Modeling Situation-Dependent Nonverbal Expressions for a Pair of Embodied Agent in a Dialogue Based on Conversations in TV Programs

Keita Okuuchi[1,*], Koh Kakusho[1], Takatsugu Kojima[2], and Daisuke Katagami[3]

[1] Kwansei Gakuin University, Sanda, Japan
{ccd77074,kakusho}@kwansei.ac.jp
[2] Shiga University of Medical Science, Otsu, Japan
[3] Tokyo Polytechnic University, Tokyo, Japan

Abstract. Mathematical model for controlling nonverbal expressions of a pair of embodied agents designed for presenting various information through their dialogue is discussed. Nonverbal expressions of a human during conversation with others depend on those of them as well as the situation of the conversation. The proposed model represents the relationship between nonverbal expressions of a pair of embodied agents in different situations of conversation by a constraint function, so that the nonverbal expression of each agent reproduces the characteristic of nonverbal expressions observed in human conversation with various situations in TV programs by minimizing the function.

Keywords: Embodied agent, Human-agent interaction, Nonverbal expression.

1 Introduction

It is proposed in some recent work on human-agent interaction to employ a dialogue by a pair of embodied agents for presenting information to users, similar to news or talk shows in TV[1][2]. Information presentation based on a dialogue is expected to make the point of the information easier to understand than that based on a monologue.

For realizing dialogues between embodied agents with nonverbal expressions, we need to consider how to maintain consistency between those nonverbal expressions, such as each agent should nod and smile when the other agent speaks with a smile. Aiming to maintain the consistency, we have proposed a mathematical model that represents the consistency by referring to social psychological studies about the relationship between nonverbal expressions of humans in a dialogue[3]. However, the degree of each nonverbal expression is not always the same but could differ with changing situations in the dialogue.

In this article, we discuss which kinds of situations actually affect the degree of nonverbal expressions in what manner, by analyzing human dialogues in TV programs, in order to propose a further extension of the mathematical model, which approximates the dependency of nonverbal expressions on those kinds of situations.

* Corresponding author.

M. Kurosu (Ed.): Human-Computer Interaction, Part IV, HCII 2013, LNCS 8007, pp. 340–347, 2013.

2 Dependency of Nonverbal Expressions on Changing Situations in a Dialogue

2.1 Relationship between Nonverbal Expressions

In the previous work on social psychology, it is known that positive correlations are observed between the amount of speech and the degree of nonverbal expressions, which include gaze, smile and nod, of humans during conversation; when the amount of speech or the degree of one of the nonverbal expressions increases or decreases, the degree of another nonverbal expression increases or decreases in a similar manner[4][5][6][7]. These positive correlations are summarized in Table 1, where each pair denoted by #1-12 is reported to show the positive correlation for their amounts in human conversation as described above.

Table 1. Correlations of nonverbal expressions

(a) Between different persons

		PersonA			
		speech	gaze	smile	nod
PersonB	speech		#4	#6	#8
	gaze	#1	#5		
	smile	#2		#7	
	nod	#3			

(b) Within the same person

	gaze	nod	smile	hand gesture
speech	#9	#10	#11	#12

2.2 Factors for Classifying Situations of Human Conversation in TV Programs

Since the above report of the previous work only describes the general tendency in the appearance of nonverbal expressions, the degree of each nonverbal expression in actual conversation could differ with various kinds of situations changing during the same dialogue while satisfying those correlations qualitatively. In order to learn about most influential situations especially in dialogues for information presentation, we made some interviews for 18 participants.

The participants are asked to watch some news and talk shows in TV while paying their attentions to the moments when they recognize the situation of the conversation changes. For each moment that they recognize the change, they are asked about the factor that cause them to recognize the change.

The result is shown in Figure 1. In this figure, the number of answers of the participants is counted by classifying the answers into three kinds: (A)role of the interlocutors (which interlocutor takes the control of the conversation), (B)structure of participation (which are the addressee of the speech by each interlocutor, the other interlocutor or the viewer), (C)atmosphere of the dialogue (humorous or serious). These three factors

are based on the viewpoint taken in the study of conversation analysis[8][9]. This result shows that most of the factors that make the participants recognize the change of situations are included in one of these three factors.

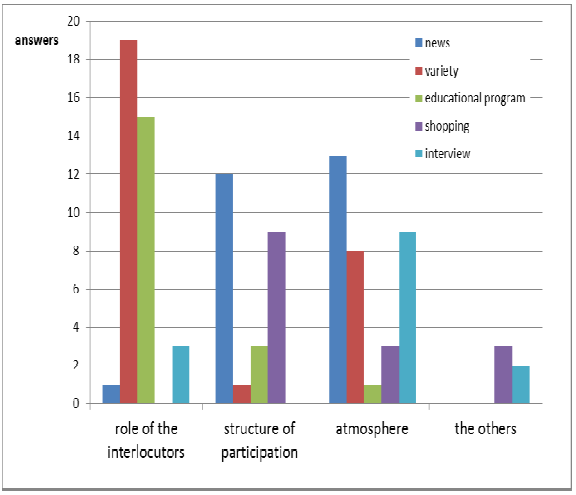


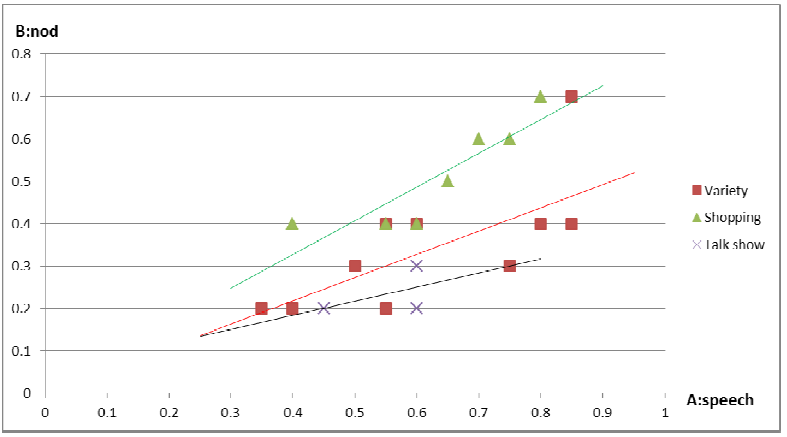
Fig. 1. Factors for recognizing different situations of TV programs

2.3 Analyzing the Dependency of Nonverbal Expressions on the Situations of Conversation

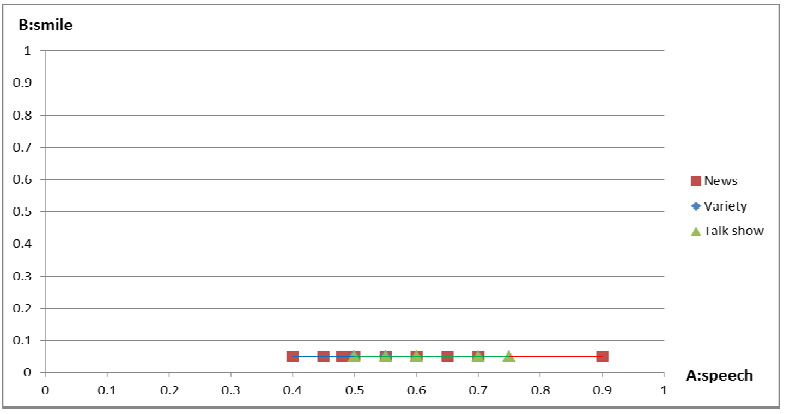
In order to analyze the dependency of the positive correlations between nonverbal expressions on situations (A)-(C) described in 2.2, we observed actual degree of each nonverbal expression in dialogues of various TV programs. If the amount of speech or the degree of a nonverbal expression of one person has positive correlation with those of the other person, the positive correlation should be approximated by a linear function describing the rate of the amount or the degrees between the positively correlated pair of speech or nonverbal expressions, and if the correlation depends on the situation of conversation, the rate should take different values for different situations.

From the result of our analysis about how each pair of the amount of speech or the degrees of nonverbal expressions, we found that some pairs can be approximated by linear functions and others cannot. Figure 2 are examples of the pairs that can be approximated by linear functions and those that cannot. For example, the degree of nod of interlocutor B has the positive correlation that can be approximated by a linear function with the amount of speech of interlocutor A for the situation of conversation with the structure of participation where the addressee of each interlocutor is the other interlocutor in humorous atmosphere, whereas the degree of smile of interlocutor B has no relation with the amount of speech of interlocutor A for the structure of participation where the addressee of each interlocutor is the other interlocutor in serious atmosphere. The pair of the amount of speech or the degrees of nonverbal expressions with positive

correlation for each situation of conversation is summarized in Table.2, where each cell with a check denotes that the corresponding pair has positive correlation.



(a) With positive correlation
("humorous" & "addressing to the other interlocutor")



(b) Without positive correlation.
("serious" & "addressing to the other interlocutor")

Fig. 2. Example of a pair of speech or nonverbal expressions with or without positive correlation for their amount or degrees

Figure 2(a) also shows, from the slopes of the lines, that the same pair of the amount of speech or the degrees of nonverbal expressions takes different ratios for their positive correlation relationship for different kinds of programs. For example, the degree of nod of interlocutor B increases more rapidly with the increase of the amount of speech of interlocutor A in shopping programs than in variety shows or talk shows for the situation of conversation with the structure of participation where the addressee of each interlocutor is the other interlocutor in humorous atmosphere.

Table 2. Correlation between nonverbal expressions in different situations

situation	Ⅰ	Ⅱ	Ⅲ	Ⅳ
(B)addressee	intelocutors	intelocutors	viewers	viewers
(C)atmosphere	humorous	serious	humorous	serious
#1	✓	✓		
#2	✓		✓	
#3	✓	✓	✓	✓
#4	✓	✓		
#5	✓	✓	✓	✓
#6	✓		✓	
#7	✓	✓	✓	✓
#8	✓	✓	✓	✓
#9	✓	✓		
#10	✓	✓	✓	✓
#11	✓		✓	
#12	✓	✓	✓	✓

3 Mathematical Model for Situation–Dependent Correlations

In order to reproduce the situation-dependent positive correlation between nonverbal expressions in TV programs described in section 2 by a pair of embodied agent in a dialogue following a predetermined scenario for information presentation, we represent the positive correlation by a constraint function.

First of all, we represent the basic relationship of positive correlation in the amount of speech or the degrees of nonverbal expressions for pairs #1-12 by constraint function E as follows:

$$E \equiv \sum_{i=1}^{16} E_i = \sum_{i=1}^{16} (x_i^X - y_i^Y) = 0 \tag{1}$$

where x_i^X and y_i^Y denote the amount of speech or the degrees of the nonverbal expressions for pair #*i* (i=1,...,12) to be produced by embodied agents *X* and *Y*, which are variables to denote one of the two embodied agents A and B ($X, Y \in \{A, B\}$).

As already shown in Table 2, it depends on the situation of communication whether the positive correlation is actually observed or not for each pair #1-12. In order to represent the situation-dependency for the existence of positive correlation for each pair #1-12, we modify equation (1) by introducing an additional variable that represents the existence of positive correlation for each pair as follows:

$$E' \equiv \sum_{i=1}^{16} l_i E_i = \sum_{i=1}^{16} l_i (x_i^X - y_i^Y) = 0 \tag{2}$$

where variable l_i denotes the binary flag that represents the existence of the positive correlation between the amount of speech or the degrees of nonverbal expressions denoted by pair #*i*. This value is predetermined based on Table.2.

Moreover, as already discussed in section 2.2, the ratio between the values for the amount of speech or the degrees of nonverbal expressions depends on the kind of TV program. In order to represent this dependency, we modify equation (2) by introducing another additional variable that represents the ratio as follows:

$$E'' \equiv \sum_{i=1}^{16} E_i' = \sum_{i=1}^{16} (x_i^X - \alpha_i y_i^Y) = 0 \tag{3}$$

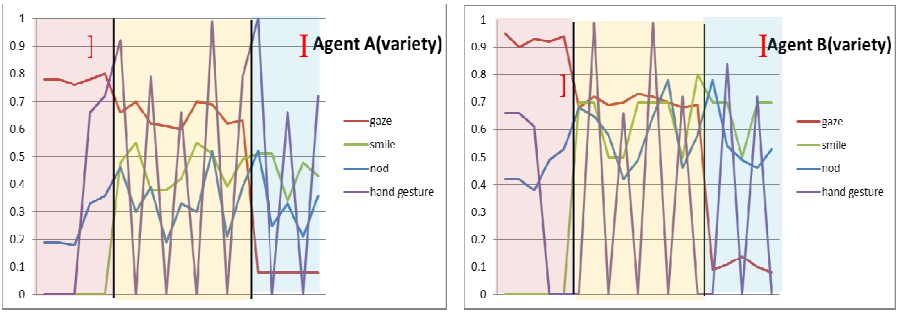
where variable α_i denotes the ratio between the value for the amount of speech or the degrees of nonverbal expressions for pair #i. This value is estimated in advance so that function E'' is minimized when the degrees of nonverbal expressions observed in each situation of actual TV programs are given as the value of x_i^X and y_i^Y.

For controlling nonverbal expressions of embodied agents A and B in each situation of conversation, the values of x_i^X and y_i^Y for each moment of the dialogue are calculated by minimizing function E'' by setting the degree of the speech of each embodied agent at the moment based on the scenario of the dialogue, after the values of l_i and α_i for the situation are obtained as described above.

4 Experimental Results

The degrees of nonverbal expressions is calculated for different situation of conversation by minimizing constraint function E" in equation (3) after setting the amount of speech of interlocutors of actual TV program used in section 2 for the values of variables x_i^X and y_i^Y corresponding to speech. The resultant values for the degrees of nonverbal expressions for a variety show are shown in Figure 3. As shown in the figure, the tendency for degrees of nonverbal expressions changes with situations of communication I – III in Table.2. The reason why the degrees of each nonverbal expression are not the same even in the same situation is because it is also affected by the amount of speech at each moment.

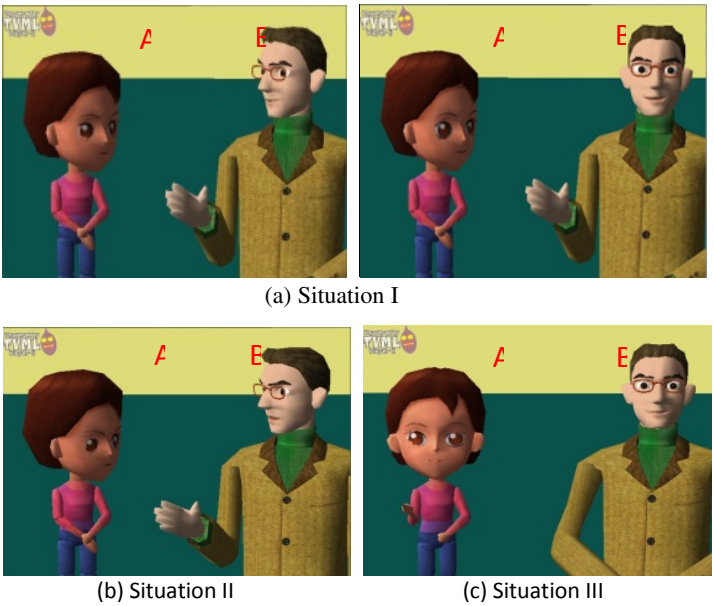
In order to realize a dialogue by a pair of embodied agents with speech and nonverbal expressions obtained above, we employed TVML (TV program Making Language)[10], which is developed by NHK (Japan Broadcasting Cooperation) for producing TV program by a script language. Sample scenes in situations I - III are shown in Figure 4. As shown in the figure, the degrees of nonverbal expressions of each embodied agent are controlled depending on changing situations of the dialogue. For example, the agents gaze at each other in the situation with the structure of participation where the addressee of each agent is the other agent (situation I, II), whereas they gaze at the viewer in the situation where the addressee is the viewer (situation III). Also, the smile of the agents increases for humorous atmosphere (situation I, III), whereas it becomes zero for serious atmosphere (situation II). In addition, the



(a) Agent A (b) Agent B

Fig. 3. The degrees of nonverbal expression reproduced by the proposed model for different situations of communication

agents do not gaze at each other all the time but also gaze sometimes at the viewer even in the situation with the structure of participation where the addressee of each agent is the other agent (situation I).



(a) Situation I

(b) Situation II (c) Situation III

Fig. 4. Sample scenes of the dialogue by the embodied agents in different situations

5 Conclusion

We proposed a mathematical model that approximates situation-dependency of positive correlation between nonverbal expressions in human conversation in TV program for a pair of embodied agent in a dialogue for information presentation to users. As shown in the experimental results, our model reproduces the characteristic of nonverbal expressions in human conversation depending on the situation of the structure of participation and atmosphere. However, for increasing the precision for reproducing the characteristic of nonverbal expression in human conversation, we need to consider further improvement of our model as one of our future steps.

In our model, the degree of nonverbal expressions of embodied agents are driven by the speech, which is given as the script for their dialogue, based on its constraint of on the nonverbal expressions represented by the constraint function. This constraint only employs the amount of speech without consideration of the content of the speech. However, the nonverbal expression in human conversation should also depend on the content of the speech. In order to extend our model towards the capability for considering the dependency of nonverbal expression on the content of the speech, we plan to include some tags that represent various situations dependent on the content of the speech in the script of the dialogue of the agents.

References

1. Kubota, H., Yamashita, K., Nishida, T.: Conversational Contents Making a Comment Automatically. In: Int. Conf. on Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies (KES), pp. 1326–1330 (2002)
2. Takahashi, T., Katagami, D.: Agent Design As An Information Interface Assumed For Full-time Operation. In: Human-Agent Interaction Symposium (2011) (in Japanese)
3. Kakusho, K., Itou, J., Minoh, M.: An Embodied Agent that Sends Nonverbal Conversational Signals Consistent with those of the Partner during a Dialogue. In: IEEE Workshop on Robot and Human Interactive Communication (RO-MAN), pp. 247–252 (2003)
4. Beattie, G.W.: Sequential Temporal Patterns of Speech and Gaze in Dialogue, Semiotica (1978)
5. Dimberg, U.: Facial Reactions to Facial Expressions, psychophysiology. psychophysiology 19(6), 643–647 (1982)
6. Kendon, A.: Some functions of gaze direction in social interaction. Acta Psychologica 26, 22–63 (1967)
7. Matarazzo, J.D., Saslow, G., Wiens, A.N., Weitman, M., Allen, B.V.: Interviewer Head Nodding and Interviewee Speech Durations. Psychotberapy: Theory, Research and Practice 1, 54–63 (1964)
8. Clark, H.H.: Using Language. Cambridge University Press (1996)
9. Gatica-Perez, D.: Automatic nonverbal analysis of social interaction in small groups. Journal Image and Vision Computing 27(12) (2009)
10. `http://www.nhk.or.jp/strl/tvml/`