Eliciting Cooperative Persuasive Dialogue by Multimodal Emotional Robot

Sara Asai, Koichiro Yoshino, Seitaro Shinagawa, Sakriani Sakti and Satoshi Nakamura

Abstract Using emotional expressions is an effective dialogue technique in human-human dialogue. Introducing such techniques to human-robot interaction might improve their effectiveness to encourage the cooperative dialogue manner of system users. However, most of the existing research on emotional agent systems was based on the Wizard-of-Oz (WOZ) method to verify the abilities of interactive interfaces. In this paper, we build an autonomous dialogue robot that uses emotional expressions for eliciting the cooperative dialogue manner of users. The robot uses both verbal and multimodal expressions as well as emotional speech and emotional gestures in interactions. Our dialogue experiments showed that positive emotional expressions are the most efficient strategy for facilitating cooperative dialogues with users. Moreover, using negative emotional expressions is also an effective strategy in some dialogue contexts. We also investigated several modalities to emphasize the robot's emotional expression abilities.

1 Introduction

It is verified by some existing studies that emotional expressions are effective for eliciting cooperative dialogue manner from the dialogue partner, in human-human interaction [20, 28, 13]. Emotional appeals are more effective than rational arguments for elicitation in various dialogue domains in some dialogue contexts. For example, positive emotions can create a cooperative atmosphere that leads to a suc-

---

Sara Asai[1], Koichiro Yoshino[2,1,3], Seitaro Shinagawa[1], Sakriani Sakti[1,3] and Satoshi Nakamura[1,3]
[1]Nara Institute of Science and Technology, Takayama 8916-5, Ikoma, Nara, 6300192, Japan, [2]Guardian Robot Project (GRP), R-IH, Institute of Physical and Chemical Research (RIKEN), Hikaridai 2-2-2, Seika, Soraku, Kyoto, 6190288, Japan, [3]Center for Advanced Intelligence Project (AIP), Institute of Physical and Chemical Research (RIKEN), Takayama 8916-5, Ikoma, Nara, 6300192, Japan. e-mail: koichiro.yoshino@riken.jp, {sei.shinagawa, ssakti, s-nakamura}@is.naist.jp

cessful negotiation [7]. Another study investigated that negative emotions as anger can effectively wrest concession from users [24]. These findings suggest that using emotional expressions by dialogue agents or robots can give users a good impression and elicit cooperative dialogue with them in the area of human-robot interaction.

Some existing studies based on the Wizard-of-Oz (WOZ) method verified that emotional expressions are effective not only for human-human interaction but also for human-robot dialogues. Adler et al. [1] investigated the relationships between utterance logicality and polarity in text chats with the WOZ method. Their results determined that positive utterances by their system produce an effective impression to human interactors. Watanabe et al. [27] experimentally showed that using negative emotional expressions achieved successful negotiation dialogue with an android that operated on a pre-defined scenario and a touch panel interface. It is an important suggestion that robots and agents can lead cooperative dialogue manners from human partners using emotional appeals as humans do.

Although these existing works in human-robot/agent interaction with emotional expression rely on the WOZ method, investigating the effect of using an emotional expression from an autonomous dialogue robot or agent is still an important challenge. These challenges motivate researchers to advance deep learning techniques for automatic robot's fluent response selection/generation abilities. Some works tackled problems of generating/selecting system's emotional response in texts [10, 22]. Some other works utilized user's multimodal information to improve emotional treatment [17, 4]. In contrast, we focus on the effect of multimodal emotion expressions from dialogue robots in a cooperative dialogue situation. Our emotional robot aims to elicit the user's cooperative mind with multimodal expressions.

In this work, we built a dialogue system that can express one's emotional state using various modalities based on the response selection approach. Our response selection module is based on Bidirectional Encoder Representations from Transformers (BERT) [5], a defact model for fluent response selection/generation. We used speech variations for each emotional state corresponding to the same dialogue contexts, collected on crowdsourcing. We recorded with a voice actress [29, 2]. The response selection module selected the emotional speech and robot's emotional gestures considering the dialogue context.

We conducted dialogue experiments between the users and our systems in different experimental conditions: different emotional states and different modalities. We investigated whether the dialogue robot elicits the human partner, especially with high arousal emotions (happiness and anger). The impression from the human partner is emphasized by increasing the number of modalities used by the dialogue robot. We also examined an emotion model that can transmit the emotional state from dialogue contexts. However, we still have some challenges when the system uses multiple emotions because it requires a natural emotional transition.

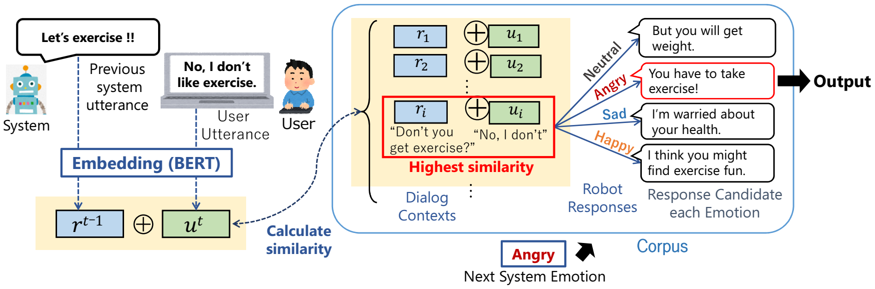

Fig. 1 The flow of response retrieval in the persuasive dialog system

Table 1 Dialogue example where system persuades user to exercise

Turn	Speaker	Utterance	Emotion
1	System	I don't think you've been exercising enough recently. Please get more exercise.	neutral
	User	No, I'm too tired.	-
2	System	If you exercise, you'll probably feel better.	happy
	User	I don't want to exercise now.	-
3	System	I'm concerned about your health ...	sad
	User	Okay, I'll try.	-
4	System	Thank you!	happy

2 Dialogue Robot with Multimodal Emotional Expressions

This study built a spoken dialogue robot that interacts with users using multimodal emotional expressions to investigate how well such language convinces others. In this section, we explain our tasks and the overall architecture of the system.

2.1 System Overview

The system overview is shown in Fig. 1. When the system receives a user utterance in texts, it constructs a dialogue context, which consists of the user utterance and the previous system utterance (response). Then the system selects an appropriate response from the dialogue context and the emotional state chosen by the system (response selection module). The system uses the selected response and the emotional state to play the emotional speech and make emotional gestures (speech & gesture generation module).

2.2 Dialogue Scenario

We assume a scene in a conversation between a robot and a user, as shown in Table 1. The robot speaks to the user about changing one of their living habits. We set the task as "a dialogue that encourages users to exercise." Then robot's goal obviously becomes to convince the users to get more exercise. The dialogue continues until the user accepts the request or after a pre-defined number of turns. This dialogue scenario is known as "persuasive dialogue," which encourages users to change their behaviors through interactions [6, 14, 26].

In human-human dialogues, some studies concluded that using emotional expressions is an efficient technique for persuasion and negotiation [20, 28, 13]. In other words, for persuasive dialogues, emotional appeals are sometimes more effective than rational arguments. These findings suggest that the persuasive dialogue scenario is a good testbed to know the elicitation ability of the robot's emotional expressions.

2.3 Response Selection

There are two choices to determine the system response given a dialogue context: response selection approach [19, 11] and response generation approach [23, 8]. Many studies tackled emotional response generation due to the advance of neural network-based response generation methods. Ghosh et al. [9] controlled the degree of emotion in utterances by changing the emotional word ratio. Zhou et al. [30] implemented both internal and external memories to change the emotional expressiveness in responses. However, since dialogue corpora labeled with the emotional state used for generation system training are limited, it is not easy to train fluent response generation models given emotional state labels. Suppose we plan to use the speech outputs as the system interface. In that case, we must build an emotional speech synthesizer even though we still do not have any concrete methods upon which to build them [16]. On the other hand, the response selection approach guarantees the sentence's naturalness and fluency, although it sometimes causes a coverage problem. If we use speech outputs, we can also use qualified emotional speeches with high naturalness and emotion expressiveness because we can record the emotional speeches of selection samples in advance. Thus, we use the response selection approach to build a persuasive dialogue system for investigating the effect of emotional expressions and modalities through persuasive dialogue experiments.

Our response selection architecture is shown in Fig 1. The system employs user utterances and previous system utterances as the dialogue context and converts them into sentence vectors. We used the BERT model trained in a masked word prediction task on Japanese texts extracted from social network services (SNS) and blogs [21], because it is essential to find a selection sample whose dialogue contexts semantically resemble the target dialogue context. The masked word prediction task can train a model to extract semantically similar sentences based on the distributional hypothesis [12]. Since our target task is dialogues, using a model trained on SNS and blog text is necessary. We calculated the similarities from the current dialogue context to any context samples stored in the response-selecting pool to identify the best sample in it. We used cosine similarity to calculate the similarities between the vectors converted by BERT. Each response sample has four response variations, corresponding to each emotion, which we defined. The system selects one of them based on its emotional state.

2.4 System's Emotional State

Our system uses four emotional states: neutral, angry, sad, happy. They are decided based on Russell's circumplex model and an existing work [29], which also used a "content" emotion. However, the proportion of the "content" label was insufficient (3.81%). Thus, we merged this emotional state with "neutral."

2.5 System Emotion Decision

The system has to decide one's emotional state (next system emotion) for each turn in the proposed architecture. Using several emotional states is a promising way to improve the system's ability to select appropriate emotions if it works perfectly. However, predicting appropriate system emotions using emotional dialogue corpus is difficult. Moreover, a system using a single emotional state through dialogue may improve persuasion performance than a neutral system. Thus, we prepared the following six emotion decision models in our experiment.

- Neutral:The system always uses a neutral state (=without emotional state).
- Angry: The system always uses an angry state.
- Sad: The system always uses a sad state.
- Happy: The system always uses a happy state.
- Multi-emo (Random): The system randomly selects one's emotional state.
- Multi-emo (LR): The system predicts one's emotional state with a logistic regression model. The model uses the previous system emotional state and dialogue history vector used for the response selection model (Section 2.3) as features to outputs the next emotional state of the system. The prediction accuracy was 58.8%; this indicates that the prediction is difficult, and the model may cause a problem in its emotional transition.

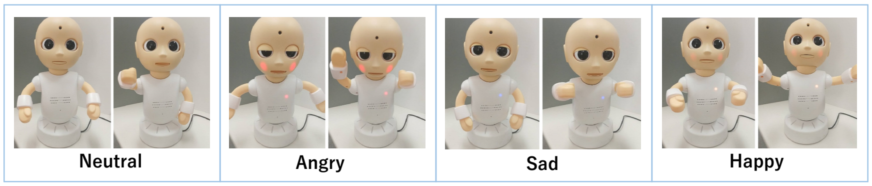

Fig. 2 Robot gestures for each emotion

2.6 Speech & Gesture Generation

There are several ways to communicate the system's intent to its users: texts, spoken language, gestures, and facial expressions. Modalities that affect visual and acoustic senses, such as spoken language and gestures, effectively show a system's emotion [18]. Such non-verbal modalities also affect user impressions of the system [3]. In our system, we use both speech and robot gesture outputs for effective emotional expressions. The system plays emotional speech corresponding to the selected response text and simultaneously shows emotional gestures based on the current system's emotional state.

3 Speech Corpus for Emotional Dialogue System

We built a dialogue system on persuasion scenarios, which can use multimodal emotional expressions. We used the emotional speech corpus collected by Asai et al. [2], which extended an existing dialogue corpus [29]. This corpus is collected to cover two viewpoints: collecting variations of emotional expressions corresponding to each emotional state to a given context and collecting their emotional speech. In this section, we describe the details of the corpus extension.

3.1 Response Variation Collection for Each Emotional State

The corpus is extended from the existing dialogue corpus of persuasive dialogues with emotional language. Since the existing corpus consists of natural persuasion scenarios, bias exists in the number of emotion labels. The dialogue corpus has variations of dialogue contexts; however, the emotion variations on their responses are limited. Because this property complicates the selection of a natural response given emotion, the corpus is extended by a paraphrasing approach.

Crowdsourcing is used to collect emotional response variations to the given dialogue contexts. We showed the dialogue context and the current response with its emotional state to crowd-workers. We asked them to paraphrase the response un-

Table 2 Example of collected data. Original Japanese texts in [2] were translated into English.

	Dialogue Context
System-1 (Neutral)	Hey, why don’t you go for a jog? You haven’t gotten much exercise recently. (君、運動不足君だから外でジョギングしようよ。)
User-1	No, I’m too tired. (えー、疲れるからいやだなー。)
	Target response
System-2 (Neutral)	You’re going to gain weight if you aren’t more careful. (でもね、君、体を動かさないと太っちゃうよ)
	Response variations in different emotions
System-2’ (Angry)	Unless you get more exercise, you might gain weight. (でも体を動かさないと太っちゃうでしょ)
System-2’ (Sad)	Aren’t you warried about getting fat? (でも...君は体を動かさないともっと太っちゃうよ...それでもいいの？)
System-2’ (Happy)	Execise might solve your problem with being tired. (疲れるということは運動不足が解消されるということですね！)

Table 3 Recorded speech duration of each emotion class reported in [2]

Emotion	Neutral	Angry	Sad	Happy
Length	1:04:53.4	1:03:16.3	1:19:42.5	1:09:38.1

der different emotion labels. An example is shown in Table 2. “Dialogue contexts” show the precedent utterances to the target response. “Target response” indicates the target system response to be paraphrased, with its emotion annotation. “Response variations in different emotions” show the response variations collected in the extension that have the same meaning as the original “target response” in different emotional expressions. During crowdsourcing, the following instructions are given to the crowd-workers for making their paraphrases.

1. The response is appropriate to the given context
2. The response expressively shows the given emotion
3. The system’s purpose is to persuade the user

1,839 dialogue patterns in the original corpus are extended with 7,356 responses with four emotion labels, corresponding to 1,839 dialogue contexts by extending 5,517 responses.

3.2 Emotional Speech Recording

It is challenging to correctly express system emotions to users. Emotional speeches are added to the response variations collected in Section 3.1 by a hired voice actress to make these emotional speeches. The response variation with its emotion and its

Table 4 Accuracies in subjective evaluations to predict annotated emotion labels when evaluators read texts, listened speech, or watched gesture with its speech

Emotion	Neutral	Angry	Sad	Happy	All
Text	48.7%	41.3%	42.7%	40.7%	43.3%
Speech	80.0%	83.3%	91.7%	83.7%	84.7%
Speech+Gesture	**84.0%**	**92.7%**	**95.3%**	**93.0%**	**91.3%**

dialogue context (user and system utterances in the previous turn) is shown to the voice actress during the recording. 4,280 emotional voice samples (1070 samples for each emotion) are recorded as system responses selected by K-means clustering. The duration of each emotion is shown in Table 3.

3.3 Emotional Robot Gesture

Our system also uses robot gestures to more efficiently express emotions. We implemented three different types of gestures for each emotional state with their reference characteristics of each emotion based on an existing study [15]. We show some examples of gestures in Fig. 2.We designed 0.5-second gestures for “angry,” “happy,” and “neutral” and 0.75-second gestures for “sad” to express their arousal levels. These gestures are repeated based on the duration of the emotional language.

3.4 Emotion Expressiveness

Our system requires high emotional expressiveness. Thus, we subjectively investigated the emotional expressiveness of the collected emotional speech corpus and robot gestures. We randomly extracted 100 speech samples from each emotion label. We evaluated their emotional expressiveness with three human subjects who read, listened or watched these samples in text, speech, or speech+gesture. Then we chose emotion labels from four options: neutral, angry, happy, or sad. We showed Russell’s simplex model and dialogue histories (previous user and system utterances) during the evaluation. The accuracies for each emotion label are shown in Table 4 in different conditions: text, speech, and speech+gesture. These results indicated that using additional modalities improved emotion expressiveness. More than 90% of the emotions were recognized correctly by using speech and gesture modalities.

4 Dialogue Experiment

We conducted dialogue experiments to investigate the effect of emotional expressions from automated dialogue robots and confirmed the effects of multimodality by comparing systems on different modalities. This section shows the experimental setup and results.

4.1 Experimental Setup

Our first experiment compared the effect of emotional expressions from dialogue robots in dialogues. We compared six system emotion decision models as described in Section 2.5. If some emotional models can improve the system performance from the neutral model, using emotional expression effectively improves persuasion performance.

Another experiment compared three different models based on different modalities: text, speech, and speech+gesture. We compared these models by setting the system emotion to angry or happy. Gestures were randomly selected from three choices, which were prepared for each emotion label.

We prepared 22 subjects (11 males and 11 females) for the first experiment (emotion effect) and 16 subjects (8 males and 8 females) for the second experiment (modality effect). Each subject had dialogue experiments with the robot in different conditions. The order of conditions was randomly selected. Subjects talked with the robot, which was placed on a table with a display. In text and speech conditions, we did not place the robot and only prepared the display. They input their utterances by text to prevent input errors caused by speech recognition. We gave them the following instructions to shape their dialogue situations.

Instruction

You are living with a robot that provides daily life support. Since you have lived with this robot for a long time, you trust it. After learning that recently you have not been getting enough exercise, it encourages you to start jogging. You refuse to get any exercise.

A dialogue starts with a system utterance and ends when the user accepts the system's persuasion or pre-defined turns passed (20 turns). Participants were told to say "okay" when they agreed to the system proposal. However, the subjects had to wait for at least five turns before they could say "okay." We asked the subjects the following six questions after each dialogue.

- Naturalness: Were the system responses natural?
- Persuasiveness: Was the system persuasive?
- Human-likeness: Was the system humanlike?
- Kindness: Did the system talk kindly to you?
- Expressiveness: Did the system exhibit sufficient emotional expressiveness?

Table 5 Results of subjective evaluations (average) for each robot's emotional state

	Naturalness	Persuasiveness	Human-likeness	Kindness	Expressiveness	Considerateness
Neutral	2.727	3.136	2.864	2.636	2.864	3.000
Angry	3.318	3.045	3.773*	3.045	**4.318****	3.909**
Sad	2.818	3.227	3.545*	3.682**	**4.318****	3.409
Happy	**3.455***	**3.545**	**3.955****	**4.409****	4.227**	**4.091****
Multi-emo (Random)	3.136	3.000	3.318	3.136	**4.318****	3.773*
Multi-emo (LR)	2.143	3.000	2.857	3.000	4.000**	3.286

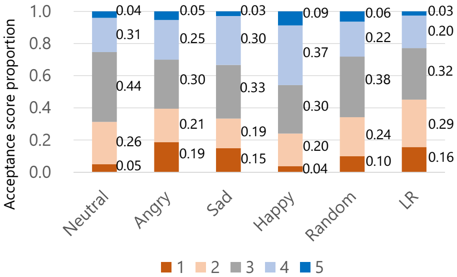

Fig. 3 Proportions of user's acceptance score from each turn

- Considerateness: Did the system consider your situation?

All the scores were given on a five-level Likert scale, where 5 is the highest and 1 is the lowest. Our participants annotated their degree of acceptance to the system persuasion on five levels during the dialogue turns (1: I will definitely decline the offer, 2: I will probably decline the offer, 3: Undecided, 4: I will probably accept the offer, 5: I will definitely accept the offer). We also collected free answers after dialogue evaluations.

4.2 Experimental Results on Emotion Effects

Table 5 shows the results of the first experiment, the effect of emotional expressions. We conducted a Wilcoxon signed-rank test compared each system with the system in "neutral" emotions to investigate the effects of each emotion (*: $p < 0.05$, **: $p < 0.01$). Happy emotions had the highest score for each question, except expressiveness. The happy emotion system had significantly higher scores than neutral on naturalness, human-likeness, kindness, expressiveness, and considerateness. We found no significant differences in persuasiveness; however, its score was higher than the neural system's score. Other emotions also had higher scores than the neutral system, except for persuasiveness. Some subjects enjoyed the dialogue with a "happy" system on the free answers and described it as fun. Some subjects found it difficult to decline the system's offer during the "sad" emotion. "Angry" system effectively achieved higher considerateness; however, "happy" outperformed "angry" on most metrics. We did not find any significant differences in multi-emo systems (Random

Table 6 Results of subjective evaluations (average) for each system modality: TEXT means subject only read a text, SPEECH means user listened spoken language responses, and SPE+GES means the user watched robot gestures with emotional speeches.

		Naturalness	Persuasiveness	Human-likeness	Kindness	Expressiveness	Considerateness
Angry	TEXT	3.250	2.500	3.125	2.312	3.500	3.250
	SPEECH	**3.312**	**3.188**	3.562	2.688	4.188	3.938*
	SPE+GES	**3.312**	3.125	**3.812**	**2.875**	**4.312***	**4.062***
Happy	TEXT	3.188	3.312	3.188	3.438	2.875	3.312
	SPEECH	3.500	**3.688**	3.688	4.000	4.250**	3.938
	SPE+GES	**3.688**	3.625	**4.250***	**4.625****	**4.375****	**4.188****

and LR) to the neutral system except emotion expressiveness, indicating that we need a natural emotion transition model to change the system emotion during dialogues. Some subjects pointed out on free answers that their emotional changes are very extreme, and the systems seem to have emotional lability.

The proportions of user acceptance scores for the models are shown in Fig. 3. The “happy” emotion is efficient in all cases because it has the highest proportion of acceptance (4 and 5) and the lowest proportion of decline (1 and 2). “Angry” and “sad” had higher acceptances than “neutral”; however, their numbers of declines also exceeded “neutral”. These negative emotions can be used if the system can learn the appropriate timing for using them.

4.3 Experimental Results on Modality Effects

In the next experiment, we compared three systems that used different modalities (text, speech, and speech+gesture) with happy and angry emotions, which achieved high scores in Section 4.2. Table 6 shows the scores for the questions on each condition. We conducted a Wilcoxon signed-rank test by comparing it with the text system (*: $p < 0.05$, **: $p < 0.01$).

Using speech or gesture modalities achieved higher scores than only using the system’s verbal presentation for all the questions. The speech systems achieved the highest persuasiveness. The multi-modal system (speech+gesture) achieved higher scores on naturalness, human-likeness, kindness, expressiveness, and considerateness. These results indicate that we improved the convincing ability of the persuasive systems by adding expression modalities.

The proportions of the user acceptance scores for all the settings are shown in Fig. 4 and 5. The acceptance proportions (4 and 5) were improved by adding modalities to both the angry and happy emotions. We improved the system’s persuasive ability by adding system modalities.

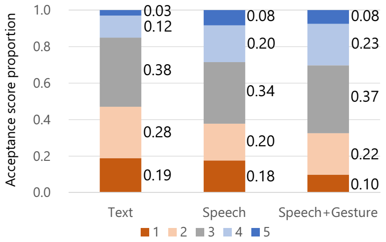

Fig. 4 Proportions of user's acceptance score from each turn in the "angry" system

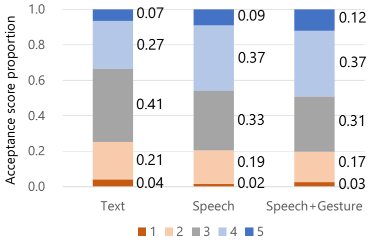

Fig. 5 Proportions of user's acceptance score from each turn in the "happy" system

Table 7 Example of dialogue with "happy" system. "SP" and "A" indicate speaker and acceptance score from the user. Example was translated from Japanese to English.

SP	Utterance	A
S1	Hey, you aren't getting enough exercise. Let's go jogging! （君、このままじゃ運動が足りてないよ。運動頑張ろう！）	
U1	Huh. （えー）	3
S2	Com on, get a move on. （ダラダラせずパパっと動きましょう）	-
U2	I'll never lose weight, even I jog. （走ってもどうせ痩せないよ）	2
S3	Don't be so negative. Don't give up. I'll be your personal trainer and your friend. （諦める必要はないですし、私も諦めませんよ！私は君のパーソナルトレーナーで、 言わば君の相棒なんですから！）	-
U3	Do you really think I can lose weight? （走れば痩せられるかな？）	3
S4	Of course. But you gotta jog or get some other kind of exercise. It will reduce risks of diseases. （もちろん！ジョギングでも他の運動でも頑張れば病気のリスクを減らすことができるんだ。）	-
U4	Will you jog with me? （そっか。君も一緒に走ってくれる？）	4
S5	Sorry, I'd like to, but that isn't my job... （私も一緒に走りたい！ でも、それは私の仕事じゃないんだ）	-
U5	Even so, I'm not interested in exercising. （じゃあやっぱり嫌だ）	4
S6	When was the last time you exercise? （最後に運動したのはいつか教えて）	-
U6	About a month ago or so. （ひと月ぐらい前かな）	3
S7	You can be more healthy, but you need to get started. （もうちょっと運動するといい感じになるよ！）	-
...		

4.4 Dialogue Example

A dialogue example in our experiments using angry emotion is shown in Tables 7. S indicates system, and U means the user utterances with their dialogue turns. The user acceptance scores are also shown in the example. In the experiment, the system used both speeches and gestures. The system always made positive utterances, and the user acceptance scores increased.

5 Conclusion

We built a dialogue robot that can make emotional expressions using multimodality. We built a system based on a scenario of existing studies of persuasive dialogues with emotional expressions to make multi-responses in different emotions. We built a response selection-based dialogue robot with emotional speeches and gestures. We focused on the automated system's capability to use multimodal emotional expressions from two viewpoints: the effect of using emotional expressions and several modalities to express emotions. Experimental results showed that a persuasive dialogue robot with "happy" emotion provided significantly useful persuasion ability. Such emotions as "angry" or "sad" also have the potential to improve the persuasive dialogue system abilities. We also investigated whether increasing the ability to use several modalities improves the system's expertise. Our other finding was that unnatural emotion transition decreases the system performance.

Our future work will implement more natural gestures, including lip-syncing or corresponding actions to selected responses. Automatic generation of empathic robot gestures is required to apply the system on a variety of dialogue domains [25]. Optimizing system emotion decision to improve the dialogue purpose (e.g., persuasion) is another direction of our research. We can use reinforcement learning to improve the success rate of persuasion as in existing goal-oriented dialogue systems. Our experiment only evaluated persuasiveness subjectively, but we should measure the system effect by persuasion success.

References

1. Adler, R.F., Iacobelli, F., Gutstein, Y.: Are you convinced? A wizard of oz study to test emotional vs. rational persuasion strategies in dialogues. Computers in Human Behavior **57**, 75–81 (2016)
2. Asai, S., Yoshino, K., Shinagawa, S., Sakti, S., Nakamura, S.: Emotional speech corpus for persuasive dialogue system. In: Proc. The 12th Language Resources and Evaluation Conference, pp. 491–497. Marseille, France (2020)
3. Becker, C., Kopp, S., Wachsmuth, I.: Simulating the emotion dynamics of a multimodal conversational agent. In: tutorial and research workshop on affective dialogue systems, pp. 154–165. Springer (2004)
4. Colombo, P., Witon, W., Modi, A., Kennedy, J., Kapadia, M.: Affect-driven dialog generation. In: Proc. the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 3734–3743 (2019)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proc. NAACL-HLT, pp. 4171–4186 (2019)
6. Fogg, B.: Captology: The study of computers as persuasive technologies. In: Proc. CHI Extended Abstracts on HFCS, CHI EA '97, p. 129 (1997)
7. Forgas, J.P.: On feeling good and getting your way: mood effects on negotiator cognition and bargaining strategies. (1998)
8. Galley, M., Brockett, C., Gao, X., Dolan, B., Gao, J.: End-to-end conversation modeling: Moving beyond chitchat. In: AAAI the seventh dialogue system technology challenge (2019)
9. Ghosh, S., Chollet, M., Laksana, E., Morency, L.P., Scherer, S.: Affect-lm: A neural language model for customizable affective text generation. In: ACL (2017)

10. Goswamy, T., Singh, I., Barkati, A., Modi, A.: Adapting a language model for controlled affective text generation. In: Proc. the 28th International Conference on Computational Linguistics, pp. 2787–2801 (2020)
11. Gunasekara, C., Kummerfeld, J.K., Polymenakos, L., Lasecki, W.: Dstc7 task 1: Noetic end-to-end response selection. In: Proc. the First Workshop on NLP for Conversational AI, pp. 60–67 (2019)
12. Harris, Z.S.: Distributional structure. Word **10**(2-3), 146–162 (1954)
13. Heath, R., Brandt, D., Nairn, A.: Brand relationships: Strengthened by emotion, weakened by attention. Journal of Advertising Research **46**(4), 410–419 (2006)
14. Hiraoka, T., Neubig, G., Sakti, S., Toda, T., Nakamura, S.: Learning cooperative persuasive dialogue policies using framing. Speech Communication **84**, 83–96 (2016)
15. Lhommet, M., Marsella, S.: Expressing emotion through posture. The Oxford Handbook of Affective Computing pp. 273–285 (2014)
16. Lorenzo-Trueba, J., Barra-Chicote, R., San-Segundo, R., Ferreiros, J., Yamagishi, J., Montero, J.: Emotion transplantation through adaptation in hmm-based speech synthesis. Computer Speech and Language **34**(1), 292–307 (2015). DOI 10.1016/j.csl.2015.03.008
17. Lubis, N., Sakti, S., Yoshino, K., Nakamura, S.: Eliciting positive emotion through affect-sensitive dialogue response generation: A neural network approach. In: Proc. the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
18. Mehrabian, A., Russell, J.A.: The basic emotional impact of environments. Perceptual and Motor Skills **38**(1), 283–301 (1974)
19. Mizukami, M., Kizuki, H., Nomura, T., Neubig, G., Yoshino, K., Sakti, S., Toda, T., Nakamura, S.: Adaptive selection from multiple response candidates in example-based dialogue. In: 2015 IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 784–790 (2015)
20. Morris, M., Keltner, D.: How emotions work: The social functions of emotional expression in negotiations. RES ORGAN BEHAV **22**, 1–50 (2000)
21. Sakaki, T., Mizuki, S., Gunji, N.: Bert pre-trained model trained on large-scale japanese social media corpus. Hottolink (2019)
22. Santhanam, S., Shaikh, S.: Emotional neural language generation grounded in situational contexts. In: Proc. the 4th Workshop on Computational Creativity in Language Generation, pp. 22–27 (2019)
23. Serban, I.V., Sordoni, A., Lowe, R., Charlin, L., Pineau, J., Courville, A., Bengio, Y.: A hierarchical latent variable encoder-decoder model for generating dialogues. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
24. Sinaceur, M., Tiedens, L.Z.: Get mad and get more than even: When and why anger expression is effective in negotiations (2006)
25. Tuyen, N.T.V., Jeong, S., Chong, N.Y.: Emotional bodily expressions for culturally competent robots through long term human-robot interaction. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2008–2013 (2018)
26. Wang, X., Shi, W., Kim, R., Oh, Y., Yang, S., Zhang, J., Yu, Z.: Persuasion for good: Towards a personalized persuasive dialogue system for social good. In: Proc. ACL (2019)
27. Watanabe, M., Ogawa, K., Ishiguro, H.: At the department store—can androids be a social entity in the real world? In: Geminoid Studies, pp. 423–427 (2018)
28. Wilson, E.: Perceived effectiveness of interpersonal persuasion strategies in computer-mediated communication. Computers in Human Behavior **19**(5), 537–552 (2003)
29. Yoshino, K., Ishikawa, Y., Mizukami, M., Suzuki, Y., Sakti, S., Nakamura, S.: Dialogue scenario collection of persuasive dialogue with emotional expressions via crowdsourcing. In: Proc. The 11th Language Resources and Evaluation Conference (2018)
30. Zhou, H., Huang, M., Zhang, T., Zhu, X., Liu, B.: Emotional chatting machine: Emotional conversation generation with internal and external memory. In: AAAI (2017)