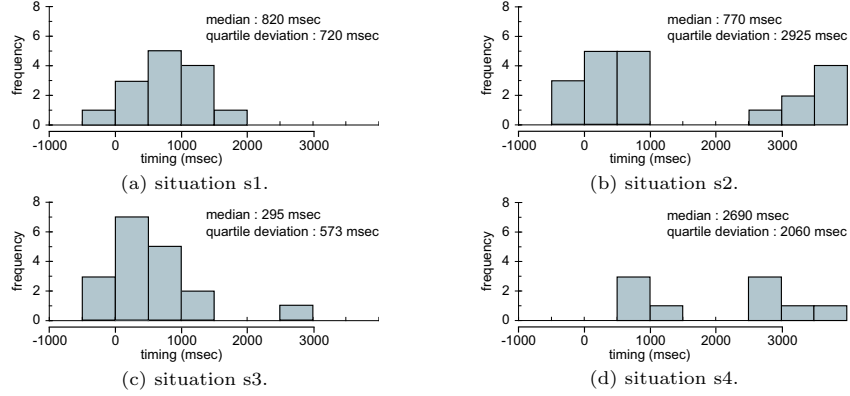
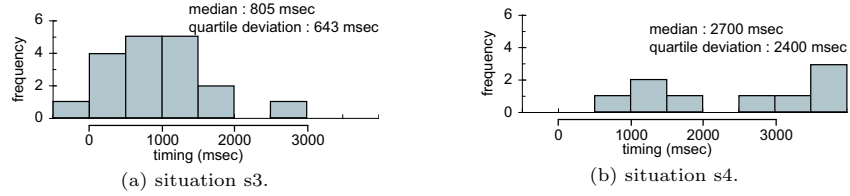
The timing of disagreement-*Reaction* for *Confirmation* with *face-turning* was later than that of agreement, as well as the timing structure without *face-turning*. The result of the median test shows significant difference between medians of their timings ( $p = 0.013$ ). The difference was significantly increased by *face-turning* because of not significant difference in case of *Confirmation* without *face-turning*. This result supports **Hypothesis 2**.

## 4.2 Analysis of the Timing Structures on the Reaction by Body Motion

We analyze the timing structures  $I_{UM}$  and  $I_{MM}$  on body motion of *Reaction*. The body motion did not occur for every utterance of *Confirmation* in the whole of the experiments. Table 1 shows frequency of the occurrence in each experimental situation.

### Distributions of $I_{UM}$

The distributions of  $I_{UM}$  are shown in Fig.9. Agreement-*Reaction* occurred more frequently between 0 and 2000msec. Disagreement-*Reaction* tended to occur frequently in two parts of span (0 ~ 1000msec and 2500msec ~).

Fig. 9. Histograms of  $I_{UM}$ .Fig. 10. Histograms of  $I_{MM}$ .

The medians of timings of agreement- and disagreement-*Reaction* for *Confirmation* without *face-turning* were 820msec and 770msec, respectively. The quartile deviations of their timings were 720msec and 2925msec, respectively. And the medians of timings of agreement- and disagreement-*Reaction* for *Confirmation* with *face-turning* were 295msec and 2690msec, respectively. The quartile deviations of their timings were 573msec and 2060msec, respectively.

### Distributions of $I_{MM}$

The distributions of  $I_{MM}$  are shown in Fig.10.  $I_{MM}$  has distributions and statistics similar to  $I_{UM}$  because we regulated  $I_{cnf}$ .

### Test of Hypothesis 3

Although distributions of  $I_{UM}$  and  $I_{MM}$  were similar to  $I_{UU}$  and  $I_{MU}$ , sample numbers of the distributions were less and uneven as shown in Table 1. We shelve test of **Hypothesis 3** by the statistical significant difference.

Table 1 shows that the body motion tends to occur as below.

- ★ For *Confirmation* without *face-turning*, the respondent more frequently produces body motion in case of disagreement-*Reaction* than agreement-*Reaction*.
- ★ For *Confirmation* with *face-turning*, the respondent more frequently produces body motion in case of agreement-*Reaction* than disagreement-*Reaction*.

We discuss the reason of this tendency in the next section.

## 5 Discussion -What does the Face-Turning Action Imply in Consensus Building Communication?

We first consider the reason why agreement-*Reaction* occurred at early timing, based on the discussion by Fujiwara *et al.* [4]. If the respondent reacts at late timing, his partner may imagine that he is not sure which is better. The respondent, therefore, wants to convey agreement to his partner soon. In case of disagreement-*Reaction*, the respondent needs to choose his words with full respect for his partner and think how to *Reaction* in order to make the partner understand his intention. Therefore, disagreement-*Reaction* occurred at late timing. Also, the timing structure must be an implicit rule to express agreement/disagreement.

We confirmed that *face-turning* of *Confirmation* increased the difference between the timing of agreement- and disagreement-*Reaction*. Based on the approach-avoidance model by Argyle *et al.* [2], we consider that the *face-turning* suggests the approach action. Therefore, the respondent produced agreement-*Reaction*, which is regarded as the approach response, at earlier timing and he produced disagreement-*Reaction*, which is regarded as the avoidance response, at later timing. Also, the *face-turning* often induces the reaction. This inducibility can be confirmed from a result that the quartile deviation of timings of disagreement-*Reaction* for *Confirmation* with *face-turning* (Fig.7(d)) was much smaller than that without *face-turning* (Fig.7(b)). The *face-turning* has the potency inducing the partner to react by utterance within a certain temporal interval. Through these results, we consider that the *face-turning* controls the timing of partner's reaction and elicits his internal state. The dialogists can build up a sense of dialogue timing by producing the explicit event of *face-turning*.

As shown in Table 1, the *face-turning* promoted the body motion suggesting agreement and restrained the motion suggesting disagreement. When the listener receives *Confirmation* without *face-turning*, he is not pressured by partner's observation. Therefore, he does not need to express agreement by using body motion excessively. On the other hand, he produces the motion suggesting disagreement unconsciously. In case of *Confirmation* with *face-turning*, it is a kind of *Reaction* with disagreement that he does not produce the motion. He nods to convey agreement to the partner excessively, and he often suggests disagreement without the motion. *Face-turning* opens new dialogue channel of body motion, and it is possible to change the meaning of producing the motion.

It is difficult to detect false agreement/disagreement, i.e. lie, when confirming partner's intention without *face-turning*, because the timing of agreement-reaction is almost the same as that of disagreement-reaction. By producing the *face-turning*, the dialogists can observe the reaction by body motion and estimate partner's real agreement/disagreement by sensing the timing of the reaction. We actually turn our face toward partner when we cannot understand partner's internal state and want to understand the state surely in daily communication. *Face-tuning* is an action to probe partner's mind <sup>5</sup>.

<sup>5</sup> We have proposed a new proactive human-machine interaction model, *Mind Probing*.

## 6 Conclusion

We made some hypotheses about the relations between the timing of the listener's reaction for the confirmation of his internal state by the speaker and the state. We analyzed some consensus building dialogues by an experimental cooperator and a subject to test the hypotheses. Especially, we focused on the action that the speaker turns his face toward the listener as the confirming action and analyzed how it influences the timing structures.

As the results, the timing of the reaction by utterance suggesting disagreement was later than that of agreement, and the *face-turning* increased the difference between the timing of the reaction suggesting agreement and disagreement. We consider that the *face-turning* is a trigger to make the dialogists build up a sense of dialogue timing. The *face-turning* also promoted the body motion suggesting agreement and restrained the motion suggesting disagreement. It was contrary to result for the confirmation without *face-turning*. This result must be a feature of dialogue opening the dialogue channel of body motion.

In future work, we need to increase the size of the dialogue corpus and evaluate reliability of these results. We will design the human-machine interaction system able to estimate user's internal state based on findings of this work.

**Acknowledgments.** This work is in part supported by Grant-in-Aid for Scientific Research of the Ministry of Education, Culture, Sports, Science and Technology of Japan under the contract of 18049046.

## References

1. T. J. Jacobs. On Unconscious Communications and Covert Enactments: Some Reflections on Their Role in the Analytic Situation. *Psychoanalytic Inquiry*, 21:4–23, 2001.
2. M. Argyle and J. Dean. Eye Contact, Distance and Affiliation. *Sociometry*, 28:289–304, 1965.
3. A. Kendon. Some Function of Gaze-direction in Social Interaction. *Acta Psychologica*, 26:22–63, 1967.
4. N. Fujiwara, T. Itoh, and K. Araki. Analysis of Changes in Dialogue Rhythm due to Dialogue Acts in Task-oriented Dialogues. *Springer-Verlag Lecture Notes in Artificial Intelligence(LNAI)*, 4629:564–573, 2007.
5. M. Yoshida and Y. Miyake. Relationship between Utterance Dynamics and Pragmatics in the Conversation of Consensus Building Process. *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication*, 641–645, 2006.
6. T.F.Cootes, G.J.Edwards, and C.J.Taylor. Active Appearance Model. *Proceedings of European Conference on Computer Vision*, 2:484–498, 1998.
7. R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. *Cambridge University Press*, 2004.
8. K.S.Arun, T.S.Huang, and S.D.Blostein. Least-squares Fitting of Two 3-d Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–701, 1987.
9. M. Kipp. Gesture Generation by Imitation - From Human Behavior to Computer Character Animation. *Boca Raton, Florida: Dissertation.com*, 2004.