Non-verbal Communication for Correlational Characters

Marco Gillies and Mel Slater
Department of Computer Science, University College London, London WC1E 6BT, UK
{m.gillies, m.slater}@cs.ucl.ac.uk

Abstract

Social interaction is a key element of modern virtual environments. This paper discusses how non-verbal communication (or body language) is vital to real world social interaction, and how it is important to carry it over to virtual environments. It is not sufficient for a character to passively exhibit non-verbal communication; non-verbal communication should be a genuine interaction between a real and virtual person. To this aim the behaviour of the character should correlate realistically with that of the real person. We hypothesise that this sort of correlational non-verbal behaviour enhances presence and outline work in progress to investigate this hypothesis. We present a virtual character that exhibits this type of correlational behaviour in an immersive virtual environment.

1. Introduction

Perhaps the most interesting virtual environments for participants are social ones, where participants commonly share the VE, both with other real people, each represented by their own graphical character, or avatar, and with completely virtual people, that are entirely computer controlled. Since humans are social animals these other inhabitants of the virtual environment become a focus of interest, and VEs become a venue for social interaction. This means that such social interaction is a vitally important issue for presence research.

Though most social interaction among humans takes the form of conversation, there is a large sub-text to any interaction that is not captured by a literal transcription of the words that are said. Tone of voice can transform the meaning of a statement from angry, to sarcastic or playful. Posture can indicate keen engagement in the subject of discussion or bored disengagement, by leaning forward or slumping in a chair. Gestures can help clarify a path to be taken when giving directions. Facial expression can be smiling, and encouraging or indicate displeasure at what is being said. How close people stand to each other can indicate a lot about their relationship.

All of these factors go beyond the verbal aspects of speech and are called Non-Verbal Communication (often referred to by the popular term "body language"). Non-Verbal Communication (NVC) is a key element of human social interaction. Certain aspects of communication such as the expression of emotion or of attitude toward, and relationship, with other people are much more readily expressed non-verbally than verbally. Communication that lacks non-verbal elements can be limited and ambiguous, as demonstrated by the problems of interpreting the emotional tone of emails. In particular virtual characters that do not display NVC during conversation are less likely to be judged as realistic or to elicit presence.

However, it is not enough to display realistic postures, gestures, facial expressions etc, if these do not represent a genuine interaction with participants. In a recent review of the literature Sanchez-Vives and Slater[14] defined presence in a VE as successful replacement of real by virtually generated sensory data. Here 'successful' means that the participants respond to the sensory data as if it were real, where response is at every level from physiological through to cognitive. One element in this is the response of the environment to behaviours of the participant, and suggests that one of the most important factors in eliciting presence is form of interaction, particularly whole body, natural interaction. It is therefore important that social interaction occurs through natural bodily interaction, i.e. through NVC. This should be a true interaction, not merely a real and virtual human independently producing NVC.

Under what circumstances are people likely to find themselves responding to virtual characters as if they are real? Our hypothesis is that this would occur if the virtual characters respond to people *as if they are real*! Specifically what this means is that a kind of correlational dance is established in which actions of one person are refected in the actions of the other, which are reflected in the actions of the other, and so on. Moreover, people naturally attempt to find correlations between their own behaviour and that of their environment. This is particularly true of interaction with other people, people naturally interpret the behaviour of others in terms of their own actions and state. This occurs even when interaction with virtual characters whose behaviour is pre-recorded, and therefore is not related in any way[13] . This leads us to the Correlational Presence hypothesis, that presence is enhanced by producing this type of correlation between a person's behaviour and that of the VE, and will therefore be enhanced if correlations are included as part of the environment. This work focuses on correlational presence during social interaction with virtual characters. This entails creating characters who not only autonomously produce behaviour, but behaviour, and in particular NVC, that is correlated realistically with full body behaviour of real participants.

Thus we come to the central hypothesis of this paper: correlational NVC is a key determining factor for presence during social interaction with virtual characters, or mediated via avatars. The remainder of this paper describes current work in progress to test this hypothesis, and in particular to create characters that display correlational NVC. Our characters have been created to run in a Cave-like immersive virtual reality system[4] , which allows natural interaction with a life-size virtual character.

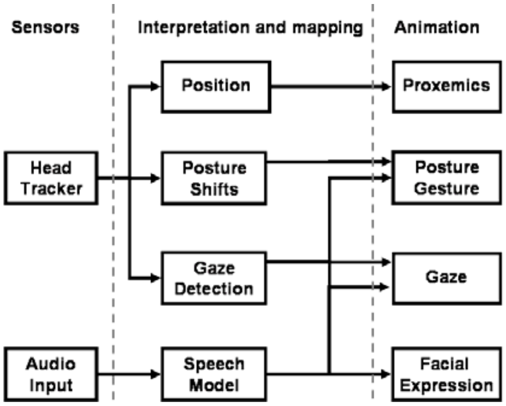

Figure 1: Mapping between sensor data and animation

Virtual characters require three basic elements in order to display NVC that correlates with a participant, as shown in figure 1. The first is an animation system that is able to generate realistic non-verbal behaviour, this is described in section 3. The character must also be able to sense the behaviour of the user. In our current system we have chosen to use the sensors commonly available in immersive virtual reality systems, particularly Cave-like systems. Thus we have restricted ourselves to a single head tracker, and to audio input via a microphone. In the future it would be interesting to look into more complex tracking systems, but this would reduce the general applicability of this work. Mediating between these two elements is a module that interprets the sensor data and maps the results to behaviour. The sensing, interpretation and mapping aspects of this work are described in section 4.

2. Non-verbal Communication

As described in the introduction non-verbal communication takes many forms, or modalities. Argyle[1] lists the following modalities of NVC: "facial expression; gaze (and pupil dilation); gestures, and other bodily movements; posture; bodily contact; spatial behaviour; clothes, and other aspects of appearance; non-verbal vocalizations, and smell". This work is restricted to modalities that involve bodily movements, avoiding non-bodily modalities such as vocalizations or smell, and static modalities such as appearance or clothing. We therefore use five main modalities: posture, gestures, facial expression, gaze and proxemics (spatial behaviour, personal space).

Our work on correlational NVC builds on a large body of work on animating NVC, for example Cassell *et al.* [3] , Guye-Vullième *et al.*[8] and Pelachaud and Bilvi[1] . We use the Demeanour framework[6] [7] to generate animated non-verbal behaviour. Demeanour consists of a number of animation modules that display the behaviour (described below), and a declarative behaviour language for specifying rules for what behaviour should be displayed. The behaviour language is used to specify mappings from input variables to output behaviour. The input variables come from sensing the user, and other contextual factors and described in section 4.The general aim of the behaviour generated is to give a generally favorable and friendly impression of our character (shown in figure 2). Thus most of the behaviour will display a generally friendly attitude towards the participant. The rest of this section will describe the modalities we use.

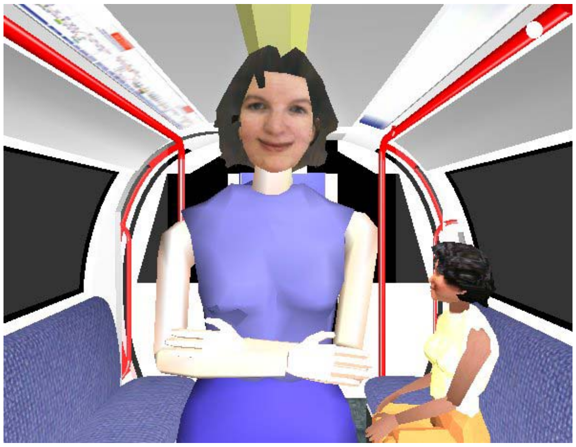

Figure 2: An example character

2.1 Posture and Gesture

Posture is the long-lasting static pose of the body whereas gestures are more transitory movements, mostly of the arms and head that commonly accompany speech. While people always have a posture, gestures are a purely conversational phenomenon, and seem intimately connected with speech, people gesture while talking on the telephone even though no one can see them.

Though posture and gesture are distinct communicative phenomena they use the same body parts, and as such there is a single animation module for both. Postures and gestures and generated from a set of basis poses (which are static) and animations (which are body movements). New postures or gestures are generated by a weighted interpolation over these bases. In order to vary the postures or gestures in response to the participant's behaviour while maintaining a large variety of behaviour, we group the bases into different types. Different types of behaviour are generated depending on the participant's behaviour, but each type can exhibit a variety of different behaviour by choosing different interpolation weights for the members of that type.

2.2 Facial Expression

The facial animation module is based on morph targets. The face is represented as a mesh, and each facial expression is represented as a set of displacements from this mesh (a morph target). The face is animated by giving weights to the morph targets. The displacements of each morph target are scaled by its weight and added to the face mesh, generating a new facial expression. The facial animation module works in the same way as the body animation module, having a number of bases which are interpolated to produce new animations. The bases can either be static facial expressions (morph targets, for example a smile) or facial animations (time varying weights over the morph targets, for example open and closing the mouth for speech). As with body motions the facial bases are grouped by type. Facial expression is not currently used to react to the behaviour of the participant, we always use a friendly smiling expression (see figure 2). Facial expression is also used to represent speech and blinking.

2.3 Gaze

The gaze animation module determines where the character is looking. At any given time the character is looking at a single gaze target, which might be the participant, an object in the environment or a location. The character moves its eyes, head and body to look at the target. It looks at the target for a set duration and after the end of that duration a new target is determined based on rules as described in section 4.

2.4 Proxemics

Proxemics are spatial relationships between people. People tend to maintain a comfortable distance between themselves. This distance depends on a number of factors such as culture and the relationship between the people. The proxemics animation module maintains this comfortable distance. If the distance between the character and participant is too large the character steps towards the participant and vice versa. The distance itself can be varied to make it a comfortable distance for the participant, or an uncomfortably distance (too close, for example) in order to elicit a behavioural response from the participant.

3. Interaction

For truly correlational behaviour the character must be able to detect the behaviour of a real person in order to react to it. The work is targeted at standard Cave-like systems and other similar immersive systems. As such, participant sensing is limited to the types of sensor that are normally available on this type of system. In fact, we only use two sensors, a 3-degrees-of-freedom head tracker (InterSense IS900) and audio input from a standard radio microphone. We attempt to extract enough information from these limited sensors to give a strong sense of correlation. The use of these limited sensors has the obvious advantage that they are relatively cheap but also that they are less intrusive and bulky than full body tracking. It is important to avoid overly intrusive trackers as they can be uncomfortable for the user and reduce the naturalness of their behaviour. This is particularly true of the subtle behaviours that make up non-verbal communication. The rest of this section describes how the sensor information is mapped to the character's behaviour, figure 1 gives an overview of this process.

3.1 Head Position

The most basic information that can be obtained from the head tracker is the current position of the participant. This is used by the proxemics module to determine the current distance of the character to the participant. In order maintain a comfortable distance as described in section 2.4. The head position is also used by the gaze module to enable the character to look appropriately at the participant.

3.2 Interactional synchrony

It is also possible to obtain more complex information from the head tracker. Kendon [10] has shown that when people engage in conversation and have a certain rapport, their behaviour will tend to become synchronised, an effect he calls 'interactional synchrony'. This is particularly true of a listener synchronizing their behaviour with a speaker. This can take many forms, two of which we simulate. The first is that a listener will tend to move or shift posture at the same type as the speaker (but not necessarily have the same posture). This can be implemented very simply using a single head tracker. We detect the participant's posture shift when the tracker moves above a threshold. When a shift is detected the character will also perform a shift. The other form of interactional synchrony noted by Kendon that we simulate is a listener synchronizing their movements with important moment in the speaker's speech. As we detect when the participant is speaking (see section 3.4) it is possible can detect the start and end of their speech. The character performs a posture shift at these two important moments in the conversation.

3.3 Head orientation

The head tracker also gives the orientation of the head. This can give an approximate direction of gaze for the participant. This is used to implement gaze following. A powerful cue for social understanding is that a one person will look in the same direction as another[9] . This displays shared attention, that they both share an interest in a common object, and they both understand that the object is important to the other and to the conversation. Thus a character that follows the gaze of the participant gives a powerful cue that they are understanding the participant's conversation and that they empathise, to some degree, with the participant. This only works when the participant is looking at something relevant, so the character cannot follow the participant's gaze arbitrarily. Otherwise the character will appear to be constantly looking at irrelevant objects, and seem stupid. To avoid this problem certain objects in the environment and defined to be *salient objects*, when the participant appears to be looking at one of these the character will follow gaze, but not otherwise.

3.4 Speech

As this work deals mostly with social behaviour a good model of speech and conversation, is needed. This model depends on a conversational state, which can have one of three states: *character talking*, *participant talking* and *neither*. The character's own conversation is handled in a wizard-of-oz manner, a number of audio clips, can be triggered by a confederate. It is thus trivial to know if the character is talking. The participant has a radio microphone

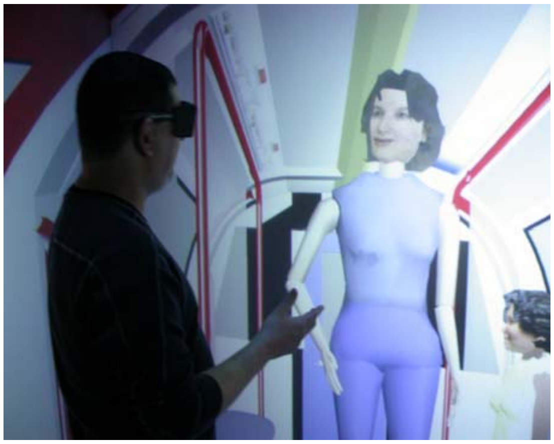

Figure 3: Social Interaction between a real and virtual human

which is used to detect when they are talking (simply based on a threshold for the amplitude of the signal). The behaviour associated with speech is consists in gesture, gaze and posture shifts (describe in section 3.2).

Gesture behaviour is intimately connected with speech. There are two basic types of gesture, normal gestures that accompany speech, and "back channel" gestures that occur when listening, and aim to encourage the talker. Normal gestures are modeled based on a number of basis gestures as described in section 3.1, and only occur in the *character talking* state. The characters mouth is also animated during the *character talking* state to show that they are talking. The most common back channel gestures in western culture are head nodding to show agreement and encouragement, and shaking the head to show disagreement. As the character's behaviour is designed to be favorable towards the participant, only head nodding is shown.

The character's gaze is driven, based on speech, by a model by Garau *et al.*[5] Vinayagamoorthy *et al.*[15] and Lee, Badler and Badler[11] , which are ultimately based on the work of Argyle and Cook[2] . In this model the character looks either at their conversational partner (the participant) or at other locations in the world. The length of any look is determined at random based on mean lengths determined from data from observation of conversations. The mean length of looking at the participant is greater when listening than when talking (as is consistent with numerous Argyle's observations of conversations). When the character is not looking at the participant then the locations chosen are determined based on statistics by Lee, Badler, and Badler.

5. Conclusions

This paper has described work in progress in developing correlational non-verbal behaviour in virtual characters. The aim of this work is to enhance presence in social interactions with virtual characters by simulating a key element of real human social interactions. We are currently planning a study to test the effects of this work. The study will involve the subjects holding a conversation with a character controlled by the behaviour model described, compared with a character that exhibits the same beahviour but without it being correlated to the behaviour of the user. . The scenario chosen is one of a London Underground train, with the character being a tourist asking directions (the environment and character are shown in figures 2 and 3).

Acknowledgements

This work is supported by the Presencia EU FET project.

References

[1] Argyle, M. Bodily Communication. Routledge. 1975

[2] Argyle, M., and Cook, M. 1976. Gaze and Mutual Gaze. Cambridge University Press,

[3] Cassell J, Bickmore T, Campbell L, Chang K, Vilhjàlmsson H, Yan H Embodiment in Conversational Interfaces: Rea. *Proceedings ACM SIGCHI*, 520-527: ACM Press. 1999

[4] Cruz-Neira, C., Sandin, D.J., DeFanti, T.A. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. *ACM Computer Graphics*, 27/2 , 135-142. 1993.

[5] Garau, M., Slater,M., Bee, S. and Sasse, M.A. (2001). The impact of eye gaze on communication using humanoid avatars. *Proceedings SIG-CHI conference on Human factors in computing systems*, 309-316. 2001

[6] Gillies M, Ballin D. Integrating autonomous behaviour and user control for believable agents. *Proceedings Third international joint conference on autonomous agents and mult-agent systems.* New York USA. 2004

[7] Gillies M, Crabtree B, Ballin D Expressive characters and a text chat interface. *Proceeings AISB workshop on Language, Speech and Gesture for Expressive Characters* (Olivier P, Aylett R, eds). University of Leeds, UK 2004

[8] Guye-Vuillème A, T.K.Capin, I.S.Pandzic, Magnenat-Thalmann N, D.Thalmann Non-verbal Communication Interface for Collaborative Virtual Environments. *The Virtual Reality Journal* 4, 49-59. 199

[9] Johnson, S.C. Detecting agents. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 358, 549-559. 2003

[10] Kendon, A. Movement coordination in social interaction: some examples described. *Acta Psychologic*. 32, 100-125. 1970

[11] Lee S. P., Badler, J. B. and Badler, N. I. Eyes Alive, *ACM Transactions on Graphics, (Proceedings of ACM SIGGRAPH 2002)* 21/3, 637-644, July 2002

[12] Pelachaud C, Bilvi M. Modelling gaze behavior for conversational agents. *Proceedings Intelligent Virtual Agents*, pp 93-100. 2003

[13] Pertaub DP, Slater M, Barker C. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence-Teleoperators and Virtual Environments* 11, 68-78. 2002

[14] Sanchez-Vives MV, Slater M. From Presence to Consciousness Through Virtual Reality. *Nature Reviews Neuroscience* 6, 8-16. 2005

[15] Vinayagamoorthy V, Garau M, Steed A, Slater M An eye gaze model for dyadic interaction in an immersive virtual environment: Practice and experience. *Computer Graphics Forum* 23, 1-11. 2004