Modalities for Automated Health Counseling by Conversational Agents

Prasanth Murali*, Teresa O' Leary*, Ameneh Shamekhi, Timothy Bickmore

Northeastern University

Boston, MA

prvrasin, tkoleary, ameneh, bickmore @ ccs.neu.edu

1 INTRODUCTION

Several conversational agents have now been developed for automated health education and health behavior change counseling, in areas as diverse as exercise promotion [1], substance abuse counseling [2], and chronic disease self-care management [3]. For practical reasons, most have been deployed as animated characters rather than robots, and for safety reasons most have used fully constrained user input so that health advice can be validated [4].

We are currently exploring alternative input and output modalities for health counseling agents to determine whether they have significantly different impacts on user perceptions and health outcomes, assuming the other limitations described above can be addressed. Related results in other areas suggest that robotic embodiments can lead to a greater sense of presence and engagement compared to animated agents [5, 6]. Our own pilot studies also revealed a user preference for unconstrained speech input.

In this paper, we discuss preliminary results from a study investigating alternative input and output modalities for a health counseling agent in the domain of breastfeeding promotion. Our goal is to find modalities that lead to the greatest user engagement with and trust in the agent, as well as the greatest health outcomes. Breastfeeding is an important public health topic: many major US public health and medical organizations have been actively promoting breastfeeding and recommend exclusive breastfeeding for the first six months of life [7]. Despite these recommendations, only 16.3% of mothers in the US follow this recommendation, leading to a range of interventions to improve breastfeeding rates [8].

2 RELATED WORK

In this section, we outline studies comparing different embodiment and input modalities for conversational agents in health counseling.

2.1 Agent Embodiment Modality Studies

Bainbridge et al., (2011) [9] state that users respond to an unusual request and afford personal space to a robot than a live video feed of the robot doing the same tasks. Fasola and Mataric [10] show that older adults prefer a physical robot over a virtual coach in terms of enjoyableness, helpfulness and social interaction in an interactive intervention. Kidd, et al., (2008) [5] show participants in a weight loss program interacting with a touch screen with a physical head that can see and talk to the user continue with the weight-loss program for twice as long to a similar touch-screen-only device. Powers et al. [11], in a study on conversations about basic health habits, where users interact with between robots and similar computer-simulated agents that engaged participants in a conversation about found the robots were rated as more helpful, more lifelike, and possessing more positive personality traits than the computer-based agents. Jung et al. [12], also show physical embodiment as a positive in the feeling of social presence of robots and the importance of tactile communication in human robot interaction.

2.2. User Input Modality Studies

Bickmore, et al. [13], compared restricted user input against allowing users to express unconstrained feelings. Investigators found that an agent with greater empathetic accuracy was more effective at comforting users even at the cost of restricting user input. Suggesting that users ultimately preferred accurate agent reactions to their constrained input versus general agent reactions to unconstrained input.

Bickmore, et al. [4], investigated the potential safety issues that can arise when consumers use unconstrained speech input to ask conversational assistants non-trivial everyday medical questions. In this study, questions asked using unconstrained speech input led users to take actions that could have resulted in harm or death. The results of this study indicate that unconstrained speech input when used for consumer facing health systems could have harmful effects and could negatively impact user health outcomes. Thus, one must carefully consider and assess the tradeoffs when designing input modalities for therapeutic agents in healthcare domains.

3 MODALITY COMPARISON STUDY

In order to evaluate various input and output modalities for a health counseling agent, we developed five variants of a conversational agent originally developed for breastfeeding promotion. The agent (Tanya) was developed to play the role of a virtual lactation educator. We designed a 20-minute interaction intended to motivate women in their third trimester to follow the US Centers for Disease Control (CDC) recommendations. The topics covered in the intervention dialogue include: greeting; asking the user about her most important breastfeeding topic; review of the CDC

*These authors contributed equally to the work.

recommendations; review of the benefits of breastfeeding for the baby; breastfeeding benefits for the mother; breastfeeding "101" (latching, basic nursing positions); another review of the CDC recommendations; and farewell.

We evaluated the following five modality variants in a between-subjects experiment:

Agent Touch – Users interact with a virtual animated character (Figure 1) ("agent") by selecting what they want to say on a touch screen displaying constrained menu options.

Agent Constrained Speech – Users interact with the agent and select what they want to say by choosing one of the displayed constrained menu options and saying it out loud, in a Wizard Of Oz("WoZ") experiment [14].

Robot Touch – Users interact with Furhat [15] (a human like robot head) (Figure 2) by selecting what they want to say on a touch screen displaying constrained menu options.

Robot Constrained Speech – Users interact with Furhat by saying one of the constrained menu options out loud in a WoZ setup.

Robot Unconstrained Speech – Users interact with Furhat by expressing their feelings in an unconstrained way in a WoZ setup.

We originally designed this study to evaluate effects of input (touch vs. constrained speech) and output (virtual agent vs robot) modalities on user satisfaction and trust in health-related conversation on breastfeeding promotion. Initial qualitative findings indicated participants' desire to be able to talk to the system in an unconstrained manner, especially in the robot condition. This finding prompted us to add the fifth arm (robot unconstrained speech) to investigate whether allowing an open conversation, where there is no restriction on what the user says to the robot, would affect their satisfaction in this context.

Figure 1: Agent

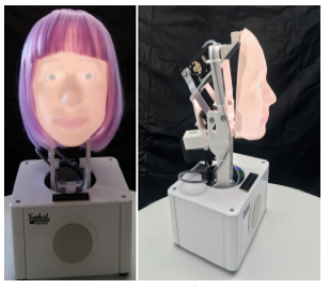

Figure 2: Furhat

3.1 Methods

Measures. We were primarily interested in evaluating user satisfaction and engagement with the agent. In addition to several single scale item measures of satisfaction, we used the bond subscale of the Working Alliance Inventory ("WAI") [16] to assess user feeling in working with the agent. WAI contains 12 items on a seven-point Likert scale (1=strongly disagree to 7=strongly agree) to measure participants' interaction with the system. Table 1 shows WAI questions.

Table 1. Working Alliance Inventory (WAI) items

WAI Questions
I feel uncomfortable with Tanya
Tanya and I understand each other
I believe Tanya likes me
I believe Tanya is genuinely concerned about my welfare
Tanya and I respect each other
I feel that Tanya is not totally honest about her feelings toward me
I am confident in Tanya's ability to help me
I feel that Tanya appreciates me
Tanya and I trust one another
My relationship with Tanya is very important to me
I have the feeling that if I say or do the wrong things
Tanya will stop working with me and I feel Tanya cares about me even when I do things that she does not approve of

Participants. 47 subjects have participated to date. Subjects were required to show interest in having children some day and able to read and speak English. They were recruited through flyers and advertisements on craigslist and compensated for their time. Participants were all female, aged 18-35.

Protocol. During the laboratory session, participants filled out a demographic questionnaire and completed all baseline questionnaires. Next, participants were introduced to either the agent or the robot and instructed on how to use the input modality, specific to their condition.

3.2 Quantitative Results

The study is ongoing. Here we report preliminary findings from the first 47 participants.

A one-way ANOVA across all 5 modality variants indicated that there were no significant differences among them on any measure (e.g., for WAI, $F(4,42)=2.0$, n.s.). Testing on robot vs. agent variants similarly revealed no significant differences (e.g., for WAI, $F(1,45)=.007$, n.s.).

However, a one-way ANOVA across the three user input modalities (ignoring differences in agent embodiment) did yield a significant difference on WAI, $F(2,44)=3.9$, $p<.05$). LSD post-hoc tests indicated that constrained speech resulted in significantly lower WAI scores compared to the other two conditions, and the other two conditions were nearly equivalent on this measure. The lower WAI score as well as the lower satisfaction score (not significant though) in the constrained speech condition compared to the touch input and unconstrained speech led us to hypothesize that the constrained speech system was in the uncanny valley [17]. We will describe this more in the discussion section.

3.3 Qualitative Results

Immediately following the interaction with the agent or robot participants were asked to share their overall thoughts and impressions during a semi-structured interview.

Initial findings from the qualitative analysis revealed that when users in the touch screen condition were asked "What would you change about the system?" participants repeatedly

requested a desire to use speech input, indicating that speech input could positively impact satisfaction.

[P12 Robot Touch] "It would have been better if it was only an interaction with her. Where she could understand what I was saying and then she could answer it. Obviously, that wouldn't be possible with everyone because different people have accents. Different ways of saying yes or no. So maybe she isn't that advanced. But yes, that's one place that I think would probably improve the interaction. That would make it 1-1 kind of a thing."

[P3 Robot Touch] " You could open up and ask more questions. I want to be able to talk to her."

However, participants in the constrained speech condition expressed frustration and disappointment. Thus, constrained speech did not lead to greater satisfaction but instead increased users' expectations of the agent's functional capabilities as well as highlighted the conversational scaffolding that inhibited user expressivity.

[P24 Robot Constrained Speech] "I think just providing more ways to interact with her. It seemed very much like a lecture like her kind of even though you could ask questions it's not like it could be that much of a conversation. I had to just stick to the script. If there was anything I would change it would be to just talk to her. But it would be difficult because at the end of the day Tanya is a robot."

[P33 Robot Constrained speech] "I liked the system overall only one frustrating part of the system was after she did finish saying her sentence I actually had to wait for a second to actually look and see the pop-up box and then say okay. It'd be nice to just talk."

4 DISCUSSION

The lower WAI scores in the constrained speech condition compared to the other two input modalities (touch input and unconstrained speech input), may be attributable to the uncanny valley effect, a theoretical framework proposed by Mashiro Mori [17]. This theory states that the reaction of a user to a humanlike robot abruptly shifts to a state of revulsion when the robot approaches, but fails to attain, a lifelike appearance. This descent is termed as the uncanny valley effect. We hypothesize that a similar "conversational uncanny valley" effect holds as users progress for very constrained forms of interaction to fully unconstrained speech input.

In our experiment, we compared three points along the conversational naturalness continuum, from fully constrained touch screen input, to constrained speech input, to fully unconstrained speech input. We hypothesize that user expectations of the overall conversational capabilities of the system increase as they progress along this continuum. Thus, one reason for their rejecting the constrained speech condition could be that it was at a point along the continuum that both raised their expectations (given the ability to speak) and then failed to meet them when they discovered that they could not say anything they wanted. Under these constraints, Tanya's inability to accept unconstrained speech resulted in revulsion.

Another explanation for these results is that people simply do not like learning or adhering to a limiting grammar for their speech[18].

This study suggests that touch input in an agent-based health counseling system performs equally as well as unconstrained speech input in increasing the user's perceived working alliance between user and agent. Whereas, constrained speech not only did not satisfy participants' desire to have a natural interaction but diminished their working alliance.

Multiple choice input to a dialogue system, while less natural and more constraining, is a widely accepted interface concept that functions in a manner consistent with participant expectations. Constrained speech, while providing necessary scaffolding so as to ensure system-level medically validated responses, seems to negatively impact users' sense of trust in the agent.

Despite overall acceptance of unconstrained speech input in this study, caution should be exercised in adopting such an input modality for health counseling agents. Unconstrained speech input could lead to errors, and in medical contexts these errors could be potentially harmful [4].

ACKNOWLEDGEMENTS

We thank Elise Masson for help running our user studies and Zachary Berwaldt for developing the animated character.

REFERENCES

[1] Bickmore, T., Silliman, R., Nelson, K., Cheng, D., Winter, M., Henaulat, L. and Paasche-Orlow, M. A Randomized Controlled Trial of an Automated Exercise Coach for Older Adults. *Journal of the American Geriatrics Society*, 61 (2013), 1676–1683.

[2] Zhou, S., Bickmore, T., Rubin, A., Yeksigian, K., Lippin-Foster, R., Heilman, M. and Simon, S. A Relational Agent for Alcohol Misuse Screening and Intervention in Primary Care. In *Proceedings of the CHI'17 Workshop on Interactive Systems in Healthcare (WISH)* (Denver, CO, 2017), [insert City of Publication],[insert 2017 of Publication].

[3] Kimani, K., Bickmore, T., Trinh, H., Ring, L., Paasche-Orlow, M. and Magnani, J. A Smartphone-based Virtual Agent for Atrial Fibrillation Education and Counseling. In *Proceedings of the International Conference on Intelligent Virtual Agents (IVA)* (Los Angeles, CA, 2016), [insert City of Publication],[insert 2016 of Publication].

[4] Bickmore, T., Trinh, H., Olafsson, S., O'Leary, T., Asadi, R., Rickles, N. and Cruz, R. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. *J Med Internet Res*, 20, 9 (2018).

[5] Kidd, C. and Breazeal, C. Effect of a robot on user perceptions. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)* (Sendai, Japan, 2004), [insert City of Publication],[insert 2004 of Publication].

[6] Wainer, J., Feil-seifer, D., Shell, D. and Matarić, M. Embodiment and human-robot interaction: A taskbased perspective. In *Proceedings of the IEEE International Conference on Robot & Human Interactive Communication, RO-MAN* (2007), [insert City of Publication],[insert 2007 of Publication].

[7] American Academy of Pediatrics (2005).

[8] Centers for Disease Control and Prevention Breastfeeding report card - United States, 2014 (2014).

[9] Bainbridge, Wilma A., Justin W. Hart, Elizabeth S. Kim, and Brian Scassellati. "The benefits of interactions with physically present robots over video-displayed agents." International Journal of Social Robotics 3, no. 1 (2011): 41-52.

[10] Fasola, Juan, and Maja J. Matarić. "A socially assistive robot exercise coach for the elderly." Journal of Human-Robot Interaction 2.2 (2013): 3-32.

[11] Kiesler, Sara, Aaron Powers, Susan R. Fussell, and Cristen Torrey. "Anthropomorphic interactions with a robot and robot–like agent." Social Cognition 26, no. 2 (2008): 169-181.

[12] Jung, Younbo, and Kwan Min Lee. "Effects of physical embodiment on social presence of social robots." Proceedings of PRESENCE (2004): 80-87.
[13] Timothy Bickmore and Daniel Schulman. "Practical approaches to comforting users with relational agents." In CHI'07 extended abstracts on Human factors in computing systems, pp. 2291-2296. ACM, 2007.
[14] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. "Wizard of Oz studies—why and how, Readings in intelligent user interfaces." (1998).
[15] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In Cognitive Behavioural Systems. Springer Berlin Heidelberg, 114-130.
[16] Horvath, A. and Greenberg, L. Development and Validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, 36, 2 (1989), 223-233.
[17] Mori, Masahiro. "The uncanny valley." *Energy* 7, no. 4 (1970): 33-35.
[18] Sidner, Candace L., and Clifton Forlines. "Subset languages for conversing with collaborative interface agents." In *Seventh International Conference on Spoken Language Processing*. 2002.