Edinburgh Research Explorer

Start Making Sense: Predicting confidence in virtual human interactions using biometric signals

Start Making Sense: Predicting confidence in virtual human interactions using biometric signals.

S. Dalzel-Job[1], R.L. Hill[1], R. Petrick[2]

1 Department of Informatics, University of Edinburgh, Edinburgh, Scotland. sdalzel@ed.ac.uk, r.l.hill@ed.ac.uk

2 School of Mathematical & Computer Sciences, Heriot-Watt University, Edinburgh, Scotland. R.Petrick@hw.ac.uk

Introduction and Aims

This project investigates the use of biometric data to predict confidence levels during task-focused interaction between humans and virtual humans. The project comprises of two main studies, the first of which examines the relationship between biometric signals – galvanic skin response (GSR), heart rate, facial expression and eye movements – and self-report levels of confidence during a task-oriented interaction between a human and a virtual human. Through the manipulation of the feedback and task demands, participants were exposed to unexpected situations and varying levels of ambiguity, resulting in a measurable range of perceived confidence as well as more implicit biometric and behavioural indicators of confidence and success. The second study utilises the paradigm and results from Experiment 1 to train an AI instruction giver to identify instances where behavioural and biometric feedback from a human signal low confidence, enabling it to modify or supplement its instructions accordingly. To ensure that the AI is acting in a useful way, and that the experimental manipulation and behavioural demonstrations of confidence are valid, the participant judges the perceived success of the interaction, as well as their trust in the AI under varying levels of feedback. This paradigm can then be adapted for use across a wide range of situations and scenarios; from interactions with virtual human avatars or agents via AR, VR, desktop or mobile devices, to fully embodied conversational agents or robots, this paradigm will enable a successful, smooth interactions between humans and Ais.

Background

Virtual humans – whether computer-controlled agents or human-controlled avatars – are widely used during online interactions. Not only are they utilised during social interaction (e.g. gaming), but also in important joint-action or task-oriented communication. Historically, virtual humans have been used in support and health [1-5], as well as in areas such as teaching and training [6-10]. To date, there has been a wealth of research into how virtual humans should behave during interactions with users in order to maximise success [11-16]. Our previous research has discovered that the optimum behaviour of a virtual human, specifically its eye movements, varies depending on the purpose of the interaction [15, 16]. This study aims to expand and extend these findings by investigating which combination of behaviours maximise positive perceptions of a virtual agent, as well as maximising any given task performance, with the aim of developing trustworthy, likeable and useful virtual humans. Furthermore, it aims to develop a system that can utilise real-time non-verbal feedback from a user to indicate confusion or occasions of uncertainty. This will enable the system to supplement or alter instructions to maximise the possibility of a smooth, successful interaction.
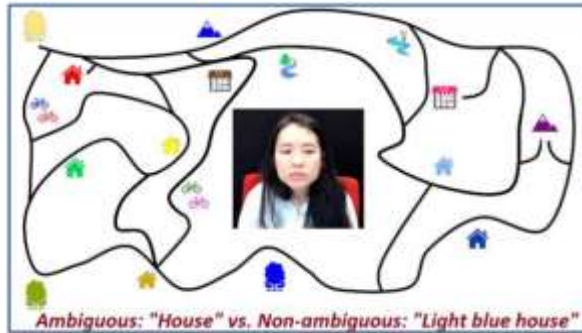
Methods

Experiment 1

Figure 1:The human follower looks at the Landmark when she has located it.

Figure 2: The speaker guides both a human and a virtual human.

A human speaker guides a listener through the map to locate a landmark. Each speaker guides both a human listener and a virtual human listener (Figure 1 and Figure 2, respectively). The listener may indicate that the target has been successfully located by looking at it (correct condition) or that an incorrect target was chosen (wrong condition). It can be seen in Figure 2 that the listener has not found the correct target, the dark blue house, but is instead looking at the other side of the map. Furthermore, the speaker may have insufficient information to uniquely identify a target, resulting in them having to choose between multiple possible choices to guide the listener towards. In Figure 1, for example, the target landmark given to the speaker may be 'House', but there is more than one house on the map, leading to an ambiguous situation where the speaker has to decide which house to guide the listener to. These manipulations deliberately generate situations of uncertainty or ambiguity. The speaker believes that the virtual human listener is either controlled by a human (avatar condition) or by a computer "AI" (agent condition). In all conditions, the listener is actually a video, and is non-interactive, although results from the study suggest that this was not identified by the participants, and that they treated the listener like an interactive virtual human.

The amount of time the listener looks at the user is also systematically varied. It has been found that the optimum amount of looking by a virtual human at a human can vary, depending on the purpose of the interaction [16]. The previous research examined the impact of looking at the user 0%, 25%, 75% and 100% of the interaction; this was adapted slightly in the current study, with the listener looking at the user during either 0%, 30% or 70% of the interaction. This was intended to identify more about the effects on the user of the listener's looking behaviour; some research has suggested that there may be a threshold amount of looking, at around 70%, at which point the social impact on the user is at its highest [16].

Measures

The users respond to questions relating to their confidence in their instructions, and they also report if they were confident that the listener found the target. They are also asked questions relating to their social perceptions of the listener.

The eye movements of the users are recorded using an Eyetribe remote eye tracker [17]. This particular device was chosen for its non-invasive nature, and its portability. Galvanic skin response (GSR) is measured using two Shimmer sensors attached to the tips of two fingers, and another sensor attached to a third finger to detect changes in heart rate [18]. Changes in GSR and heart rate can indicate a change in arousal; these changes could be positive or negative in valence (it could indicate joy or anger, happiness or frustration, for example, but in isolation the GSR data does not allow you to identify which). The facial expressions of the users are also detected during the interaction [20], allowing for a fuller understanding of the nature of the arousal. An

indication of an angry facial expression in conjunction with a large GSR peak, for example, is more informative about the effect of any stimuli on a user than the GSR alone.

iMotions is software that allows the presentation of the stimuli and collation, time-stamping and processing of all the behavioural, biometric and survey data in preparation for analysis [19]. Examining these behaviours and biometrics together rather than independently allows the identification of the behaviour, or combination of behaviours associated with varying levels of confidence, as indicated by their relationship with the responses to the survey. This combination of objective and subjective measures enables us to begin to build a model of how users respond and adjust to different types of feedback, and to use this information to design behaviourally appropriate virtual humans, responding in real-time to non-verbal feedback that may indicate anything other than a smooth interaction.

Experiment 2

The non-verbal behaviours identified in Experiment 1, which are associated with confidence – in self and in the interlocutor – can be used by a planning system to identify instances of confusion, or where the user may require extra information. The facial expressions and eye movements are fed into the planning system in real-time, and upon breaching a pre-specified threshold, the system responds accordingly, signalling the system to provide extra information where low confidence or confusion is indicated, and continuing without additional clarification when biometric responses suggest that the interaction is going well.

Outcomes and Applications

This research can be applied to several different situations: wherever it is desirable for a system to respond in real-time to a user's emotional state, the system can be trained to identify signals of confusion or uncertainty and respond immediately to remedy the situation. With the advent of more mobile eye trackers and the increasing popularity and affordability of smart watches, as well as other devices that already detect heart rate, which could potentially be developed to identify changes in GSR, this paradigm presents the possibility of interactive systems responding in real-time to behavioural and biometric cues provided via our everyday devices. Interactive and ubiquitous, virtual companions, advisors, teachers, coaches or even mediators could soon be available to provide customisable, interactive, responsive and truly trustworthy, effective virtual humans.

Ethical Statement

Ethics approval for this study was granted by the Informatics Ethics Committee, University of Edinburgh (rt #3690),

References

1. Robillard, G., et al., (2010). Using virtual humans to alleviate social anxiety: preliminary report from a comparative outcome study. *Stud Health Technol Inform* **154**: 57-60.

2. Kang, S.-H., et al. (2008). Does the contingency of agents' nonverbal feedback affect users' social anxiety? in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1. International Foundation for Autonomous Agents and Multiagent Systems*.

3. Lok, B., (2006). Teaching communication skills with virtual humans. *IEEE Computer Graphics and Applications* **26(3)**: 10-13.

4. Yuen, E.K., et al., (2013). Treatment of social anxiety disorder using online virtual environments in second life. *Behavior therapy* 44(1): p. 51-61.

5. Kenny, P., et al. (2007). Virtual patients for clinical therapist skills training. in *International Workshop on Intelligent Virtual Agents*. Springer.

6.   Kim, Y., J. Thayne, and Q. Wei (2017) An embodied agent helps anxious students in mathematics learning. *Educational Technology Research and Development* **65**(1): 219-235.

7.   Johnson, W.L. and J. Rickel, (1997) Steve: An Animated Pedagogical Agent for Procedural Training in Virtual Environments. *Sigart Bulletin* **8(1-4)**: p. 12-16.

8.   Johnson, W.L., J.W. Rickel, and J.C. Lester (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education* **11**: p. 47-78.

9.   Johnson, W.L., et al. (2003). Evolution of User Interaction: The Case of Agent Adele in *8th International Conference on Intelligent user interfaces* (Miami, Florida, USA).

10.   Rickel, J. and W.L. Johnson (1999). Virtual humans for team training in virtual reality in *Proceedings of the ninth international conference on artificial intelligence in education*.

11.   Cassell, J., J. Sullivan, and S.e. Prevost (1999). Embodied Conversational Agents, ed. M. Cambridge.: MIT Press.

12.   Cassell, J. and K.R. Thorisson, (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* **13(4-5)**: 519-538.

13.   Dalzel-Job, S., C. Nicol, and J. Oberlander (2008). Comparing behavioural and self-report measures of engagement with an embodied conversational agent: A first report on eye tracking in Second Life. in *The 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia)

14.   Dalzel-Job, S., J. Oberlander, and T.J. Smith (2011) Contested staring: issues and the use of mutual gaze as an on-line measure of social presence.

15.   Dalzel-Job, S., J. Oberlander, and T.J. Smith (2011). Don't Look Now: The relationship between mutual gaze, task performance and staring in Second Life. in *The 33rd Annual Conference of the Cognitive Science* (Boston, Massachusetts, USA).

16.   Dalzel-Job, S., (2015). Social interaction in virtual environments: the relationship between mutual gaze, task performance and social presence.

17.   Eyetribe. *The Eyetribe*. [cited 2020 30.01.2020]; Available from: https://theeyetribe.com/theeyetribe.com/about/index.html.

18.   *Shimmer*. [cited 2020 07/02/2020]; Available from: https://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor.

19.   *iMotions*. [cited 2020 07/02/2020]; Available from: https://imotions.com/.

20.   *Affectiva*. [cited 2020 07/02/2020]; Available from: https://www.affectiva.com/.