Toward a Virtual Companion for the Elderly: Exploring the Behaviors that Potentially Achieve Rapport in Human Communication

Sayumi Shibusawa[1], Hung-Hsuan Huang[1,*], Yugo Hayashi[2], and Kyoji Kawagoe[1]

[1] College of Information Science and Engineering, Ritsumeikan University, Noji-higashi 1-1-1, Kusatsu City, Shiga 525-8577, Japan
[2] Faculty of Library, Information and Media Science, Tsukuba University, Kasuga 1-2, Tsukuba City, Ibaraki 305-8550, Japan
hhhuang@acm.org

Abstract. The elderly who live alone are increasing rapidly in these years. For their mental health, it is reported useful to maintain their social life with others. This work is aiming to develop a companion agent who can engage long-term relationship with the elderly users. This paper presents our first step to explore the rapport occurred in a human-human communication which is considered to be essential in keeping social relationship with others. We analyzed the corpus collected in a human-human dyadic conversation experiment from three view points, the speaker (potential user), the listener, and the third person who did not participate the conversation. Encouraging results that may provide the hints of agent development were found in the analysis: the attitude of the conversation can have an influence on the speaker's mood, the mood of the speaker can be potentially observed by another person, and the third person can detect speaker's attitude.

Keywords: elderly support, active listening, conversational agent, rapport.

1 Introduction

In these past years, the population of elderly people has grown rapidly. If they do not maintain social life with others, they may feel loneliness and anxiety. For their mental health, it is reported effective to keep social relationship with others, for example, the conversation with their caregivers. There are already some non-profit organizations recruiting volunteers to engage "active listening" with the elderly. Active listening is a communication technique that the listener listens to the speaker carefully and attentively by confirming or asking for more details about what they heard. This kind of support helps to make the elderly feel cared and to relieve their anxiety and loneliness. However, due to the lack of the number of volunteers comparing to that of the elderly who are living alone, the volunteers may not be always available when they are needed. In order to

* Corresponding author.

P.L.P. Rau (Ed.): CCD/HCII 2013, Part II, LNCS 8024, pp. 147–156, 2013.

improve the effect, always-available and trustable conversational partners in enough number are demanded.

The ultimate goal of this study is the development of a virtual companion agent who can engage active listening and maintain a long-term relationship with elderly users. In order to conduct successful active listening, it is considered essential for the listener to establish the rapport from the speaker (elderly user). Rapport is a mood which a person feels the connection and harmony with another person when (s) he is engaged in a pleasant relationship with him / her, and it helps to keep long-term relationships [1, 2].

We assume that the state of rapport can be approximated by keeping positive mood in active listening dialogue. Therefore, the task of the active listener (a human volunteer or the agent) is to maintain the speaker's mood as good as possible and as long as possible. In order to do this, like a human listener, the listener agent has to observe the listener's attitude, to estimate the listener's mood from the observation, and to predict the change of listener's mood caused by his / her own behaviors both verbally and non-verbally [3]. However, estimating someone's mood and engaging this kind of mental interaction can be considered difficult or impossible even for a human, it would be more difficult for a machine, like the rapport agent. This paper presents our first step of this study, a human-human conversation experiment to validate whether it is possible to implement such an agent. The collected dialogue corpus was evaluated in the aspects of positive / negative state of attitude and mood of the subjects by themselves. The results were compared with that done by a third person who did not have premise about the subjects and prior knowledge about the dialogue contents. The third person view was used instead of an autonomous agent because they shared similar ability in evaluating the subjects. In this paper, the experiment settings and the relationship of the evaluation from the speaker, listener, and third person were reported. Base on the analysis of the corpus, we would like to implement a companion agent who automatically measures the speaker's attitude (approximated mood) and reacts to it, for the Japanese elderly users.

2 Related Works

The research works on making robots and agents to be the partners of the elderly and dementia patients have been getting popularity. One of the way to mitigate the progression of dementia is "coimagination" method proposed by Otake et al. [4]. It is a method by using pictures as the references for the topics in a group talking. All participants have equal chance to listen, to talk, to ask questions, and to answer questions. It is reported that the elderly who participated this activity talked and smiled more fluently than before. However, this method has the limitation that all of the participants have to meet at one single place which may be difficult in practical.

Bickmore et al. [5] proposed a companion agent to ease the anxiousness of elderly inpatients. Huang et al. [2] developed a rapport agent which analyzes facial expressions, backchannel feedbacks, and eye gazes of the user. The agent is designed to show behaviors which are supposed to elicit rapport. However, it does not try to

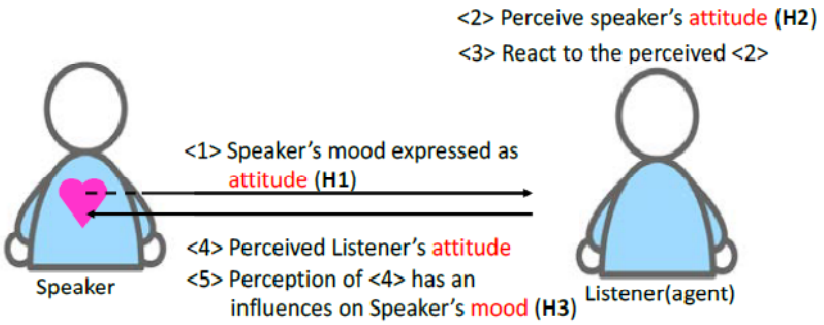

Fig. 1. Conceptual diagram of the proposed modeling of the interaction between the speaker and the listener (agent) during active listening

estimate and react to the user's mood. For example, when the user looks in bad mood, showing the agent's concern on the user by saying "Are you OK?" like human do. The SEMAINE project [6, 7] was launched to build a Sensitive Artificial Listener (SAL). SAL is a multimodal dialogue system with the social interaction skills needed for a sustained conversation with the user. They focused on realizing "really natural language processing [8] " which aims to allow users to talk with machines as they would talk with another person.

These projects were developed base on the subject studies in the U.S. or in other western countries where the subjects' communication style may diverge from that of Japanese ones [9]. In this study, we collected an active listening corpus of Japanese subjects and analyzed Japanese style verbal / non-verbal behaviors which potentially achieve the speaker's rapport toward the listener. On the other hand, previous studies in developing companion agents were usually started already with the premise that agent's behaviors have influence on the building of rapport. The analysis of this study bases on the self-evaluation on the subjects' mood and attitude in numerical way at fine granularity. The detailed comparison is then conducted to explore whether it is actually possible for the agent to do that.

3 Modeling of Active Listening

Since the goal of this study is to build a virtual listener agent which can establish rapport with elderly users, two functions of the agent can be considered essential: the agent's attitude perceived by the user has an influence on speaker's mood, and the agent can estimate speaker's mood from the speaker's attitude perceived by the agent. As the discussion in section 1, we formalize rapport as the interaction between the interlocutors' "mood" and "attitude" in this work. Here, we redefine these two general terms in the following way:

Mood: someone's internal mental status. It lasts for relatively longer time (say 10 or 20 minutes) than emotion and is difficult to be observed by another person.

Attitude: someone's mood expressed in the way how he or she behaves toward another person. These behaviors include verbal utterances and non-verbal ones like gestures, postures, change of voice tone, and facial expressions. It is supposed to be able to be detected by another person.

Figure 1 shows the conceptual diagram of the proposed modeling of the interaction between the speaker and the listener (agent) during active listening. <1> The speaker's internal mental state or his / her mood is expressed as his / her attitude <2> The speaker's attitude is perceived by the listener (or the companion agent) <3> The listener interprets the speaker's attitude and estimate the speaker's current mood. The listener then decides how he / she should react to the speaker's mood. For example, says "Are you feeling bad? Do you need some rest?" if the speaker looks tired; or says "It's ok. Everybody experiences that." if the speaker is talking about some sad memory and his / her mood is going downward <4> The listener's reaction is then perceived by the speaker as the listener's attitude for the speaker <5> The interpretation on the perception of the listener's attitude then has influence on the speaker's mood. The speaker's mood is then expressed as his / her attitude. This interaction continues as a loop until the conversation ends.

To build rapport with the elderly user, the companion agent should have two fundamental functions: its behaviors (or its attitude) perceived by the speaker has an influence on the speaker's mood, and it can estimate the speaker's mood from his / her observed attitude. However, whether these functions are possible cannot be verified directly because of the agent has not been implemented yet. Before actually developing the agent, we would like to verify whether these two functions are possible to be realized. As described in section 1, a third person without premised impression about the subjects and prior knowledge about the dialogue corpus is assumed to have similar ability in the evaluation of subjects' mood and attitude as the agent, i.e. can only judge objectively. Therefore, the verification can be formulized as validating the following hypotheses:

H1: the speaker expresses his / her mood as his / her attitude.

H2: the speaker's attitude can be perceived by a person who does not know the speaker well in advance.

H3: the speaker's mood is influenced by how he / she perceived the listener's attitude.

4 Hypothesis Validation

The three hypotheses raised in last section were validated by the following procedure. At first, a human-human active listening experiment was conducted. Second, the collected corpus was evaluated by the experiment participants (speaker and listener) and another participant with third person view. Third, the correlations between the results from each two evaluator were computed.

4.1 Active Listening Experiment

Experiment setup: five pairs of participants (four male pairs and one female pair) with the same gender were recruited in Ritsumeikan University, all of them were college students and native Japanese speakers (average age: 22.1 years old). The two participants of each pair were recruited with the condition that they are close friends. This is because close friends were considered easier to talk with each other in limited experiment time. In order to simulate the situation of talking with a 2D graphical agent, the participants of each pair were separated into two rooms and talked with each other via Skype. In each session, one participant played the role as the speaker, and the other one played the role as the listener. Large lower body movements may affect the movements of upper body which is more important in communication. They were instructed to sit on a chair so that the move of their lower bodies can be controlled within a limited range. Each room was equipped with two video cameras. One was used for recording the participant from the front. The other one was used for logging the Skype window which was duplicated on another monitor. In addition to video recordings, Microsoft's Kinect depth sensors were also used to record the participants' movement for further analysis. The speaker talked with the listener who was projected on a large screen at around life-size. The height of the projected image was adjusted so that the speaker can see the listener's eyes roughly at the level for eye contact. Natural head movements and eye gazes shifts can be further analyzed. The setup of the two experiment rooms are shown in Figure 2 and 3.

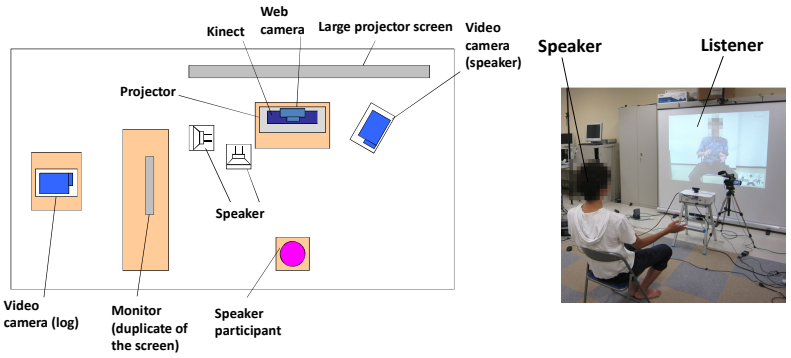

Fig. 2. Setup of the room where the speaker participant was in. The listener was projected roughly as life-size and the second monitor was used for video logging.

Experiment procedure: Each participant pair talked in four sessions. The topics of the conversation were "pleasant experience with family" or "unpleasant experience with family." These topics were chosen because they are common for almost everyone including the young experiment participants and the elderly. Each participant played the role either as the speaker or as the listener. Speaker participant initiates the session and talks to the listener about his / her family. Listener participant was instructed to try to be a good active listener. That is, listen to the speaker carefully and

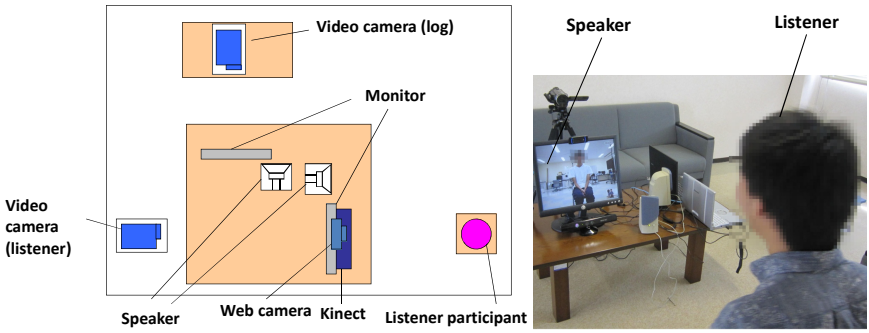

Fig. 3. Setup of the room where the listener participant was in. The second monitor was used for video logging.

attentively, follow the speaker's talk with questions or other feedbacks like nods or laugh. Table 1 shows the assignment of the order and talk topics of the two participants. They interchanged their roles in the sessions and started to talk from the pleasant experience at first because it should be easier to do. The duration of one session was set to be seven minutes because it is considered long enough for the participants to start to talk something meaningful and keep the whole experiment with a reasonable time.

Table 1. Arrangement of talk topics and the roles of the two participants (denoted as A and B)

Session	Topic	Speaker	Listener
1	Pleasant experience with family	A	B
2		B	A
3	Unpleasant experience with family	A	B
4		B	A

4.2 Evaluation of Mood and Attitude

After the end of four sessions, the participants were instructed to evaluate the mood and attitude of themselves and their partners by labeling on the recorded video corpus. The video annotation tool, ELAN [10] was used for this purpose. In order to align the granularity and label positions among different coders, the participants were instructed to label their evaluation by following the four rules:

1. The whole time line has to be labeled without blank segments
2. Starting and ending positions of the label should be aligned to utterance boundaries
3. One label can include multiple utterances
4. The maximum length of one individual label is 10 seconds

Phonetics tool, Praat [11] was used to label the boundaries of participants' utterances. Figure 4 shows a screen capture of ELAN software, the beginning and ending positions of all utterances 10-second scales are automatically labeled for the participants' easy

reference. Table 2 lists the assignment of the annotation. The experimenter who did not participate the conversation directly annotated the corpus as the third person. The mood of the speaker and the attitude of the speaker and the listener were evaluated by the speaker, the listener, the third person (the experimenter) by following criteria:

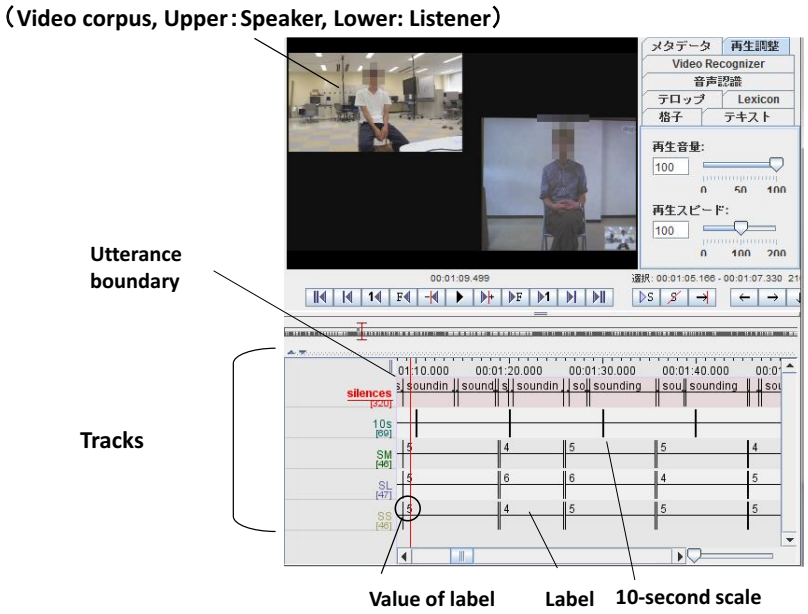

Fig. 4. Screen capture of ELAN annotation tool. 10-second scale and utterance boundaries are labeled with software tools for the convenience of the participants.

Mood: evaluated with 7-scale measure from 1 (negative) to 7 (positive). Comfort, intimacy, and sympathy are provided as positive examples in the instruction.

Attitude: evaluated with 7-scale measure from 1 (negative) to 7 (positive). Appropriate back-channel feedbacks like nods, questions, silence, agreeing opinions, smiles, or laughs were provided as positive examples in the instruction.

Table 2. Assignment of the annotation on mood and attitude. S, L, 3 at first character of the label is the abbreviation of the coder, speaker, listener, and the third person, respectively. M, L, S at the second character of the label is the abbreviation of speaker's mood, listener's attitude, speaker's attitud, respectively.

Coder	Label	Meaning
Speaker	SM	Speaker's mood
	SL	Listener's attitude
	SS	Speaker's attitude
Listener	LS	Speaker's attitude
	LL	Listener's attitude
Third person	3S	Speaker's attitude
	3L	Listener's attitude

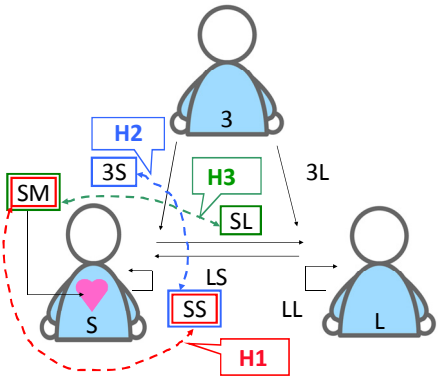

Fig. 5. Conceptual diagram of the relationship between the evaluation and the hypotheses

4.3 Computation of Correlations

The labeled mood or attitude tracks are wave-form like data sequence. In order to compare the data with different granularity and different boundaries, we cut the tracks to small slices (or resampled them). Since the shortest interval of the labels was 0.244 second, the sampling rate was set to 0.1 without losing data. Because the labeling evaluation are subject evaluation and the results among different coders cannot be compared directly. The data are then normalized with Z-score. Pearson product-moment correlation coefficient is computed between the labels which correspond to hypotheses, H1, H2, and H3 (Figure 5). The labeling results and the data number used for validating the hypotheses are shown in Table 3.

Table 4 shows the result of computing the correlation to validate the hypotheses. H3 has a strong correlation, and it means listener's attitude has an influence on the speaker's mood. H1 also has a strong correlation. H2 has lower, but shows positive tendency of correlation. It means that the third person (or the agent) can perceive the speaker's attitude and estimate his / her mood. Three hypotheses have positive correlation. Therefore, it indicates that there are some behaviors which change speaker's mood to positive or to negative state. If those behaviors are implemented to the agent, a companion agent who can build rapport with the human user should be possible.

Table 3. Summary of labeling results and the number of data used for correlation computation

Pair	Label #	SL and SM	SS and SM	SS and 3S
1	1,745	17,236	17,235	16,758
2	1,536	17,163	17,163	17,035
3	1,350	17,015	17,016	16,951
4	1,472	17,472	17,395	16,793
5	1,234	16,280	16,280	16,259

Table 4. Results of correlation hypothesis validation

Hypothesis	Label	Correlation
H1	SS and SM	0.50
H2	3S and SS	0.21
H3	SL and SM	0.48

5 Conclusion

In order to develop a companion agent which can engage active listening, two functions are essential: the listener's attitude have an influence on the speaker's mood, third person (agent) can estimate the speaker's mood from perception of the speaker's attitude. However the possibilities of these functions are not validated yet. To examine them, we proposed three hypotheses: the attitude of the conversation can have influence on the speaker's mood, the mood of the speaker can be potentially observed by another person, and third person can perceive the speaker's attitude correctly. These hypotheses were validated from a human-human conversation experiment with self-evaluations and third person view evaluation.

In the future, at first we would like to increase the corpus size with additional experiments. And then, we would like to analyze the low-level signals of listener and how they can be interpreted as the listener's attitude. Also, we would like to analyze the listener's strategy in how to react to the speaker's perceived attitude. When the technology becomes matured, we will implement this function to agent and test with the elderly.

References

1. Huang, H.H., Matsushita, H., Kawagoe, K., Sakai, Y., Nonaka, Y., Nakano, Y., Yasuda, K.: Toward a memory assistant companion for the individuals with mild memory impairment. In: 11th IEEE International Conference on Cognitive Informatics & Cognitive Computing, pp. 295–299 (2012)
2. Huang, L., Morency, L.P., Gratch, J.: Virtual rapport 2.0. In: 10th International Conference on Intelligent Virtual Agents, pp. 68–79 (2011)
3. Tickle-Degnen, L., Rosenthal, R.: The nature of rapport and its nonverbal correlates. Psychological Inquiry 1(4), 285–293 (1990)
4. Otake, M., Kato, M., Takagi, T., Asama, H.: Coimagination method: supporting interactive conversation for activation of episodic memory, division of attention, planning function and its evaluation via conversation interactivity measuring method. In: International Symposium on Early Detection and Rehabilitation Technology of Dementia, pp. 167–170 (2009)
5. Bickmore, T., Laila Bukhari, L., Vardoulakis, L., Paasche-Orlow, M., Shanahan, C.: Hospital buddy: A persistent emotional support companion agent for hospital patients. In: 12th International Conference on Intelligent Virtual Agents, pp. 492–495 (2012)

6. McKeown, G., Valstar, M.F., Cowie, R., Pantic, M.: The SEMAINE corpus of emotionally coloured character interactions. In: IEEE Intl Conf. Multimedia & Expo, pp. 1079–1084 (2011)
7. Pammi, S., Schroder, M.: Annotating meaning of listener vocalizations for speech synthesis. In: Affective Computing & Intelligent (2009)
8. Cowie, R., Schroder, M.: Piecing together the emotion jigsaw. In: Machine Learning for Multimodal Interaction, pp. 305–317 (2005)
9. Hofstede, G., Hofstede, G.J., Minkov, M.: Cultures and Organizations: Software of the Mind, 3rd edn. McGraw-Hill, New York (2010)
10. Max Planck Institute for Psycholinguistics, ELAN annotation tool, http://tla.mpi.nl/tools/tla-tools/elan/
11. Boersma, P., Weenink, D.: Praat: doing phonetics by computer, http://www.fon.hum.uva.nl/praat