Perceived Emotional Intelligence in Virtual Agents

Yang Yang
The Hong Kong University of Science and Technology
HKSAR, China
yyangag@connect.ust.hk

Xiaojuan Ma
The Hong Kong University of Science and Technology
HKSAR, China
mxj@cse.ust.hk

Pascale Fung
The Hong Kong University of Science and Technology
HKSAR, China
pascale@ece.ust.hk

CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA

ACM 978-1-4503-4656-6/17/05.

http://dx.doi.org/10.1145/3027063.3053163.

Abstract

In March 2016, several online news media reported on the inadequate emotional capabilities of interactive virtual assistants. While significant progress has been made in the general intelligence and functionality of virtual agents (VA), the emotional intelligent (EI) VA has yet been thoroughly explored. We examine user's perception of EI of virtual agents through *Zara The Supergirl*, a virtual agent that conducts question and answering type of conversational testing and counseling online. The results show that overall users perceive an emotion-expressing VA (EEVA) to be more EI than a non-emotion-expressing VA (NEEVA). However, simple affective expression may not be sufficient enough for EEVA to be perceived as fully EI.

Author Keywords

Emotional intelligence; emotions; virtual agent speech and language interaction; interactive dialog system.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

According to Miner et al.'s research published in March 2016, various world-renowned virtual assistants e.g. Siri (Apple), Google Now, S Voice (Samsung), and

Figure 1. The resting state of emotion-expressing virtual agent (EEVA).

Figure 2. An example of EEVA's positive expression.

Figure 3. An example of EEVA's negative expression.

Cortana (Microsoft) provided rather impersonal and inconsistent responses to participants' questions on mental health, domestic violence, and emergencies [19]. As a service intended to mimic human interaction, this scenario beckons a need for empathy and emotional intelligence (EI) to deal with affective information appropriately. Various studies have shown that the way humans react to virtual agents (VA) resembles how they react to real people [16, 20]. Could empathetic agents be used to handle stressful situations and would people perceive them as EI?

Research in affective computing introduced emotions to computers, robots, and VA [3, 9, 11, 15, 16]. Research shows this can alleviate user frustration with computers [16], influence behavioral changes [15], enhance learning in children [9], and endow complex emotions to VAs [3]. A healthcare VA like SimSensei [6] needs a way to ensure empathetic and sensible services are provided to users. However, none of these techniques serves to understand how people perceive EI in VA equipped with emotional expression and recognition. Thus, this research gap leaves room for novel solutions. In our research, we focused on examining EI in VA.

Zara the Supergirl

In this abstract, we present Zara the Supergirl [8, 9], an interactive dialogue system rendered in the form of a virtual agent. It can recognize user emotions, express simple emotions, and respond user emotions detected. The system runs online or locally with a simple UI that shows the virtual agent, agent's question, and user's response. The agent starts by introducing her name, her task, and what to expect from the conversation.

The agent analyzes user emotions through speech, text, and facial expressions captured real time [8, 9]. The agent expresses emotions via animations with body language and choice of words. The agent has simple affective expressions (see Figures 2 & 3) for positive (while saying things like "Fantastic," "That is awesome" etc.) and negative (while making comments like "So horrible," "Awful," etc.) emotions. To account for emotionally neutral answers from users, the agent also has a resting pose (see Figure 1) and a list of neutral feedback (i.e. "I see," "Okay").

It is important to clarify at this point that the agent does not manifest the ability to feel empathy or other emotions. Instead, it is designed with minimal elements to convey to users a sense of empathy, or "an affective response more appropriate to someone else's situation" [14]. A way humans express empathy is mimicry [2]. Mimicry happens when an observer makes a reaction similar to the observed [2]. Mimicry can also increase empathy in humans during virtual human interactions [12]. Therefore, the virtual agent (VA) in this study is designed to reciprocate user's sentiment by expressing the same sentiment. This follows the law of attraction that can increase its likability to humans [18].

The VA in this study plays the role of an assistant to evaluate the stress levels of current university students through a 5-7min, one-on-one conversation. The dialog has five subtopics modeled after Depression Anxiety Stress Scales (DASS) [13] in a manner of increased intimacy [1]. It starts with general topics (i.e. hobbies and travel), then personal topics (i.e. friends and families, current study and future plans), and finally a generic appraisal on students' wellbeing (i.e. exercise and downtime). Each topic has 2-3 open-ended short

Figure 4. NEEVA's resting state remains constant during the dialog.

questions and ~15 seconds to respond. Users receive an appraisal of their stress level at the end.

A non-emotion expressing version (NEEVA) of the agent that has all abovementioned functionalities acts as the control of this experiment. Here we shorthand the empathetic interactive dialogue system as EEVA EEVA responds to participant's answers with a short positive, negative, or neutral emotive interjection while the NEEVA moves on to ask the next question (see With the Zara system, we want to investigate how users perceive EI of a VA when it expresses simple positive and negative emotions. Our hypothesis is that users will perceive EEVA as more emotionally intelligent than NEEVA and will consider their interaction with EEVA more satisfying than that with NEEVA.

MSCEIT-Based Perceived EI Questionnaire	
Br.	**Questions**
PE	Convey a sense that it **listen**s openly to participant's emotions
	Convey a sense that it pays **attention** to user's moods during the conversation
	Identify user's emotions correctly
	Discern between different emotions
	Tell apart the **degree** of emotions present
UsE	Convey a sense that the agent can **feel** what the user is feeling
	Convey a sense that the agent understand user's point of view (**POV**)
	Respond in a way that make the user feel **sad**
	Respond in a way that make the user feel **happy**
	Respond in a way that make the user feel that they are **understood**
UnE	Convey a sense that it can be emotionally **self-aware** and insightful
	Display some knowledge of **complex** emotions
	Respond **empathetic**ally to user
	Describe/understand difficult emotions
	Give user an impression it is **attempt**ing to empathize
ME	Make **decisions** with feelings and thoughts
	Influence some of user's thoughts
	Provide psychologically-minded **advice**
	Shows some **conscious** thought before responding
	Show varying **openness** to various emotions

Table 1: 20Q adapted from MSCEIT v2.0 to assess PEI; Br. is short for "Branch".

Experiment Design and Measures

The experiment was conducted in-person as a within-subject, counter-balanced study with perceived EI (PEI) as the primary dependent variable of interest. All users talked to both EEVA and NEEVA. After each interaction, rated their experience on a questionnaire that contains 20 questions modified after Mayers-Salovey-Calousey-Emotional-Intelligence-Test (MSCEIT) v2.0 [18]. There are four branches (i.e. perceiving, using, understanding, and managing emotions), with five questions per branch, rated on a 5-pt Likert-scale (1="Never", 5="Always"). The four branches are abbreviated PE, UsE, UnE, and ME respectively. Since MSCEIT emphasizes ability, we start questions on the constructed questionnaire [18] with "The agent is able to…" and the short-forms, or attributes, are bolded.

Participants

A total of 40 participants (17 females), aged 18 to 34, were recruited via email, social media, and word-of-mouth at a local university (*Toefl score* >= 100/120). About 43% of the users had some prior experience with virtual agents and indicated infrequent interactions (M=2, SD=0.37, on a 5-pt Likert scale of 1="Never" to 5="Always"). A lack of opportunity or access to VA is the most common reason for those with no experience.

MSCEIT-Based Perceived EI Questionnaire Results						
Branches	**Attributes**	**df**	**MS**	***f***	***P-values***	η^2
Perceiving Emotions (PE)	Listen	1	1.25	1.59	.215	.039
	Attention	1	5.51	6.52	.015*	.143
	Identify	1	1.25	1.40	.243	.035
	Discern	1	.31	.38	.542	.010
	Degree	1	2.11	2.54	.119	.061
Using Emotions (UsE)	Feel	1	3.20	4.33	.044*	.100
	POV	1	2.81	4.84	.034*	.110
	Sad	1	.05	.05	.830	.001
	Happy	1	.32	.64	.430	.016
	Understood	1	.11	.17	.686	.004
Understanding Emotions (UnE)	Self-Aware	1	1.25	1.79	.200	.042
	Complex	1	1.25	2.05	.160	.050
	Empathetic	1	3.2	5.03	.031*	.114
	Describe	1	.20	.16	.694	.004
	Attempt	1	5.00	6.96	.012*	.152
Managing Emotions (ME)	Decision	1	3.61	6.75	.013*	.147
	Influence	1	.20	.30	.586	.008
	Advice	1	.61	.86	.360	.021
	Conscious	1	5.00	9.75	.003**	.200
	Openness	1	.45	.69	.412	.017

Table 1: Results of PEI on VA (*p < .05, ** p < .01)

User Expectations of Virtual Agents (VA) Interactions. Of the 40 users, 70.59% indicated that existing VAs are not empathetic or emotionally intelligent (EI) (M=2.03, SD=0.84, on a 5-pt Likert scale of 1= "Not at all" to 5="A great deal") and 73.27% expected them to be empathetic and EI in the future (M=3.23, SD=1.12, on a 5-pt Likert scale of 1="Not at all" to 5="A great deal"). A gift token was presented to each participant after completion of the test.

Results & Analysis

Manipulation Check. To ensure that users recognize and remember both emotion-expressing VA (EEVA) and non-emotion-expressing VA (NEEVA), we performed a manipulation check at the end of the experiment. The EEVA framed as being more emotion-expressing was indeed perceived by users to be more emotion-expressing (M=3.67, SD=1.29) than its NEEVA (M=1.72, SD=1.13), repeated measures MANOVA, F(1, 76.05)=29.97, p<.001, η^2 =.44.

Perceived EI. Overall, 65% of the users perceive emotion-expressing VA (EEVA) to be more EI than non-emotion-expressing VA (NEEVA) and 60% agrees that they sense a strong self-awareness from EEVA than NEEVA (see Figure 5). EEVA outperforms NEEVA in some certain attributes within each dimension (see Table 1). For results shown here, a Bonferroni correction was applied to all post-hoc contrasts.

- *PE branch.* There is a significant main effect on the Attention attribute, repeated measures MANOVA, F(1, 5.51.)= 6.52, p<.015, η^2= .143. The users who interacted with EEVA felt that it was better at paying attention (M=3.28, SD=.88), than NEEVA, (M=3.80, SD=.99), pairwise comparison (p=.015). However, we did not find any statistically significant difference in the remaining four attributes (see Figure 5).
- *UsE branch*. There are two significant main effects on the attributes 1) Feeling: repeated measures MANOVA, F(1, 3.20.)=4.33, p<.044, η^2 =.100, and 2) POV: repeated measures MANOVA, F(1, 2.81)=4.84, p<.034, η^2=.110. Users who interacted with EEVA perceived it be more capable of feeling what they were experiencing, (M= 3.55, SD=1.01), than NEEVA, (M=3.15, SD=.94), pairwise

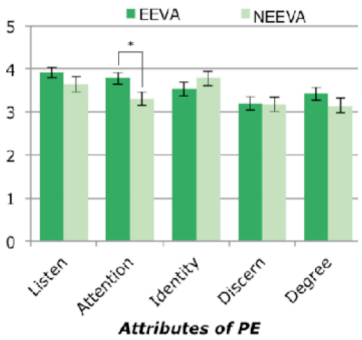

Figure 5. Avg. user rating (max=5) of attributes in PE.

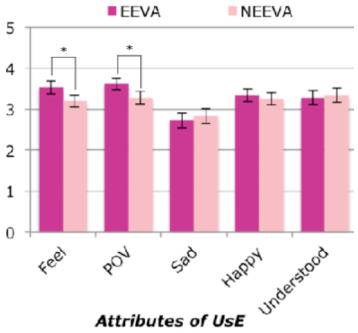

Figure 6. Avg. user rating (max=5) of attributes in UsE.

comparison (p=.044). Users also consider 3.65, SD=.88), than NEEVA, (M=3.28, SD=.98), pairwise comparison (p=.034) (see Figure 6).

- *UnE Branch*—There are 2 significant main effects on the attributes 1) Empathetic: repeated measures MANOVA, F(1, 3.20) =5.03, p<.031, η^2=.114, and 2) Attempt: repeated measures MANOVA, F(1, 5.00)=6.96, p<.01, η^2=.152. Users see EEVA as better at empathetic response, (M=3.53, SD=.95), than NEEVA, (M=3.13, SD=1.04), pairwise comparison (p=.031). Users also sense a stronger attempt to empathize from EEVA, M=3.88, SD=.80, than NEEVA, M=3.38, SD=1.08, pairwise comparison (p=.012) (see Figure 7).
- *ME Branch.* There is a significant main effect of EI on the Decision attribute, repeated measures MANOVA, F(1, 3.62.)=6.75, p<.013, η^2=.147. Users perceive EEVA to be more apt at making decisions, (M=3.90, SD=.72), than NEEVA, (M =3.78, SD= .95), pairwise comparison (p=.013). A highly significant effect of EI was found on the Conscious attribute, repeated measures MANOVA, F(1, 5.00) =9.75, p=.003, η^2 =.200. Users perceive EEVA to be much better at showing conscious thought before replying to user, (M=3.75, SD=.74), than NEEVA, (M=3.25, SD= 1.04), pairwise comparison (p=.003) (see Figure 8).

Overall, users perceive emotion-expressing VA (EEVA) to be slightly more EI than non-emotion-expressing VA (NEEVA), and rate it higher in 16 attributes.

Users Satisfaction. Users rate their satisfaction after interacting with EEVA and NEEVA. Overall, 70% either agreed or strongly agreed that they feel more satisfied talking to EEVA than NEEVA. Only 12.5% feel more satisfied with NEEVA and 17.5% expressed indifference.

Qualitative Results. To capture users' opinions about the VAs, we asked them to specify, "Why did you say the agent you chose was the EI version?" and how willingly they are to interact with a VA in the future.

In most cases, users said that EEVA was more (EI). We learnt more about why users prefer EEVA to NEEVA:

- "*It is more cheerful and responsive.*"—P(3, F, 21), P(15, M, 21), P(25, F, 21)
- "*It responds to my answers before asking a new question.*"—P(27, M, 21), P(29, M, 20), P(38, M, 20), P(8, F, 23), P(21, M, 22)
- "*Even though it misinterpreted what I was saying, it showed that [EEVA] was definitely more empathetic.*"–P(31, M, 20)

We also gained insight from the reasons users choose *not* to speak to non-emotion-expressing VA (NEEVA):

- "*It interrupts me several times.*"—P(11, F, 19), P(35, F, 24)
- "*It asked some things irrelevant to what I said.*"—P(22, M, 20), P(19, F, 24)
- "*It is just drilling all the questions.*"—P(31, F, 20)

We also discovered that 20% users find the two systems to be pretty similar and the reason why:

- "*They sound the same.*"— P(24, M, 20), P(7, M, 26), P(16, M, 21), P(17, M, 23), P(33, F, 20), P(40, F, 27)

Discussion, Limitations, and Future Work

The results broadly support our hypothesis. Users especially caught on the impression that emotion expressing virtual agents (EEVA) was more conscious with its replies. One reason could be EEVA reciprocate

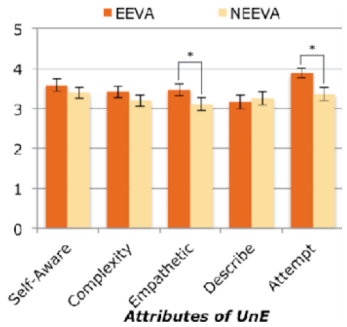

Figure 7. Avg. user rating (max=5) of attributes in UnE.

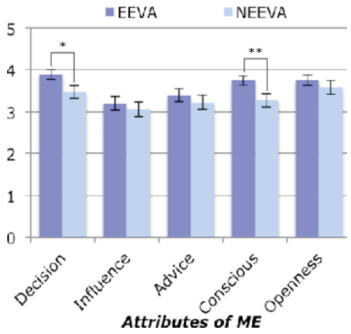

Figure 8. Avg. user rating (max=5) of attributes in ME.

user's answers before asking the following up question. This is similar to how human counselors use "reflection of feeling" and backchannel feedback to cultivate a good relationship with his/her clients [17]. However, emotion-expressing VA (EEVA) stood out as being more EI than non-emotion-expressing VA (NEEVA)in only seven of the 20 attributes. Of the participants, 20% noted that the difference between the two systems was not significantly noticeable. This may be because just as users vary in their degree of emotion expressivity in terms of mood and personality, they may also differ in emotion receptivity [4]. We see that simply adding capabilities to express emotions to a all attributes and need to consider varying degrees of user's emotion receptivity in future designs.

For qualitative results, we were surprised by users' more forgiving stance on EEVA when it misinterpreted their emotions. They were more critical toward NEEVA, pointing out that it "cannot even recognize the right meaning of what I say." –P(19, F, 24). One reason could be the "warm or cold" first impression users formed about the VA [5]. In human interactions, warm people are seen as friendly with good intentions [5, 7]. According to Frisk et al., "warmth judgments carry more weight in affective and behavioral reactions" [7] and these judgments are easier to lose and harder to regain [7]. Therefore, user's first impression of the agent's warmth may have persisted through the test and made them more forgiving toward EEVA.

Our research has its limitations with the VA's gender. Female characters are often stereotyped as emotional, empathetic, and warm. The current system's animation design with skintight clothes and red colors may also be a stereotyping factor. It is likely that this introduces expectation bias to the PEI scores. A neutral agent or a male agent could be used in the future to compare and see if similar results would surface. Future user testing can also take into account VA's gender, the types and degree of emotions expressed, and VA's personalities.

Another limitation is the singularity of the user pool. To assess the validity of the PEI scale, we need to examine it under a more diversified and bigger user pool to see if any result changes. For instance, feedback from older people (≥60 years old) who rarely use modern technologies can be one way to diversify user profiles.

Yet another limitation is the nature of the task. Stressful situations call upon EI. However, is EI necessary and suitable in all situations? An obvious example that may not need EI is factual questions like "What time is it in New York?" A possible next step is to discern different situations where an EI-equipped VA is necessary and suitable, and examine the perceived emotional intelligence (PEI) of VA in these situations.

Conclusion

Testing with an interactive virtual agent that can recognize and reciprocate user's feelings, we found that users perceive emotion-expressing virtual agents (EEVA) to be more emotional intelligent (EI), especially in the Conscious, Attention, Feel, POV, Empathetic, Attempt and Decision attributes. However, less than 50% of all attributes were significant. Therefore, we conclude that simple affective expression may not be enough for users to perceive a VA as fully EI.

Conclusion

A thank you to all developers of *Zara, the Supergirl* platform and all participantws for their time and support.

References

[1] Arthur Aron, Edward Melinat, Elaine N. Aron, Robert Darrin Vallone, and Renee J. Bator. 1997. The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin* 23, 4: 363-377.

[2] Janet Beavin Bavelas, Alex Black, Charles R. Lemery, and Jennifer Mullett. 1990. 14 Motor mimicry as primitive empathy. *Empathy and its development* 317.

[3] Christian Becker-Asano and Ipke Wachsmuth. 2010. Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems* 20, 1: 32-49.

[4] Nancy J. Briton and Judith A. Hall. 1995. Beliefs about female and male nonverbal communication. *Sex Roles* 32, 1-2: 79-90.

[5] Amy J. Cuddy, Peter Glick, and Anna Beninger. 2011. The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior* 31: 73–98/

[6] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, Louis-Philippe Morcency. 2014. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 1061-1068.

[7] Susan T. Fiske, Amy J. C. Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Science* 11, 2: 77–83.

[8] Pascale Fung, Anik Dey, Farhad Bin Siddique, Ruixi Lin, Yang Yang, Dario Bertero, Wan Yan, Ricky Chan Ho Yin, and Chien-Sheng Wu. 2016. Zara: A Virtual Interactive Dialogue System Incorporating Emotion, Sentiment and Personality Recognition. In *Proceedings of COLING 2016*, 278-281.

[9] Pascale Fung, Anik Dey, Farhad Bin Siddique, Ruixi Lin, Yang Yang, Wan Yan, and Ricky Chan Ho Yin. 2015 "Zara the Supergirl: An Empathetic Personality Recognition System." In *Proceedings of NAACL-HLT 2016 (Demonstrations)*, 87-91.

[10] Castellano, Ginevra, Ana Paiva, Arvid Kappas, Ruth Aylett, Helen Hastie, Wolmet Barendregt, Fernando Nabais, and Susan Bull. 2013. Towards empathic virtual and robotic tutors. In *International Conference on Artificial Intelligence in Education*, Springer, Berlin Heidelberg, 733-736.

[11] Jonatha Gratch, and Stacy Marsella. 2004. A domain-independent framework for modeling emotion. *Cognitive Systems Research* 5, 4:269-306.

[12] Béatrice S. Hasler, Gilad Hirschberger, Tal Shani-Sherman, and Doron A. Friedman. 2014. Virtual peacemakers: mimicry increases empathy in simulated contact with virtual outgroup members. *Cyberpsychology, Behavior, and Social Networking* 17, 12: 766-771.

[13] Julie D. Henry and John R. Crawford. 2005. The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British journal of clinical psychology* 44, 2: 227-239.

[14] Martin L. Hoffman. 2001. Empathy and moral development: Implications for caring and justice. Cambridge University Press,

[15] Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. 2013. I can help you change! An empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems (TMIS)* 4, 4: 19.

[16] Jonathan Klein, Youngme Moon and Rosalind. W. Picard. 2002. This computer responds to user frustration: Theory, design, and results. *Interacting with computers* 14, 2: 119-140.

[17] Yuka Kobayashi, Daisuke Yamamoto, Toshiyuki Koga, Sachie Yokoyama, and Miwako Doi. 2010. Design targeting voice interface robot capable of active listening. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, IEEE Press, 161-162.

[18] John D. Mayer, Peter Salovey, David R. Caruso, and Gill Sitarenios. 2003. Measuring emotional intelligence with the MSCEIT v2.0. *Emotion* 3, 1: 97.

[19] Adam S. Miner, Arnold Milstein, Stephen Schueller, Roshini Hegde, Christina Mangurian, and Eleni Linos. 2016. Smartphone-Based Conversational Agents and Responses to Questions about Mental Health, Interpersonal Violence, and Physical Health. *JAMA Internal Medicine* 176, 5: 619-625.

[20] Youngme Moon and Clifford Nass. 1996. How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication research* 23, 6: 651-674.

[21] Jennifer H. Pfeifer, Marco Iacoboni, John C. Mazziotta, and Mirella Dapretto. 2008. Mirroring others' emotions relates to empathy and interpersonal competence in children. *Neuroimage* 39, 4: 2076-2085.