2012 IEEE/SICE International Symposium on System Integration (SII)
Kyushu University, Fukuoka, Japan
December 16-18, 2012

Synchrony of utterance rhythms and context of repetitive dialogue in a cooperation game

Tomohito Yamamoto and Yoshihiro Miyake

Abstract— **Speech dialogue systems, such as Apple's "Siri," have gradually become more widespread, and in the near future, a greater number of general users will have the opportunity to communicate with such systems. To facilitate this, it is necessary for the system to communicate more naturally with users, and to realize that both verbal and nonverbal information must be taken into consideration. Therefore, a model that can select contextually appropriate nonverbal information should be developed. However, the relation between context and nonverbal information has not been sufficiently analyzed, because it is difficult to control the context of communication in an experiment. In this research, we clarify the relation between nonverbal information, such as the utterance rhythm, and the context of a dialogue by analyzing the correlation of utterance durations with the game state in a kind of "prisoner's dilemma" that we introduce to control the context. The results show diverse correlations across four game states (Cooperation-Cooperation, Cooperation-Betrayal, Betrayal-Cooperation, Betrayal-Betrayal); however, in the Cooperation-Cooperation state, a positive correlation is often observed between the duration of utterances. From these results, we discuss the mechanism of synchrony between utterance rhythms.**

I. INTRODUCTION

Speech dialogue systems have frequently been used for the audio guidance of public facilities such as a museum or car navigation systems. In addition, in recent years they have been introduced into communication systems for elderly people [1] and smartphones' speech assistants, such as Apple's "Siri." In the near future, as smartphones and tablet PCs become more common, a greater number of general users will have the opportunity to communicate with such systems. Therefore, it is necessary that speech dialogue systems can communicate naturally with their users.

Generally speaking, natural communication requires participants to consider not only verbal information, but also nonverbal information such as the utterance rhythm and prosody [2]. Thus, a model that can select contextually proper nonverbal information must be developed to realize an effective artificial communication system. However, the relation between context and nonverbal information has not yet been sufficiently analyzed. Because of the difficulty in controlling the context of communication dynamically in an experimental environment, most research into this relation has only dealt with static situations in a passive experimental approach [3]. In this research, we introduce a kind of cooperation game for controlling the context dynamically, and investigate the nonverbal information in human dialogue. Hence, we are able to clarify the relation between context and nonverbal communication.

There is a considerable amount of nonverbal information in human communication. Of this, the utterance rhythm is an important factor in natural dialogue, and has been analyzed frequently. For example, Matarazzo et al. and Webb have studied the utterance and pause rhythms in human dialogue, and found that such rhythms were synchronized between speakers [4], [5], [6]. Nagaoka et al. and Koiso et al. have also reported synchronization between speakers' pause duration and speech speed [7], [8]. This is sometimes called the "synchrony tendency" [9], and is an important index for evaluating smooth or good communication. From this point of view, some researchers have applied the synchrony phenomenon to communication systems. For example, Watanabe and collaborators have applied the entrainment of utterance and nodding rhythms to communication systems, and reported that it was effective in promoting interaction between humans and artificial systems [10], [11]. Moreover, Kitaoka et al. developed a decision tree for utterance timings from a corpus of conversations, and constructed a dialogue system that was able to speak with a proper utterance timing [12].

Our research group has also analyzed the synchrony in human dialogue [13]. In that research, we analyzed the effect of changes in intentional utterance speed on the synchrony of utterance and pause rhythms in a dialogue composed of an instruction and response utterance. The results showed that the correlation between the duration of an instruction utterance and that of the switching pause was negative and low when the change of utterance speed was too small to be noticeable. However, the correlation became positive and high when the change of utterance speed was increased. We applied these results to a communication robot and evaluated the outcome [14].

From these previous studies, it is clear that synchrony in human dialogue has already been analyzed. However, the relation between synchrony and context (in this research, we consider context to be the content of dialogue and the accompanying mind state) have rarely been analyzed. This is primarily because it is difficult to control the context of dialogue in an experimental setting. However, if we wish to realize natural communication between a human and a robot or artificial dialogue system, it is necessary to elicit the relation between context and synchrony. Therefore, in

Tomohito Yamamoto is with the Department of Information and Computer Engineering, Kanazawa Institute of Technology, 7-1 Ohgigaoka, Nonoichi, Ishikawa, Japan tyama@neptune.kanazawa-it.ac.jp

Yoshihiro Miyake is with the Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, 4259 Nagatsuta, Midori,Yokohama, Japan miyake@dis.titech.ac.jp

978-1-4673-1497-8/12/$31.00 ©2012 IEEE

this research, we introduce a similar game to the ”prisoner's dilemma” to control the context, and analyze the relation between the subject's mind state and the synchrony of utterance rhythms.

II. METHOD

A. Experimental task and procedure

The purpose of this research is to reveal the relation between the context, which is controlled by the cooperation game, and the synchrony of utterance rhythms in human dialogue. In our previous research, we analyzed synchrony using a dialogue composed of two sentences [13]. However, it was difficult to control and analyze the context of such noncontinuous dialogue. In this research, we focus on changing the mind state of the subject with short but continuous dialogue in a scenario based around game theory. Moreover, we propose a new game based on the repetitive prisoner's dilemma, and analyze subjects' dialogue while playing the game.

In our game, subjects are divided into a precedent speaker and a following speaker, who hold a repetitive dialogue. The dialogue is composed of precedent and following utterances, the content of which is only “I cooperate with you” (Watashiha kyouryoku simasu, in Japanese). There is no restriction about utterance without this dialogue content (Subjects could speak at any spead or pitch in the experiment). Of course, there a lot of possibility of dialogue content in this experiment. However, in this research, in order to form some connectivity with our previous research, we took a bottom-up approach from simple dialogue to general dialogue.

The experiment is conducted according to the following procedure:

1) The experiment controller explains the outline and strategy of the cooperative game, and performs some exercises.
2) Subjects sit down and wear the headset microphone (Figs. 1, 2).
3) The controller gives the sign to start, and the subjects say “I cooperate with you (Watashiha kyouryoku simasu)” alternately.
4) After the precedent speaker has spoken six times, one game is finished.
5) Subjects choose a “Cooperation” or “Betray” card and place it face up.
6) Finally, subjects turn the card according to the controller's instructions.

One game is composed of procedures 3-6, and an experiment is composed of 10 games. After procedure 6, subjects score points according to Table 1. Following an experiment, subjects are given a real prize corresponding to their score, with a top prize for scores above 44 points, a middle prize for scores above 40 points, and a low prize for scores above 35 points (actual price ratio of low:middle:top = 1:2:4). In this game, for subjects to win a middle prize, the best choice is the ”Cooperation” card. However, to win the top prize, subjects must choose the ”Betray” card more than once. Therefore, it is expected that subjects will sometimes choose the ”Cooperation” card, and sometimes choose the ”Betray” card. In this way, as the game proceeds, the subjects' mind state will change with the choice of card. In this research, by focusing on the utterance and pause duration, we analyze the relation between the context and the synchrony of utterance rhythms.

TABLE I
SCORE TABLE

Card type	Cooperation	Betrayal
Cooperation	(4,4)	(8,0)
Betrayal	(0,8)	(1,1)

Fig. 1. Picture of the experimental setup

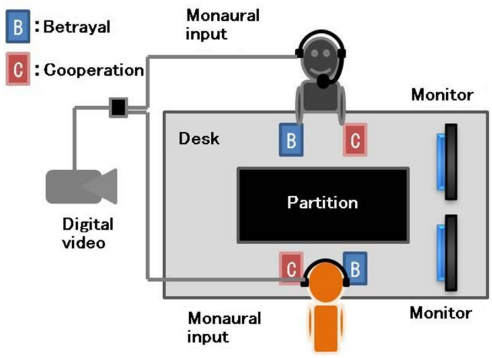

Fig. 2. Diagram of the experimental setup

To avoid any effects from other humans, it would be ideal to operate each procedure by computer system. However, as it is difficult to predict the experimental situation, human staff operated all procedures.

B. Subjects

Thirty male students (all in their 20s) were selected to participate in this experiment. The precedent and following speakers were selected randomly. If a pair of subjects had a friendship or other relationship, one was exchanged in order to exclude social effects.

C. Experimental system

Fig. 2 shows the experimental system. To record audio data, we use a video camera (SONY: DS-SR8) and a headset (Audio Technicha: PRO&HEW/P). Two monitors allow the subjects to track the state of the game (number of utterances,

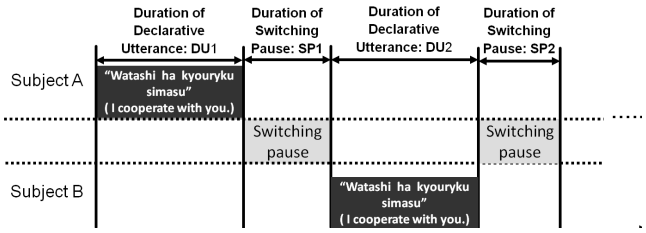

Fig. 3. Indices of dialogue

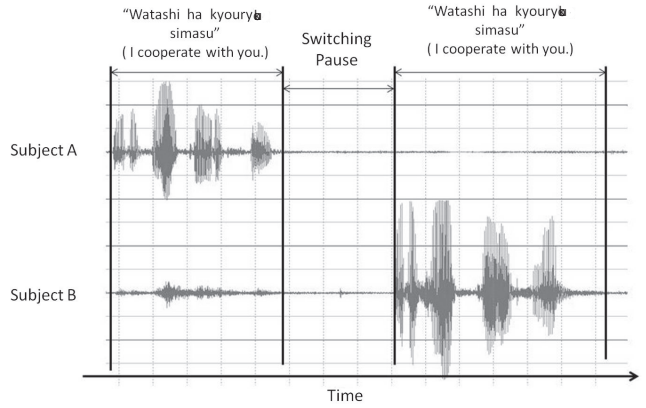

Fig. 4. Calculation of sound duration

number of games, present score). Therefore, subjects are constantly able to recognize the game situation.

D. Data analysis

Fig. 3 illustrates the indices used to analyze the dialogue time structure. The analysis items are the precedent and following speaker's duration of utterance (DU1, DU2), and the switching pause between precedent and following utterance (SP1), or between following and precedent utterance (SP2). In this research, we calculate the correlation coefficients between DU1-DU2, DU2-DU1, DU1-SP1, DU2-SP2, SP1-SP2, and SP2-SP1 from five alternate utterance and pause durations. However, for brevity, only the data from the precedent speaker (DU1-DU2, DU1-SP1, SP1-SP2) are shown in the following section.

The utterance duration is calculated as the interval of the wave data that is bigger than noise level(= 5 dB) (Fig. 4). However, if the noise of the subjects' breath is overlapped or the recording level is low, it is difficult to detect this interval. In such cases, the interval is determined by examining audio and visual information.

III. RESULTS

A. Duration of utterance and switching pause

Fig. 5 shows a typical example of the duration of an utterance, and Fig. 6 shows the duration of a switching pause. In these figures, DU1 and SP2 correspond to the precedent speaker, and DU2 and SP1 correspond to the following speaker.

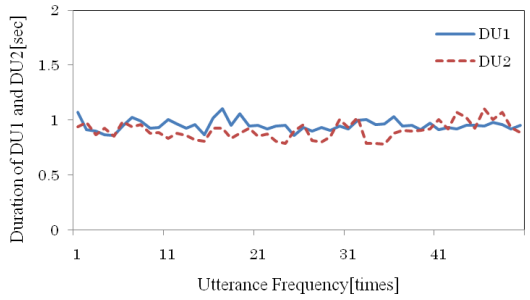

Fig. 5. Time series of declarative utterances

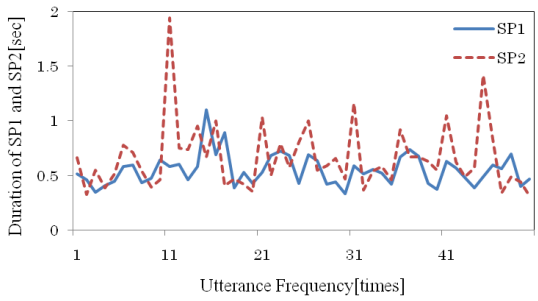

Fig. 6. Time series of switching pauses

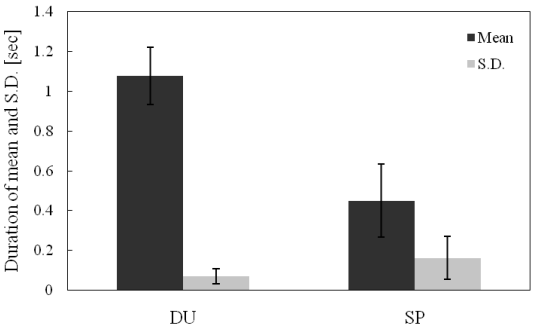

Fig. 7. Mean and S.D. of DU and SP

In Fig. 5, the time series of DU1 and DU2 are relatively flat throughout the 10 games. However, in Fig. 6, the time series fluctuate much more than the DU. Fig. 7 shows the mean and S.D. of DU and SP. These values are calculated from 15 subject pairs' data. The results of a t-test show that the mean of DU is larger than that of SP ($t(29) = 16.9, p < .01$), and the S.D. of DU is smaller than that of SP ($t(29) = -5.06, p < .01$). These results imply that SP has a smaller absolute value than DU, and is more likely to fluctuate.

B. Correlation between durations and game states

Figs. 8-10 show some examples of the time series of correlation coefficients. Correlation coefficients are calculated for each game (five alternate utterances and pause durations).

- Red:(Cooperation, Cooperation)
- Yellow:(Cooperation, Betray)

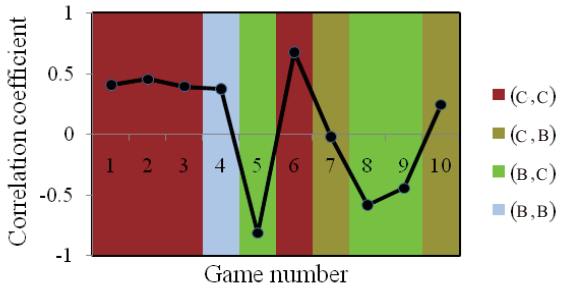

Fig. 8. Time series of correlation coefficients for DU1 and DU2. C = Cooperate, B = Betray; precedent speaker ' s decision is given on the left.

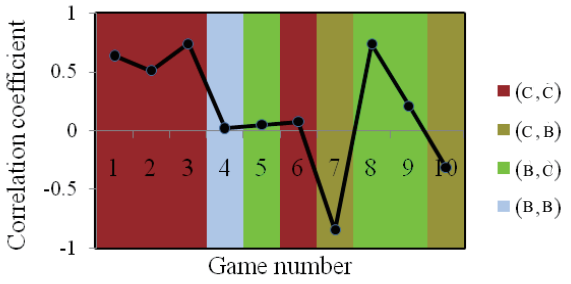

Fig. 9. Time series of correlation coefficients for DU1 and SP1

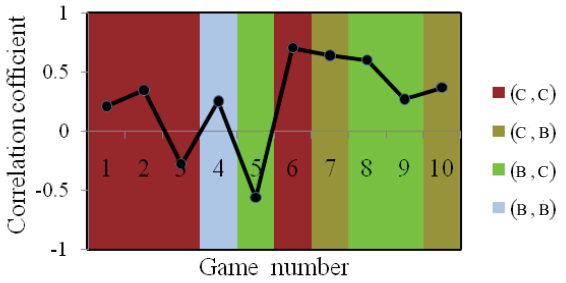

Fig. 10. Time series of correlation coefficients for SP1 and SP2

- Green:(Betray, Cooperation)
- Blue:(Betray, Betray)

In these figures, the coefficient values are changing from positive to negative. Moreover, the relation between the correlation coefficient and the game state is not always strong (e.g., when both subjects choose "Cooperation," the correlation coefficient is not always positive). To reveal the tendency of all the data, we analyze the relation between the correlation coefficient frequencies and the game states (Fig. 11: DU1-DU2, Fig. 12: DU1-SP1, and Fig. 13: SP1-SP2). The coefficient frequencies are relative because the frequency of each game state will be different.

In these figures, each of the correlation coefficients has a broad distribution from negative to positive. However, in Fig. 11, the value is biased toward the positive side when the game state is Red (C, C) or Yellow (C, B). To analyze

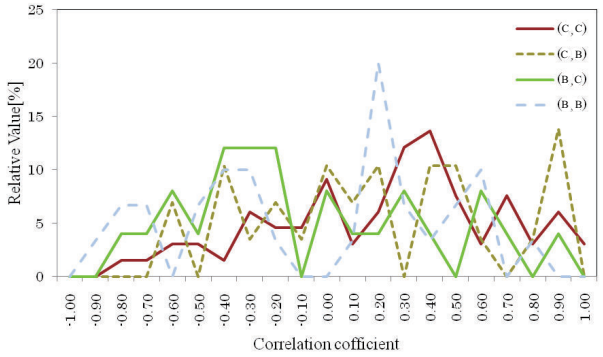

Fig. 11. Frequency plot of correlation coefficients between DU1 and DU2. C = Cooperate, B = Betray; precedent speaker ' s decision is given on the left.

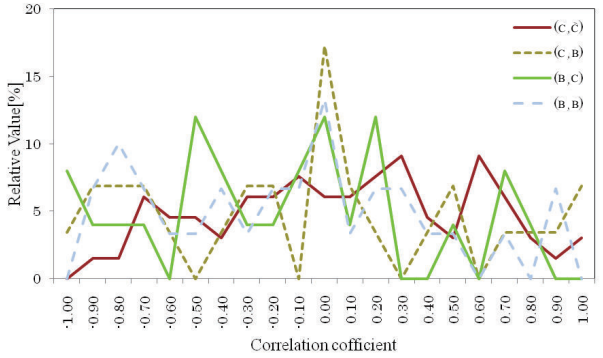

Fig. 12. Frequency plot of correlation coefficients between DU1 and SP1

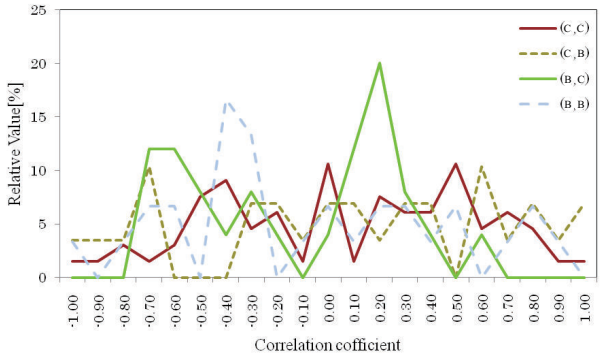

Fig. 13. Frequency plot of correlation coefficients between SP1 and SP2

this tendency quantitatively, we calculated the mean value of the correlation coefficient in each game state (Fig. 14). The results of a Kruskal-Wallis test for the median value of the correlation coefficients showed that there is significant difference in DU1-DU2 ($\chi(3) = 8.52, p < .05$). The results of a multiple comparison (Steel method) show that there is a significant difference between the correlation coefficients of Red (C, C) and Green (B, C) ($t = 2.37, p < .05$). These results mean that the correlation between DU1 and DU2 is changed by the game state, and a positive correlation is likely to appear when both subjects choose "Cooperation."

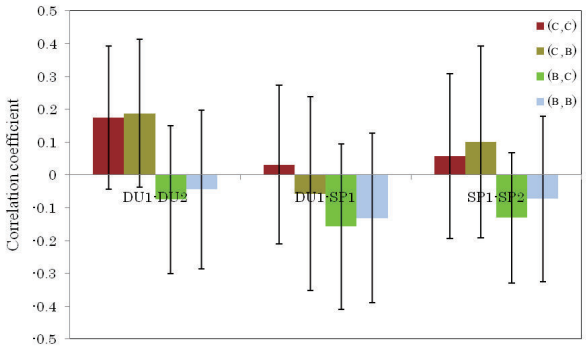

Fig. 14. Mean and S.D. of correlation coefficients in each game situation. C = Cooperate, B = Betray.

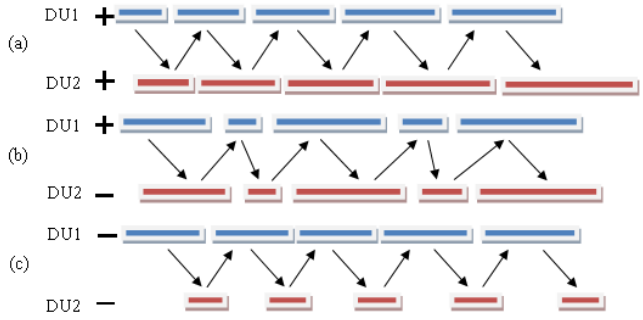

Fig. 15. Three correlation types of DU1 and DU2

IV. DISCUSSION

In this research, to reveal the relation between the context of a dialogue and the synchrony of utterance rhythms, we introduced a sort of "prisoner's dilemma" game. This enabled us to analyze the correlation between utterance and pause durations and the game state. We defined four game states (Cooperation-Cooperation, Cooperation-Betray, Betray-Cooperation, Betray-Betray), and studied the change in the subjects' mind state with the changing game states. We found that when both subjects choose the "Cooperation" card, the correlation between utterance durations is likely to become positive. This result means that the context of cooperation tends to produce synchrony in the utterance durations. In previous research [9], synchrony was assumed to be an index of smooth or good communication, and our result supports this hypothesis.

Moreover, instead of taking a passive approach to observing free conversation, we took an active approach to controlling the context in order to analyze the relation between context and synchrony. The results of our research suggest that such an approach is applicable for dialogue analysis, and necessary for applying the results to artificial dialogue systems.

We observed that the correlation coefficient between DU1-DU2, DU1-SP1, and SP1-SP2 had a broad distribution. In previous research (e.g., [4]), it was reported that these parameters enjoyed a positive correlation. However, negative correlations and large fluctuations have not previously been reported. From this point of view, the results of previous research have revealed only one aspect of synchrony. To reveal the mechanism of these positive and negative correlations, it is necessary to analyze each case of correlation in detail (DU1-DU2, DU1-SP1, SP1-SP2). However, it is difficult to analyze all of these combinations, therefore, in this paper, we focus on the correlation between DU1-DU2 and discuss its mechanism.

First, the correlation between utterance durations can be classified into three groups: (a) the correlation between precedent utterance and following utterance (DU1-DU2) is positive, and the correlation between following utterance and precedent utterance (DU2-DU1) is also positive; (b) the correlation between precedent utterance and following utterance is positive, and the correlation between following utterance and precedent utterance is negative, or vice versa; and (c) the correlation between precedent utterance and following utterance is negative, and the correlation between following utterance and precedent utterance is also negative. Fig. 15 illustrates these three relations. In the case of (a), both utterances become shorter and shorter, or longer and longer, in a dialogue. This situation has been supposed to explain the typical synchrony between utterance durations. However, if this situation continues, the utterance duration becomes extremely short or long, which rarely happens.

For correlation type (b), the following utterance becomes longer or shorter according to whether the precedent utterance is becoming longer or shorter. In this case, the utterance duration follows the pattern long, long, short, short, long, long, ..., and the duration does not become extremely short or long. In other words, a constant utterance rhythm is maintained between the speakers. In case (c), if the precedent utterance becomes long, then the following utterance becomes short, or vice versa. In this case, one speaker's utterance is always long and the other is always short, and this tendency becomes stronger and stronger.

Case (a) describes typical synchrony, and has been discussed in previous research. However, the other two cases are also interesting. For example, in case (b), two speakers seem to maintain an utterance rhythm unconsciously. This situation suggests that the mechanism that stabilizes the dialogue rhythm is built-in to human nature. In the case of (c), it is supposed that there is no interaction between speakers, and each keeps their own utterance rhythm. Actually, in Fig. 14, negative correlation values appear in the case of the (B, B) game state, suggesting that this situation reflects a poor or strained relationship.

In this discussion, we have focused on the correlation between utterance durations. However, it is necessary to consider other durations, such as the switching pause, which seems more reflective of the subject's mind state. Moreover, it is necessary to analyze not only the effect of one game state, but also that of continuous game states. In future work, to reveal the relation between context and synchrony

more clearly, we will focus on these factors and analyze the relations.

V. CONCLUSION

In this research, we studied the relation between the context of dialogue and the synchrony of utterance rhythms. We did this by analyzing the correlation between the utterance durations in a dialogue and the state of a repetitive prisoner's dilemma-type game. The results showed that various degrees of correlation could be observed in the four game states, although a positive correlation between the duration of utterances was often observed in the Cooperation-Cooperation state.

Our game restricted the dialogue to the single phrase "I cooperate with you" (Watashiha kyouryoku simasu in Japanese). Using this simple content, it was possible to analyze the micro-structure of the utterance rhythm. However, general dialogue is more complex, and it is necessary to investigate whether the results of this research can be applied to such dialogue. Therefore, in future work, we will prepare an experiment composed of plural dialogue, and investigate the relation between the context and synchrony of utterance rhythms.

REFERENCES

[1] Y. Kobayashi, D. Yamamoto, T. Koga, S. Yokoyama, and M. Doi, "Design targeting voice interface robot capable of active listening," Proceedings of the 5th ACM/IEEE International Conference on Human Robot Interaction (HRI 2010), pp.161-162, 2010

[2] M. L. Knapp, and J. A. Hall, "Nonverbal Communication in Human Interaction" (7th Ed.), Wadsworth, Cengage Learning, 2009

[3] T. Itoh, N. Kitaoka, and R. Nishimura, "Subjective experiments on influence of response timing in spoken dialogues," Proceedings of Interspeech 2009, pp.1835-1838, 2009

[4] J. D. Matarazzo, and A. N. Wiens, "Interviewer influence on durations of interviewee silence," Journal of Experimental Research in Personality, Vol.2, pp.56-69, 1967

[5] J. D. Matarazzo, M. Weitman, G.Saslow, and A. N. Wiens, "Interviewer influence on durations of interviewee speech," Journal of Verbal Learning and Verbal Behavior, Vol.1, pp.451-458, 1963

[6] J. T. Webb, "Interview synchrony: An investigation of two speech rate measures in an automated standardized interview," In B. Pope and A.W. Siegman (Eds.), Studies in dyadic communication, New York: Pergamon, pp.115-133, 1972

[7] C. Nagaoka, M. Komori, T. Nakamura, and M. R. Draguna, "Effects of receptive listening on the congruence of speakers' response latencies in dialogues," Psychological Reports, Vol.97, pp.265-274, 2005

[8] H. Koiso, A. Shimojima, and Y. Katagiri, "Collaborative Signaling of Informational Structures by Dynamic Speech Rate," Language and Speech, Vol.41, No.3-4, pp.323-350, 1998

[9] C. Nagaoka, M. Komori, and S. Yoshikawa, "Synchrony Tendency: Interactional Synchrony and Congruence of Nonverbal Behavior in Social Interaction," Proceedings of the 2005 International Conference on Active Media Technology, pp.529-534, 2005

[10] T. Watanabe, M. Okubo, M. Nakashige, and R. Danbara, "InterActor: Speech-Driven Embodied Interactive Actor," International Journal of Human-Computer Interaction, Vol.17, No.1, pp.43-60, 2004

[11] M. Yamamoto, and T. Watanabe, "Timing Control Effects of Utterance to Communicative Actions on Embodied Interaction with a Robot and CG Character," International Journal of Human-Computer Interaction, Vol. 24, No.1, pp.87-107, 2008

[12] N. Kitaoka, M. Takeuchi, R. Nishimura, and S. Nakagawa, "Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems," Transactions of the Japanese Society for Artificial Intelligence, Vol.20, No.3, pp.220-228, 2005

[13] T. Yamamoto, Y. Kobayashi, Y. Muto, K. Takano, and Y. Miyake, "Hierarchical Timing Structure of Utterance in Human Dialogue," Proceedings of the 2008 IEEE International Conference on Systems, Man and Cybernetics (SMC 2008), pp.810-813, 2008

[14] Y. Muto, S. Takasugi, T. Yamamoto, and Y. Miyake, "Timing Control of Utterance and Gesture in Interaction between Human and Humanoid robot," Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2009), pp.1022-1028, 2009