
Practical Approaches to Comforting Users with Relational Agents

Timothy Bickmore, Daniel Schulman

Northeastern University
College of Computer and Information Science
360 Huntington Avenue, WWH 202,
Boston, MA 02115
{bickmore,schulman}@ccs.neu.edu

Abstract

Interactions in which computer agents comfort users through expressed empathy have been shown to be important in alleviating user frustration and increasing user liking of the agent, and may have important healthcare applications. Given the current state of technology, designers of these systems are forced to choose between (a) allowing users to freely express their feelings, but having the agents provide imperfect empathic responses, or (b) greatly restricting how users can express themselves, but having the agents provide very accurate empathic feedback. This study investigates which of these options leads to better outcomes, in terms of comforting users and increasing user-agent social bonds. Results, on almost all measures, indicate that empathic accuracy is more important than user expressivity.

Keywords

Comforting; caring; embodied conversational agent; affective computing; social interface; relational agent; emotion, empathy.

ACM Classification Keywords

H5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology, Graphical user interfaces, Interaction styles, Natural language, Theory and methods, Voice I/O.

Introduction

The ability for computer agents to provide comfort to users has important applications in healthcare and other areas. Automated systems that comfort users could be used in pain management, palliative care, and pre-surgery anxiety reduction. In addition to the potential health applications of automated comforting, simply alleviating transitory negative affect in users may have beneficial effects on attention, memory retention, learning, creativity, and attitude towards the perceived source of the affective state [5].

Empathy—inferring the feelings and thoughts of others—is a prerequisite for comforting, and is also crucial for establishing trust and working alliance between users and therapeutic agents. These forms of patient-provider relationship have been shown to be an important determiner of patient satisfaction, regimen adherence and health outcomes in a number of studies [2].

Relational agents are interface agents that are designed to establish social bonds with users [2]. In previous work, we have demonstrated the ability of a relational agent to establish a working alliance relationship with users over multiple interactions, partly through use of empathic and comforting behavior [1,2]. In these systems, the agent communicates to users with synthesized speech and synchronized nonverbal behavior (hand gesture, posture, facial display, etc.), but users are limited to selecting their utterances from a multiple choice menu, updated during each turn of the interaction. This design choice of restricted input (rather than allowing free text or speech) was intentional, since we felt it was important in a health counseling application for the agent to unambiguously

understand the user's intent. However, many users complained about their inability to express themselves:

“She would ask a question and I would have a choice, one, two, three, four. But I could never explain...I don't feel that it gets the true feeling, the true experience of a person.”

One area in which this restriction on user input may be especially problematic is in empathic interactions. Although much research has been conducted over the last decade on detecting user affect from speech, text, and passive sensors [6], these technologies are all far from perfect, and basing the behavior of a comforting agent on them could leave users feeling more frustrated and angry than if the agent hadn't noticed their feelings at all.

This raises the central research question explored in this work: in the design of effective comforting agents, is it better to restrict what users can say so that the agent can unambiguously determine their emotional states and provide accurate comforting (low expressivity and high empathic accuracy), or is it better to allow users to express their feelings in an unconstrained manner, knowing that the agent's inferences about their emotional state will often be incorrect (high expressivity and low empathic accuracy)? Our hypothesis is that demonstrating empathic accuracy is always more important for comforting, since it demonstrates greater understanding of what the user says.

Related Work

The most similar previous study to the current one was conducted by Klein in 1999 [5]. Klein experimentally

induced frustration in study subjects playing a computer game and then evaluated three methods for alleviating their frustration. His primary finding was that computer displays of empathy via text dialogs were significantly more effective at reducing user frustration compared to similar text-based interfaces that only allowed users to express their feelings (vent) or ignored their feelings altogether. Our study is inspired by Klein's but differs in three important ways: (1) Klein's system provided no feedback whatsoever in the vent condition, whereas we are interested in the efficacy of a neutral acknowledgement of the user's feelings, which we argue is currently the best practical response given unconstrained user input; (2) Our agent ("Louise") is embodied, and uses verbal and nonverbal modalities to convey emphasis, turn-taking cues, discourse shifts, and affect (e.g., facial display of concern when expressing empathy) [2]; and (3) users can more freely express themselves than they can in text, via intonation and prosody.

Relational Agent

We are using a relational agent named "Louise" in our investigation (see Figure 1). As in our previous work, this agent speaks to users using synthesized speech and synchronized nonverbal behavior. However, in this study, rather than having users select what they want to say using a mouse or touch screen, they speak their chosen utterance to the agent. We also allow free speech input at predetermined points in the dialogue. For the purpose of the study, speech recognition was performed by a human experimenter in a wizard-of-oz configuration [3].

Experimental Method

The study is a counterbalanced, randomized, two-condition within-subjects experiment to compare two methods a relational agent can use to calm users following a negative mood induction. The mood induction procedure is identical in both conditions (pilot testing indicated that it may be used twice in one session to achieve roughly equivalent levels of anxiety), but in one condition the agent appears immediately after the induction and uses a high expressivity dialogue script (EXPRESSIVITY), whereas in the other condition the agent uses a high empathic accuracy script (EMPATHY).

The high empathic accuracy script asks users "How are you feeling right now?" (stressed, anxious, exhausted, disappointed, "I've been better.", and great), and provides appropriate empathic feedback, including a happy facial display for 'great' and a close-up concerned facial display for the other feeling states. The high expressivity script asks users the same question, but then prompts them for a free speech input, and always replies with "Really? That is interesting to hear." with a neutral facial display. We also developed a short (3 minutes) introductory script, designed to acquaint users with Louise and the two modes of speech input.

Apparatus

Subjects are seated at a desk supporting a 17" touch screen color computer monitor (experiment computer) and a microphone. Immediately to their left is a 50" color plasma display monitor positioned at eye level, used to display the life-sized relational agent (see Figure 1). The experiment computer guides subjects through all steps of the experiment except when they



Figure 1. Experimental Setup

Measure	Question	1	7
SATISFIED	How satisfied were you with Louise?	not at all	very satisfied
CONTINUE	How much would you like to continue working with	not at all	very much
LIKE	How much do you like Louise?	not at all	very much
CARES	How much do you feel that Louise cares about you?	not at all	very much

Table 1. Self-report measures of attitude towards agent

are talking to the agent. Subjects' heart rate and skin conductivity are monitored continuously, using finger-clip sensors from Thought Technology, Ltd.

The experimenter/wizard is seated in an adjoining room and can see and hear everything the subject does via a remote monitor and headphones. Subjects are led to believe that the agent can understand their spoken utterances when, in fact, the experimenter controls the agent. During free speech, the experimenter causes the agent to nod its head whenever the user pauses briefly, and, if needed, to prompt the user to continue, using the phrases "please go on" or "tell me more".

Mood Induction

The mood induction procedure is run on the experiment computer, and is designed to induce mild anxiety. The procedure utilizes a mental arithmetic task combined with failure manipulation, similar to that used in previous anxiety induction studies [4]. The induction consists of an initial practice session and two test sessions, each lasting 3 minutes. During all three sessions, the subject is required to answer arithmetic problems. The practice session has no time limit per question, although the system surreptitiously records the average time used per question. During each test session, the subject is told that the average score is 60% correct. There is a time limit per question, which

initially is the average time used during the practice session, and is continually manipulated (as in [4]), to ensure that the time available per question will be slightly less than the time a subject needs, causing a low success rate (typically 30-40%). Subjects are continuously shown their current score relative to the "average", as well as the time remaining on the current question. In addition, we provide auditory feedback whenever a subject fails to answer a question correctly, using a sound that is considered extremely annoying and stressful, according to pilot tests.

Participants

Sixteen subjects (7 male, 9 female, aged 18-30) participated; all but one were students.

Procedure

After being seated, subjects are connected to the heart rate and skin conductivity sensors. They are told that they will be conversing with the agent, that they will take tests which measure "intelligence and reasoning ability", and are to follow the instructions on the experiment computer. The experimenter leaves the room for the remainder of the session. Subjects complete the practice session of the mood induction, followed by an introductory dialogue with the agent. This is followed by two repetitions of: (1) a rest period of 2 minutes, intended to calm the subject and

Measure	EMPATHY		EXPRESSIVITY		Paired-t		
	Mean	SD	Mean	SD	df	t	p
Change in Positive Affect (PANAS)	-2.47	5.45	-4.87	4.78	14	2.18	p<.05
Change in Negative Affect (PANAS)	2.93	4.08	2.93	6.39	14	00.0	n.s.
Post Intervention Heart Rate	75.60	10.43	76.96	9.60	14	1.71	p=.11
Change in Skin Conductivity	-0.16	0.84	-0.36	0.62	14	1.07	n.s.
SATISFIED	3.20	1.86	2.60	1.60	14	2.07	p<.05
CONTINUE	3.27	2.31	2.87	1.92	14	1.57	n.s.
LIKE	3.27	2.25	2.87	1.85	14	1.38	n.s.
CARES	3.07	2.19	2.67	1.84	14	1.70	p=.11

Table 2. Study Results

establish a baseline, during which the subject is shown an affectively neutral image; (2) a pre-induction assessment; (3) a 3-minute mood induction; (4) intervention by the agent, using one of the two conditions; and (5) a post-intervention assessment. At the end of the session, the experimenter holds a brief semi-structured interview with the subject.

Measures

Affective State is measured using a computer-administered self-report questionnaire and physiological measures to assess arousal of the sympathetic nervous system. The Positive and Negative Affect Schedule (PANAS) [7] is administered by the experiment computer immediately before each induction procedure and following each comforting intervention. Heart rate and skin conductivity are recorded continuously at 256 Hz.

Attitude Towards Agent is measured using the four Likert-scale self-report questions shown in Table 1,

administered by the experiment computer at the end of each comforting intervention.

Results

Data for one subject was removed as an outlier due to several of his measures being more than 3 SD away from the mean relative to all other subjects. Motion artifacts were manually removed from the heart rate data. One-minute-averaged samples of physiological data were taken at the end of the induction and at the end of the intervention.

Results are shown in Table 2. Six of the eight outcome measures indicated that the EMPATHY condition was superior to the EXPRESSIVITY condition, and two of these results are statistically significant. The change in positive affect from pre-induction to post-intervention was greater in the EMPATHY condition than in the EXPRESSIVITY condition, paired $t(14)=2.18$, $p<.05$. In addition, subjects were more satisfied with the agent in the EMPATHY condition than in the EXPRESSIVITY condition, paired $t(14)=2.07$, $p<.05$.

When asked whether they felt one intervention was more comforting than the other, six subjects (out of eight that gave a preference), preferred the EMPATHY condition.

“I think I like that more than the, you know, talk at length, when she didn’t seem to respond to anything... I know some people need to just talk, and then they work out their thoughts in their head, and I’m not like that.”

While we did not find any significant order effects in the outcomes, we did find one significant gender effect: females expressed a greater desire to CONTINUE working with the agent in the EXPRESSIVITY condition than males, $t(13)=2.23$, $p < .05$.

Conclusion

We found that an agent with greater empathic accuracy was more efficacious at comforting users, even at the cost of restricting user input. This is an important result for designers of these systems, since it indicates that sophisticated affect recognition technology is not only unnecessary; it can actually be counter-productive.

Future work should examine what choices of possible user utterances can maximize an agent’s effectiveness. We believe that this strategy holds promise for increasing both the utility of agents in important applications, such as healthcare, and for increasing user satisfaction with these agents.

Acknowledgements

Thanks to Thomas Brown for his role as experimenter for many of the subjects. Thanks also to Jennifer Smith for her helpful comments on this paper.

References

- [1] Bickmore, T., Caruso, L., and Clough-Gorr, K., Acceptance and Usability of a Relational Agent Interface by Urban Older Adults, *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2005.
- [2] Bickmore, T. and Picard, R. Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Transactions on Computer Human Interaction*, 12, 2, (2005) 293-327.
- [3] Dahlback, N., Jonsson, A., and Ahrenberg, L., *Wizard of Oz Studies -- Why and How*. In M. T. Maybury and W. Wahlster, Eds., *Readings in Intelligent User Interfaces*, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 1998, pp. 610-619.
- [4] Dedovic, K., Renwick, R., Mahani, N., Engert, V., Lupien, S., and Pruessner, J. *The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *Journal of Psychiatry and Neuroscience* 30, 5, (2005) 319-325.
- [5] Klein, J., Moon, Y., and Picard, R. This Computer Responds to User Frustration: Theory, Design, Results, and Implications. *Interacting with Computers*, 14, (2002) 119-140.
- [6] Picard, R., *Affective Computing*, MIT Press, Cambridge, MA, 1997.
- [7] Watson, D., Clark, L., and Tellegen, A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J personality and Social Psychology*, 54, (1988) 1063-1070.