Real-time testing of non-verbal interaction: An experimental method and platform

Tom Gurion, Patrick G.T. Healey and Julian Hough
Cognitive Science Research Group
School of Electronic Engineering and Computer Science
Queen Mary University of London
{t.gurion,p.healey,j.hough}@qmul.ac.uk

Abstract

We present an immersive multi-person game developed for testing models of non-verbal behaviour in conversation. People interact in a virtual environment using avatars that are driven, by default, by their real-time head and hand movements. However, on the press of a button each participant's real movements can be substituted by 'fake' avatar movements generated by algorithms. The object of the game is to score points in two ways a) by faking without being detected and b) by detecting when others are faking. This enables what amounts to a non-verbal Turing test in which the effectiveness of different algorithms for controlling non-verbal behaviour can be directly tested and evaluated in live interaction.

1 Introduction

Experimental studies of conversation have primarily focused on verbal exchange, though it is now widely recognised that non-verbal communication is important for successful interaction. For example, listeners gesture to demonstrate attention to a speaker (Goffman, 1955) and their readiness to take the floor (Hadar et al., 1985); mutual eye-gaze, or its absence, affects speech fluency (Goodwin, 1979) and when listeners fail to provide timely and appropriate concurrent feedback, a speaker's performance is disrupted (Bavelas et al., 2000). Currently there is a paucity of experimental approaches for studying these processes.

Recently, research using virtual reality (VR) technologies has begun to address this need. VR can eliminate the need for confederates that are otherwise common in studies of social interaction, and are known to be problematic (Kuhlen and Brennan, 2013). It can also be used to test scenarios that are hard (e.g. physical danger) or impossible (e.g. body transfer) to recreate in the lab (Pan and Hamilton, 2018). VR studies are also increasingly easy to reproduce. They often rely on commonly available hardware, and standard software components can support most, if not all, of the basic experimental procedures and are easy to share. In addition, the VR application can log all movement information directly for further analysis (Fox et al., 2009).

One issue, common to many experimental studies of interaction, is the strategy of restricting the conversation to obtain greater experimental control; for example assigning the speaker and listener roles in advance or using restricted tasks (Bailenson and Yee, 2005; Gratch et al., 2007; Hale and Hamilton, 2016). This strategy makes it easier to isolate the effects of a manipulation and can provide simple outcome measures. A second issue is the measures of the effects of manipulating avatar behaviours are typically indirect. For example, asking participants to retrospectively rate the friendliness or persuasiveness of an agent on a Likert scale. One difficulty here is that there are known dissociations between what people say about their own (and other's) behaviour and the factors influencing those behaviours (Nisbett and Wilson, 1977; Haidt, 2001).

This paper describes a method and associated software platform that can more effectively leverage the potential of VR for testing models of non-verbal interaction. Building on previous work on intervening manipulations of live text-based dialogue (Healey et al., 2003) and live graphical interaction (Healey

Figure 1: A view of the virtual environment.

et al., 2002), this approach involves free interaction but still provides a high level of control over the experimental manipulation. Importantly, a game element is introduced that ensures continual real-time testing of the effectiveness of each manipulation of non-verbal behaviour.

2 The System

The system is inspired by standard social VR applications (Wallis, 2016). It allows groups of remote users to interact in the same virtual environment. However, users can also press a button that initiates automatic algorithmic control over their avatar's movements. This behaviour is presented to the users as "faking attention". During faking users can engage in other activities, while their avatar continues to present socially appropriate responses. Importantly, participants are encouraged to detect when other people faking and, if they accuse them correctly are awarded points. This creates a situation in which we can make direct experimental tests of different models of non-verbal behaviours, implemented as alternative algorithms for controlling the avatars.

The system is implemented on standard commercial hardware (HTC Vive[1]) which combines a head mounted display and two hand-held controllers. These components are tracked in 3-dimensional space to recreate live head and hand movement in the virtual environment. The microphone and headphones' connection on the headset are used for a voice chat between the users. The system animates mouth movement directly from speech to compensate for the lack of actual tracking and to help players to identify the current speaker. The main application, consisting of a server and game clients, is developed in Unity3D,[2] a game engine commonly used to create VR experiences.[3]

A game context is used to incentivise participants through a scoring mechanism. Participants see their own score in a floating message in front of them. When they fake attention a 'Snake' game[4] pops up above the floating message. Collecting a snake's food pellet increases the player's score by one point. Another way to get points is by accusing other players for faking. A correct accusation is worth one point, but an incorrect accusation loses a point. The specific moments when points are accumulated provide a fine-grained assessment of how effective each faking period is.

Players start faking by pressing and holding a button on the left hand-held controller with their index finger. While faking a model of non-verbal listening behaviour takes control over the player's avatar, making player's real behaviour invisible to the rest of the group. Fakers are also muted from the chat so they hear everything but are cannot take part in the conversation. While faking, the joystick like button for the left thumb is used to control the snake game. Players accuse each other of faking by looking at them and using a button on the right hand-held controller. Note that there is no need to point at players

[1] https://www.vive.com/uk/

[2] https://unity3d.com/

[3] The source code for the system is open and available online at https://github.com/Nagasaki45/UnsocialVR. A video demonstrating the environment can be found at https://youtu.be/OOp1pARFM8I.

[4] https://en.wikipedia.org/wiki/Snake_(video_game_genre)

to accuse them, as this "pointing and shooting" gesture might interfere with the social dynamics.

Figure 1 shows the avatars design. They are cartoon-ish gender-neutral head and hands figures, similar to those use in commercial social VR products like Facebook Spaces (Tauziet, 2017).[5]

3 Possible Applications

This system enables new experimental approaches to a variety of questions in non-verbal interaction. For example, backchannel responses, the concurrent head nods and "uh-huh" utterances produced by listeners during speaker turns (Yngve, 1970), have been modelled in different theories. Some models use a single feature, like speech prosody, and a set of simple rules to predict backchannels (Ward and Tsukahara, 2000). Others combine more features, including the speaker's head movement (Gratch et al., 2007), speaker-listener eye contact, or even the speaker's smile (Huang et al., 2011). Most of these studies, however, evaluate their models on corpus data. The approach we introduce here enables direct causal tests of the relative effectiveness of each model.

Similarly, there are a number of different predictions about where side-participants should look in multi-party conversation. Some studies suggest a side-participant is equally likely to look at the speaker as to look at the addressee (Healey et al., 2013); others suggest that side-participants usually gaze towards the speaker (Fujie et al., 2009). Another possibility is that side-participants follow the speaker gaze. These alternative hypotheses can be directly tested using the approach described here.

4 Discussion

While this method opens up new possibilities, it also has limitations. First, social interaction in VR might be significantly different from face-to-face conversations. This is essentially an empirical question and the answer will change as the capabilities of the technology change. We note however that social VR is an increasingly important mode of communication in its own right (Wallis, 2016). Studying communication in social VR might help us understand and build better virtual agents and environments even if it does not reliably generalise to the physical world.

A contingent limitation of the current system is that it uses data from specific hardware with specific capabilities: tracking a head mounted display and two hand-held controllers in 3-dimensional space. This implies that only behaviours that are tracked by the system can be generated by the models and checked for their credibility. For example, facial expressions, eye gaze, fine fingers movement, and torso pose, are not tracked by the system, and cannot be tested. More advanced sensing hardware, however, might improve this in the future.

Finally, we found that theories are often underspecified. Implementing computational models for these introduce subtle complications. For example, studies of backchannel responses often concentrate on triggering the response in the correct timing but doesn't describe the response itself. Subtle differences in head nods, for example, might have different interactional functions (Hadar et al., 1985).

5 Conclusion

We have presented a system for comparing models of non-verbal behaviour, suggested example applications and highlighted some limitations. This system provides several benefits compared to existing methods and practices in the field of multi-modal communication research. It can be used to test non-verbal models of communication in natural social interaction, without restricting the conversation. The credibility of the models is assessed by the participants during the interaction (as opposed to post-experiment questionnaires), based on direct perceived-plausibility ratings. Lastly, it provides easy means to compare competing models.

References

Jeremy N Bailenson and Nick Yee. 2005. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological science*, 16(10):814–819.

[5] `https://www.facebook.com/spaces`

Janet Beavin Bavelas, Linda Coates, and Trudy Johnson. 2000. Listeners as co-narrators. *Journal of personality and social psychology*, 79(6):941–952.

Jesse Fox, Dylan Arena, and Jeremy N Bailenson. 2009. Virtual reality: A survival guide for the social scientist. *Journal of Media Psychology*, 21(3):95–113.

Shinya Fujie, Yoichi Matsuyama, Hikaru Taniyama, and Tetsunori Kobayashi. 2009. Conversation robot participating in and activating a group communication. In *Tenth Annual Conference of the International Speech Communication Association*.

Erving Goffman. 1955. On face-work: An analysis of ritual elements in social interaction. *Psychiatry*, 18(3):213–231.

Charles Goodwin. 1979. The interactive construction of a sentence in natural conversation. *Everyday language: Studies in ethnomethodology*, pages 97–121.

Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *Intelligent Virtual Agents*, pages 125–138. Springer.

Uri Hadar, Timothy J Steiner, and F Clifford Rose. 1985. Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, 9(4):214–228.

Jonathan Haidt. 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review*, 108(4):1024–1046.

Joanna Hale and Antonia F de C Hamilton. 2016. Testing the relationship between mimicry, trust and rapport in virtual reality conversations. *Scientific reports*, 6.

Patrick G.T. Healey, Nik Swoboda, Ichiro Umata, and Yasuhiro Katagiri. 2002. Graphical representation in graphical dialogue. *International Journal of Human-Computer Studies*, 57(4):375–395.

Patrick G.T. Healey, Matthew Purver, James King, Jonathan Ginzburg, and Greg J Mills. 2003. Experimenting with clarification in dialogue. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 25, pages 539–544. LEA.

Patrick GT Healey, Mary Lavelle, Christine Howes, Stuart Adam Battersby, and Rosemarie McCabe. 2013. How listeners respond to speaker's troubles. In *CogSci*, pages 2506–2511.

Lixing Huang, Louis-Philippe Morency, and Jonathan Gratch. 2011. Virtual rapport 2.0. In *Intelligent Virtual Agents*, pages 68–79. Springer.

Anna K Kuhlen and Susan E Brennan. 2013. Language in dialogue: when confederates might be hazardous to your data. *Psychonomic bulletin & review*, 20(1):54–72.

Richard E Nisbett and Timothy D Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological review*, 84(3):231.

Xueni Pan and Antonia F de C Hamilton. 2018. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*.

Christophe Tauziet. 2017. Designing facebook spaces (part 2) - presence & immersion. `https://medium.com/@christauziet/designing-facebook-spaces-part-2-presence-immersion-35eb3c96a4cc`.

Thomas Wallis. 2016. What is social vr? `https://www.vr-intelligence.com/social-vr-101`.

Nigel Ward and Wataru Tsukahara. 2000. Prosodic features which cue back-channel responses in english and japanese. *Journal of pragmatics*, 32(8):1177–1207.

Victor H Yngve. 1970. On getting a word in edgewise. In *Chicago Linguistics Society, 6th Meeting*, pages 567–578.