Getting to the point: Contrasting Directness and Warmth in Motivational Embodied Conversational Agents

Michael O'Mahony, Cathy Ennis, Robert Ross
School of Computer Science, Technological University Dublin, Ireland
michael.t.omahony@mytudublin.ie, {cathy.ennis, robert.ross}@tudublin.ie

Abstract

Enhancing long-term engagement with conversational agents remains a significant challenge. Controlling the perceived warmth or directness of an agent's personality through the style of its generated text could be used to increase user likeability. This paper reports an investigation of a Wizard-of-Oz (WoZ) mediated study of two variants of a motivational embodied conversational agent to measure user perception of and attitudes towards warmth in interaction style[1]. Results show a significant effect of users preferring an agent with a "more direct" personality for this scenario, though this effect is in many ways nuanced.

1 Introduction

There have been many advancements in data-to-text generation in recent years, especially through the use of neural networks (Lin et al., 2020; Su et al., 2021), and more recently, Large Language Models (LLMs). Most of this work has focused on content fidelity rather than text style (Lin et al., 2020; Su et al., 2021). However, LLMs have significantly improved our ability to style content. Since the release of ChatGPT in November 2022, dialogue systems have been applied to many more tasks, but the challenge of keeping a user engaged with an agent and understanding how the nuances of the agent can be tailored to enhance specific conversational goals, for example around motivation intervention, remains a very real research challenge. Moreover, the style of generated text can change the perceived personality of an agent and hence impact likeability and engagement.

Text style is an important aspect of generated text as a wide variety of applications require that information is given in a certain way. Considering that text generation plays a large role in the user satisfaction of a dialogue system (Peng et al., 2020), dialogue systems that aim to imitate a human agent can appear to have a consistent personality through a reliable, controllable style of conversation.

While style, personality and its relationship to engagement and likeability is of relevance to semantics and pragmatics study in general, in this paper we are particularly focused on the domain of Motivational Interview (MI) agents. MI is a counselling technique used to increase a person's motivation to change their behaviour. Some other studies have researched the impact of MIs delivered by intelligent agents on users.

We hypothesise that agent "warmness" vs "directness" will impact participant likeability leading to differences in responses to the general agent ratings (see section 3). We also hypothesise that, from these ratings, there will be a preference of one simulated personality over the other.

2 Related Work

The impact of personality variations in the healthcare domain has already been the subject of significant study. Many of these works look at agent empathy (Barange et al., 2022; Chauvin et al., 2023) but there are also works focused on other aspects such as humour (Olafsson et al., 2020) and adaptivity (Egede et al., 2021).

A few studies have employed an agent to deliver MIs to participants to increase their motivation to eat healthier (Olafsson et al., 2020, 2019), exercise more (Olafsson et al., 2020, 2019; Galvão Gomes da Silva et al., 2020, 2018; Chauvin et al., 2023), or quit unhealthy behaviours such as excessive alcohol consumption (Olafsson et al., 2023). Some of these studies used animated conversational agents (Olafsson et al., 2019, 2020; Chauvin et al., 2023), a NAO robot (Galvão Gomes da Silva et al., 2018), or video recordings of human actors (Galvão Gomes da Silva et al., 2020) to build the ECA.

[1]Data will be available at https://github.com/Michael-OMahony/getting-to-the-point-data/ after 01/01/2025.

3 Experiment Design and Methodology

To investigate the perception and impact of directness variation in MI agents, we conducted an online WoZ between-subjects user study to measure the likeability of the ECA. The interaction scenario was an MI delivered by a virtual agent to increase users' motivation to change their exercise behaviour. Participants interacted with the ECA via voice though mediated through an online interface. We recruited 25 participants from local communities. Participants were also given a questionnaire, and a number of concrete metrics were collected alongside recordings of the interactions.

The interview script was adapted from an earlier study that used a NAO robot to deliver an MI to participants to help increase their motivation to change their exercise habits (Galvão Gomes da Silva et al., 2018). While our experiment used a WoZ setup, in the original study the participant would control when the next utterance was delivered by pressing a button on the robot's head. The authors designed the script so that each question should make sense to the user, irrespective of how they answered previous questions. In practice this method mostly worked, but there were instances where a somewhat broad question lead to some confusion.

Building on the existing corpus, we created two conditions by altering the text style of parts of the original script using ChatGPT to create "warmer" and "more direct" versions of the agent script. In practice, we only changed the beginning and end of the script, aside from a minor change in the first question for clarity, we did not alter any of the questions as designing a counselling intervention was outside the scope of this work, and we believe the start and end of an interaction are influential on user satisfaction. There were no options to change the next utterance based on the participants responses, but we could repeat the last question upon request.

Participants answered a questionnaire before and after the interaction. The pre-interaction questionnaire included demographics, exercise frequency, Ten-Item Personality Inventory (TIPI) for the participant (Gosling et al., 2003), familiarity and attitudes towards virtual agents. The post-interaction questionnaire included the TIPI for the virtual agent (Gosling et al., 2003), general agent ratings (Olafsson et al., 2019, 2020), and an open-ended feedback box. The general agent rating questions were Q1:"I am satisfied with the agent", Q2:"I would continue talking with the agent", Q3:"I trust the agent", Q4:"I like the agent", Q5:"The agent was knowledgeable", Q6:"The conversation was natural", Q7:"I have a good relationship with the agent", and Q8:"I am similar to the agent". Participants rated the agent using a five point Likert scale.

As the focus of our work was on embodied agents rather than text or speech only based interaction, the agent was given a virtual appearance. For this we used the Unity game engine, along with a Ready Player Me avatar. Moreover, we used the Talking With Hands dataset for the talking gestures (Lee et al., 2019), Ready Player Me animation library for the idle animation, and Salsa Lip Sync. Google's Cloud AI text-to-speech was used for the agent's voice, where we selected a female avatar as some studies suggest that men slightly prefer a female therapist to a male one or do not care, and women are much more likely to prefer a female therapist (Liddon et al., 2018; Seidler et al., 2022). In general, each experiment lasted 20-30 minutes, with the interaction lasting 5-15 minutes.

4 Results

The interaction times between the "warmer" and "more direct" conditions were not statistically significant (means=680, 646s). Table 1 presents mean Likert Ratings for each of the key likeability questions. The means for the responses to every question were higher for the "more direct" condition though when analysed on a question by question basis, the only question which demonstrated statistically significant difference was Q7. Potential limitations were the sample size, and our inability to alter most of the agent script. Future work will focus on nuancing the qualities of directness and warmness in speech and embodying these in a more automated agent with evaluation of effectivness as well as engagement.

Q	Warm	Direct	Difference	Sig.
1	3.46	3.83	-0.37	0.4494
2	3.08	3.33	-0.26	0.5729
3	3.15	3.50	-0.35	0.4058
4	3.53	4.00	-0.46	0.0953
5	3.23	3.50	-0.27	0.5685
6	2.85	3.00	-0.15	0.5495
7	**2.85**	**3.42**	**-0.57**	**0.0379***
8	2.15	2.25	-0.10	0.8193

Table 1: Means and significance for each question.

Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue, September, 11–12, 2024, Trento, Italy.

5 Acknowledgements

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6183. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

Mukesh Barange, Sandratra Rasendrasoa, Maël Bouabdelli, Julien Saunier, and Alexandre Pauchet. 2022. Impact of adaptive multimodal empathic behavior on the user interaction. In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, IVA '22, New York, NY, USA. Association for Computing Machinery.

Rachel Chauvin, Céline Clavel, Nicolas Sabouret, and Brian Ravenet. 2023. A virtual coach with more or less empathy: impact on older adults' engagement to exercise. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, IVA '23, New York, NY, USA. Association for Computing Machinery.

Joy Egede, Maria J. Galvez Trigo, Adrian Hazzard, Martin Porcheron, Edgar Bodiaj, Joel E. Fischer, Chris Greenhalgh, and Michel Valstar. 2021. Designing an adaptive embodied conversational agent for health literacy: a user study. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, IVA '21, page 112–119, New York, NY, USA. Association for Computing Machinery.

Joana Galvão Gomes da Silva, David J Kavanagh, Tony Belpaeme, Lloyd Taylor, Konna Beeson, and Jackie Andrade. 2018. Experiences of a motivational interview delivered by a robot: Qualitative study. In *J Med Internet Res*.

Joana Galvão Gomes da Silva, David J. Kavanagh, Jon May, and Jackie Andrade. 2020. Say it aloud: Measuring change talk and user perceptions in an automated, technology-delivered adaptation of motivational interviewing delivered by video-counsellor. *Internet Interventions*, 21:100332.

Samuel D Gosling, Peter J Rentfrow, and William B Swann. 2003. A very brief measure of the big-five personality domains. *Journal of Research in Personality*, 37(6):504–528.

Gilwoo Lee, Zhiwei Deng, Shugao Ma, Takaaki Shiratori, Siddhartha Srinivasa, and Yaser Sheikh. 2019. Talking with hands 16.2m: A large-scale dataset of synchronized body-finger motion and audio for conversational motion analysis and synthesis. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 763–772.

Louise Liddon, Roger Kingerlee, and John A. Barry. 2018. Gender differences in preferences for psychological treatment, coping strategies, and triggers to help-seeking. *British Journal of Clinical Psychology*, 57(1):42–58.

Shuai Lin, Wentao Wang, Zichao Yang, Xiaodan Liang, Frank F. Xu, Eric Xing, and Zhiting Hu. 2020. Data-to-text generation with style imitation. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1589–1598, Online. Association for Computational Linguistics.

Stefan Olafsson, Teresa O'Leary, and Timothy Bickmore. 2019. Coerced change-talk with conversational agents promotes confidence in behavior change. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth'19, page 31–40, New York, NY, USA. Association for Computing Machinery.

Stefan Olafsson, Teresa K. O'Leary, and Timothy W. Bickmore. 2020. Motivating health behavior change with humorous virtual agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, IVA '20, New York, NY, USA. Association for Computing Machinery.

Stefan Olafsson, Paola Pedrelli, Byron C. Wallace, and Timothy Bickmore. 2023. Accomodating user expressivity while maintaining safety for a virtual alcohol misuse counselor. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, IVA '23, New York, NY, USA. Association for Computing Machinery.

Baolin Peng, Chenguang Zhu, Chunyuan Li, Xiujun Li, Jinchao Li, Michael Zeng, and Jianfeng Gao. 2020. Few-shot natural language generation for task-oriented dialog. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, page 172–182, Online. Association for Computational Linguistics.

Zac E. Seidler, Michael J. Wilson, David Kealy, John L. Oliffe, John S. Ogrodniczuk, and Simon M. Rice. 2022. Men's preferences for therapist gender: Predictors and impact on satisfaction with therapy. *Counselling Psychology Quarterly*, 35(1):173–189.

Yixuan Su, David Vandyke, Sihui Wang, Yimai Fang, and Nigel Collier. 2021. Plan-then-generate: Controlled data-to-text generation via planning. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, page 895–909, Punta Cana, Dominican Republic. Association for Computational Linguistics.