

Refining Personal and Social Presence in Virtual Meetings

Jesse Dean, Mark Apperley, Bill Rogers

Department of Computer Science, University of Waikato
Hamilton, New Zealand

m.apperley@waikato.ac.nz

Abstract

Virtual worlds show promise for conducting meetings and conferences without the need for physical travel. Current experience suggests the major limitation to the more widespread adoption and acceptance of virtual conferences is the failure of existing environments to provide a sense of immersion and engagement, or of ‘being there’. These limitations are largely related to the appearance and control of avatars, and to the absence of means to convey non-verbal cues of facial expression and body language. This paper reports on a study involving the use of a mass-market motion sensor (Kinect™) and the mapping of participant action in the real world to avatar behaviour in the virtual world. This is coupled with full-motion video representation of participant’s faces on their avatars to resolve both identity and facial expression issues. The outcomes of a small-group trial meeting based on this technology show a very positive reaction from participants, and the potential for further exploration of these concepts.

Keywords: virtual meeting, motion sensor, Kinect, avatar

1 Introduction

Video-conferencing technology, once a stuttering impediment to group discussion (Egido, 1988) has moved forward significantly over the past 25 years. High-speed internet connections and audio-visual integration with even low-cost personal and mobile computers, mean that reasonable quality video calls are available to all, and that the centralised technology of the video-conferencing suite is fast becoming an anachronism.

In hand with this ability to readily communicate via everyday technology, at one’s desk, or even on the move, is the growing motivation to use this technology as a substitute or alternative to physical meetings. Not only is the budgetary cost of having people travel to meetings a factor in this shift (Lindeman, Reiners & Steed, 2009; Erikson, Shami, Kellogg & Levine, 2011) but there is increasing awareness of the cost to the environment of such travel (Arnfolk & Kogg, 2003). Personal travel accounts for almost one-third of all energy consumption in the developed world, and that energy is based almost exclusively on fossil fuels, a fast-dwindling resource, and accountable for the majority of greenhouse gas emissions (Mackay, 2008).

While Voice Over Internet Protocol (VOIP) audio/video telephony services such as Skype™ are widely and successfully used, they are essentially one-to-one, and although they can extend to multi-party conference calls, there is an absence of a sense of *being* anywhere (other than where the caller is) or of *being together*; the normal telephone sense of speaking-across-a-distance remains (see, for example, Steuer, 1992). On the other hand, virtual worlds, such as Second Life™ do provide a sense of place, and of being somewhere, and

have successfully been used to provide virtual conference venues (Lindeman *et al*, 2009; Al Qahtani, 2010; Erikson *et al*, 2011). Meeting participants are represented by avatars, which they control, in a shared space, which may resemble a virtual meeting room, with furniture and projection screens. However, a significant shortcoming of using such environments for meetings is the major overhead for the participant of managing and controlling their avatar, and their view of the shared meeting space. These overheads become the dominant activity and significantly detract from the sense of presence, and engagement with the meeting (Al Qahtani, 2010).

This paper describes the development and evaluation of software (VMX) to reduce this overhead, and to refine the sense of personal and social presence in the virtual world (Dean, 2012). This work acknowledges the fact that the typical participant will be physically located in their own personal space (quite likely an office), and will be seated in front of a computer screen and keyboard. It uses a 3D motion-sensing input device to capture participant actions in this real space, which it then maps to view-controlling operations, such as panning and zooming, and avatar actions in the virtual space. Some avatar actions are directly mapped from the user’s movement (eg, turning one’s head), and others indirectly mapped (eg, pointing at a screen). Specific facilities are provided for presentations in the virtual space. The real-time image of the participant’s face is superimposed on the avatar, so providing for the subtleties of facial expression.

The preliminary evaluation of VMX has shown the combination of these facilities are remarkably successful in promoting a sense of engagement with the meeting, and of “being there”.

Section 2 of this paper provides a brief overview of relevant prior research and experience in relation to virtual meetings, personal and social presence in virtual environments, and the capabilities and use of mass-market motion sensing technologies in virtual world interaction. Section 3 then develops the requirements and explains the design for software to interpret participant

actions and map these to avatar and camera control, recognizing the capabilities of the current technology. Details of the implementation of the prototype VMX software system based on the Kinect™ (Kinect) motion sensor, which incorporates and explores these ideas, and limitations and issues with this implementation, are summarized in Section 4. A trial evaluation of the environment is described in Section 5, including a discussion of the feedback from participants, and some further ideas which arose from the experience. In the conclusion (Section 6) the overall findings are discussed, and directions for future research with this technology are proposed.

2 Background

Research relating to virtual meetings and collaboration at a distance, be it through telephone, video or computer/communications technology, has a history of more than forty years (eg, Chapanis, Ochsman, Parrish & Weeks, 1972) and crosses a range of disciplines including technology, psychology, sociology, education and management/organisation. This paper is specifically concerned with the use of virtual worlds as places to conduct meetings or gatherings, which may range from small group discussions to large group presentations or performances. Within this context, it sets out to develop software based on mass-market motion sensing technologies to refine the concepts of personal and social presence in a virtual world, and by reducing the additional overheads of virtual participation, enhance the level of engagement.

2.1 Meetings in virtual worlds

In 2009, an IEEE Virtual Reality conference program committee meeting was conducted in Second Life (Lindeman *et al*, 2009). Some 39 members of the committee participated during the meeting, and on average attended about 60% of the approximately 9 hours total. Attendees were widely distributed across the globe. Although the overall reaction was favorable (in preference to travelling to a two-day meeting), participants felt it was not the same as a face-to-face meeting, with personal engagement less satisfactory in the virtual world. They found it difficult to identify other participants amongst the ‘plastic’ avatars, found the absence of body language an issue, and the document presentation tools less than ideal.

A somewhat larger and more ambitious meeting held in Second Life is described by Erikson *et al* (2011). A large corporate meeting spread over 3 days was attended by approximately 500 participants from around the world. It was a single track conference, with a schedule designed to accommodate attendees from all time zones. Overall, participants felt the technology worked well, but criticisms included: keynotes were wasted in the virtual environment, and would have been better simply streamed; avatars in general tended to look the same, and were difficult to distinguish or identify; feelings relating to social events were mixed, with some suggestion they need to be either highly structured or focused on quite small groups. In spite of these shortcomings, the authors

present a positive view of the future potential for large group meetings in virtual environments.

A further small-scale study has provided much of the motivation for the work described in this paper (Al Qahtani, 2010). A trial meeting involving seven participants was held in a hired virtual meeting room in Second Life. The meeting involved a slide presentation followed by a discussion. Reactions were generally positive: participants found the virtual conference more immersive than a conventional video conference. However, largely in common with the studies described above, the following issues were identified:

- (i) the need for a participant to overtly control their avatar is a distraction, and can dominate other activity;
- (ii) providing gestures is laborious, and the gestures, which are stylized carbon-copies are unsatisfactory;
- (iii) individual avatars can be difficult to identify, even in a small group;
- (iv) there is an absence of any sense of audience response or level of engagement, such as subtle non-verbal cues like facial expression or eye-gaze direction;
- (v) ‘in-world’ presentations are difficult to deliver for the avatar-as-presenter, and not easy to view by the participant;
- (vi) general navigation within the world (getting to the conference room, sitting down, controlling one’s view, etc) seems overly complex, particularly for those participants with no prior experience of Second Life;

These six points form the key issues which our VMX system attempts to address.

2.2 Presence

Although motivated by the potential use of virtual reality in psychological therapy, Schuemie *et al* (2001) provide a relevant and comprehensive review of research and issues related to the concept of presence in virtual worlds. Although their review is not conclusive, and they suggest that a sense of *presence* alone may not be the key to the successful virtual environment, the research they review does identify factors such as immersion, engagement, social interaction, naturalness, social reality and interpersonal communication cues as being important precursors to that sense of presence. A simple dichotomy between “real” and “virtual” worlds has also been questioned (Taylor, 2002), with a stress on the significance of the persona and appearance of the avatar in the virtual world contributing to the sense of presence and reality.

Discussions of presence often focus on the exchange of non-verbal communication with an emphasis on the use of facial expression to establish trust, particularly in negotiation situations (Purdy & Nye, 2000; Bekkering & Shim, 2006). In Al Qahtani’s (2010) study, where the participants were all known to one another, it was identified that non-verbal communication would have been helpful for two purposes. First, to establish whether another participant was paying attention; in a virtual meeting there is always the possibility that a participant is AFK (away from their keyboard), or doing something not connected with the meeting, whereas a ‘live’ avatar imposes some social pressure to take part. Second,

although sound location can be helpful in determining who is talking (DiVincenzi, 2011), it gives no idea whether their speech is being directed at someone specifically; an avatar with directional gaze could provide this cue.

2.3 Human motion sensing

Computer-based systems for sensing and analyzing human motion have evolved over several decades (Dean, 2012). Most recently applicable technology has evolved from the once complex, cumbersome, intrusive, inconvenient and expensive, to simple, low-cost, non-intrusive systems based on depth sensing cameras, of which Microsoft's Kinect is probably the most pervasive and well-supported.

Kinect and its supporting software was introduced by Microsoft as part of the Xbox 360 gaming system in late 2010 (Kinect). It is designed to track a human player, to allow them to interact with games by verbal command, body movement and gesture. The unit incorporates several components: microphones which pick up voice commands; an infrared projector which displays a pattern of infrared light; and two video cameras, one tuned to visible light and one to infrared. The visible light camera produces an ordinary image. The infrared camera provides data which can be converted to a depth view of the scene – for each pixel, measuring the distance between the camera and a point in the scene. Further analysis of depth information allows estimation of body position and configuration in 3D.

The utility of Kinect in improving video conferencing facilities has been demonstrated with the Kinected Conference project (DeVincenzi *et al*, 2011). This project has implemented several features that take advantage of the ability of the Kinect audio array to determine the position of a speaker in the field of view, and the ability of the Kinect depth camera to identify the spatial location of objects in the view of the video camera. By using this information the software is able to perform a number of visual enhancements to the video feed being sent to the remotely connected participants of the meeting, such as focusing the camera on speakers, freezing parts of the camera image, and overlaying spatially contextual graphics.

2.4 Background summary

Although conferences held in virtual worlds show promise, they are unlikely to become an unqualified alternative to face-to-face meetings. The limitations are very much associated with appearance and control of avatars (in subtle, fine and gross movement), as well as camera control. For a real sense of engagement, participants need to be relieved of the burden of overt avatar control (being puppeteers), yet be confident that the avatar is a reasonable representation of themselves and projection of their body language. Also, in order to be able to focus on conference participation, gross avatar movement and camera control (controlling your view) need to be as unobtrusive and natural as possible. Current consumer level motion sensing technology has the potential to assist in meeting these requirements.

3. VMX Design and Development

As noted earlier, there are many possible scenarios for virtual meetings. One possibility is to simulate a large conference, or a performance (Yong, 2003). In this situation a presenter might want information about audience response to their material. For an audience member, view of the participant is probably best provided by conventional streamed video (Erikson *et al*, 2011). The sense of being part of an audience, hearing others gasp or laugh and possibly exchanging occasional comments with an immediate neighbour, could be benefits of a more immersive virtual environment. For such scenarios, technical issues concerning the multiplexing and combination of large numbers of location/movement/video/audio feeds are challenging, while issues of maximising the benefit of detailed exchange are not so interesting. On the other hand, a small meeting (perhaps 4 to 10 people), provides a situation in which technical communication issues are manageable with current technology, and the intimacy and familiarity of such settings enables us to focus on the challenges of supporting issues of personal and social presence in a convincing manner. Accordingly, our work has focused on virtualising a small meeting. However, even in small meetings, seeing and talking to other participants is usually not enough; for example, frequently, one member will lead discussion by presenting a report or plan. We have therefore developed a system (VMX) to support a small meeting, with provision for members to share documents and make presentations. The work focuses on addressing the issues (as listed in Section 2.1) identified by Al Qahtani (2010) in experiments with meeting in Second Life.

3.1 Avatar control

The first issue to be addressed is that of avatar control. In Second Life, participants must be 'puppeteers', using keyboard and mouse controls to operate their avatars. The Second Life system gives good control of gross movement, allowing users to explore their world, albeit requiring some familiarity and experience with those controls. However, it provides poor fine control. While it is possible to orient the whole body, determining the gaze direction, it is not possible to refine gaze direction by eye movement. While it is possible to perform a number of animations; for example, to wave or jump up and down, by selecting from a menu, it is not possible to provide detail in that interaction, such as pointing to an item of interest. Issuing commands itself can be slow, although speed can improve if the participant has practiced with the software. For example, if a speaker asks for a show of hands in a vote, participants must search through a list of animations to find the appropriate one, and that may take some time. At best their speed of response is an indication of their speed of menu search, rather than a sign of their enthusiasm. The lack of fine detail means that there is no expressiveness in the gesture itself, as might distinguish a reluctant partial raise of the hand from something more enthusiastic. If the player does not issue commands, nothing happens to their avatars; they remain stationary. In particular it is impossible to distinguish the avatars of people quietly

listening from those of people who have stopped participating altogether. We hypothesised that tracking a participant's body and using body motion to animate their avatar could have three advantages. First, it would not be necessary for the participant to issue commands, freeing them to focus on the content of the meeting. Second it would make fine grain animation possible, both in timing and motion, limited only by the speed and detail in which body movement was captured. Finally, it would help to show what a person was doing, even when they were not intentionally communicating (eg looking away from their screen).

We note that one of the advantages some perceive with a virtual meeting setting is the anonymity conferred. Appearance can be arbitrarily mapped onto an avatar, meaning that people are not required to dress appropriately, wear make-up, have neat hair, etc. They might value the option of doing other things whilst 'attending' the meeting (perhaps reading email, or playing games on their computer), safe in the knowledge that others don't know what they are really doing. However, a feature of real meetings is that there is social pressure to participate properly. Others will notice if someone isn't paying attention. Whilst there may be circumstances in which some degree of anonymity might be valuable, our work focuses on exploring a strong sense of social presence, and this requires that participants are willing to allow others to be closely aware of their actions.

3.2 Mapping from the real to the virtual

A virtual meeting involves participants in a real world setting (*not* the meeting), with information being captured and mapped to a virtual setting (*the* virtual meeting). The simplest model is that people sit at a desk, in front of a computer, quite likely their normal work scenario. This maps nicely to a virtual environment in which they are depicted as sitting at a conference table. Potentially, such a system provides good possibilities for measuring fine movement. We simply need a sensor on the computer to capture an image of the person (see Figure 1). The Kinect sensor is a good fit for this environment.

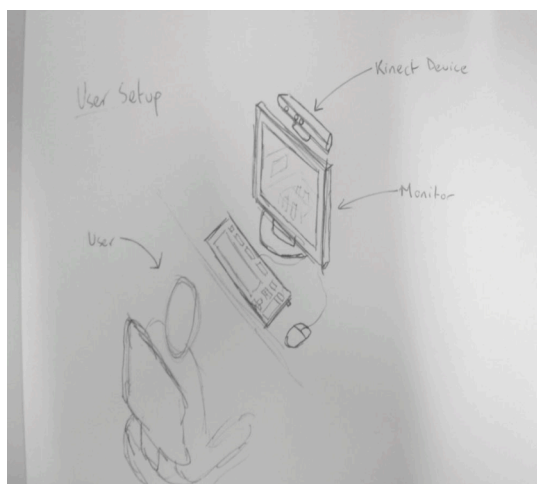


Figure 1: The participant in their real-world environment

However, with this scenario, there is little possibility for mapping gross movement; the participant cannot move in such a way as to naturally map to their avatar moving around the meeting room. In fact that need not be problem; people usually stay in their chairs during meetings. If they get up to get a drink, for example, they are effectively leaving the meeting temporarily. Perhaps only common exception to the fixed seating rule is that someone may move to the front in order to make a presentation.

A minor extension to our real world setting accommodates presentations as shown in Figure 2. As before, a participant normally sits in a chair in front of a computer. A sensor mounted on the computer monitors them, and they continue to view the virtual world on the computer display. In addition however, a large display screen (or a whiteboard) is arranged approximately 2 metres back from the computer, fully within the viewing angle of the sensor. The space is sufficient to allow the user to stand and present at the screen, and the sensor can capture a full body view allowing the user to step close to the screen and gesture. This modified arrangement is still suited to the capabilities of the Kinect device and still requires only a single sensor. The user can shift between presentation and sitting modes simply by moving. There are a number of options for mapping the movement to a presentation position. One possibility is to have a presentation screen behind each participant in the virtual world; another is to animate walking to the front of the room. In our experimental system, we have chosen to simply jump the avatar to a presentation position when the presenter moves to their real-world screen.

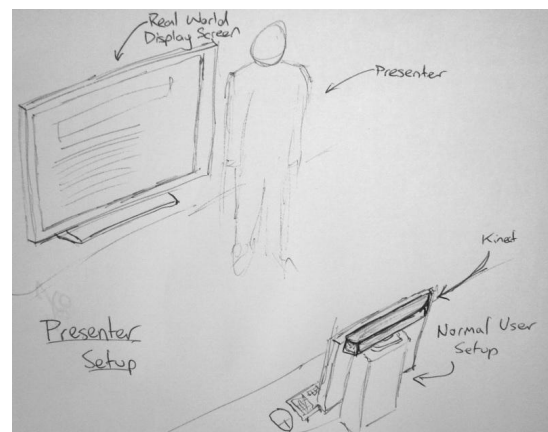


Figure 2: The modified real-world environment allowing for presentations by the participant

3.3 Avatar appearance

The second issue to be addressed is that of avatar appearance and facial expression. VMX avatars have been implemented as 'pipe' models based on the 'bones' deduced by the Kinect body position recognition system, with the head shown as a torus containing a live video image of the participant's face; Figure 3 shows an early implementation view. In a real meeting there is little interest in looking at the back of someone's head, so the video of the face was made one-sided; from behind the head is an empty torus, but the orientation makes it possible to work out the general direction in which the

person is facing. When a person is facing away from the observer, their head does not completely block the view.

Using a live video image satisfies both of the requirements of participant identification and conveying facial expression.

An interesting question was the size of the avatar head (and video image). Figure 4 shows three different head sizes. While the smaller head gives a better view around the room, larger heads give better opportunity to observe facial expressions. VMX provides a keyboard control to allow the user to adjust head size. The default is a view that is roughly anatomically correct.



Figure 3: One participant mapped around the conference table. This is an early screen shot, before chairs were implemented. Note the arm gesture, the face, and the back-of-head transparency.



Figure 4: Effect of different avatar head sizes

A final detail in avatar display is that the avatar ‘pipe’ body is coloured by automatic selection of a colour from the user’s clothing – from near the centre of their torso.

An avatar’s position at the meeting table is determined at present by login order. Once the location is determined, the system maps orientation and position automatically. Small movements by the user in the real world are mapped to appropriate movements at the virtual table. However, in doing this mapping, a problem arises.

It makes sense that a participant might look around the table by turning their head. To look at the neighbour to the left, the virtual head would be turned 90 degrees to the left. Unfortunately, the participant’s view of the virtual world is provided by an ordinary size computer monitor in front of the user. Turning by 90 degrees means that they are no longer facing the monitor, and therefore cannot see anything, although their avatar turns correctly for the intended view. The VMX system solves this problem by making avatar head orientation an

amplified copy of body orientation. Users are sitting in ordinary office chairs, which permit swivel. A small turn of the chair (and hence the body) is mapped to a larger head movement on the avatar, which is in turn reflected by change in view orientation in the image of the virtual scene. This system allows a user to turn their avatar, and thus view, plus or minus 90 degrees while continuing to comfortably view their screen. User testing showed this to be a remarkably natural mechanism to which people adapted very quickly.

3.4 Presentations at the virtual meeting

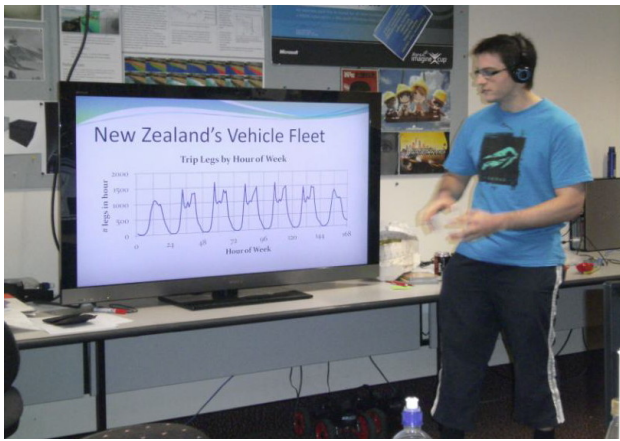
Presentation mode is based on a user standing in front of a large, but not huge, display (Figure 2). In our experimental setup, the presentation display in a participant’s real-world environment is a 52 inch video display; large in a personal setting, similar to a medium sized office whiteboard, and therefore appropriate. It is a comfortable size for hand gestures, although gesturing may at times mean that the presenter is obscuring part of the display. It is however, not as big as a typical conference room display, and our experiments showed that the display in the virtual meeting room should be larger than life size.

There are two options for managing content on the display screen. One is to use video of the real display. That would nicely accommodate a whiteboard, but would lead to lighting, resolution and occlusion problems. The alternative is to use a digital image. This latter option was chosen. A PowerPoint slide show could be displayed on the real large screen, and independently mapped onto the virtual display screen. The result is a crisp image on both, with the virtual display larger than the real one, relative to the room and the avatars (see Figure 5).

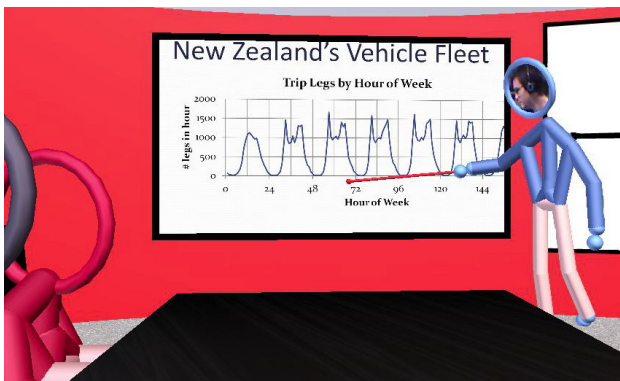
At this point, a new complexity arises. Being able to point to features of a displayed document was an important goal of the project. The scale remapping makes this difficult. The VMX implementation addresses the problem as follows. Video of the presenter’s hand is tracked in the real world by the Kinect camera. If the user makes a pointing gesture, with their index finger, the system determines the point targeted on the real screen, and in the virtual world draws a virtual pointer from the avatar’s hand to the corresponding point on the virtual display screen. Figure 5 illustrates this situation. (Note also the blue avatar body, matching the presenter’s T-shirt colour.) Evaluation showed that this approach worked with acceptable accuracy.

A benefit of the screen mapping and virtual pointer system is that the software has some control over avatar placement, and if necessary could automatically ensure that the avatar didn’t occlude any of the display, although this feature is not implemented in the current VMX system. Note also that the user doesn’t need to hold anything. The pointer is generated whenever the user makes a pointing gesture and that gesture is in the direction of the screen.

The fact that the presenter’s hands are free leaves open the possibility of further gesture usage. VMX implements gestures to scale, pan and page the display document, although testing has shown that these are not easy to use.



(a)



(b)

Figure 5: Mapping of a presenter and a presentation; (a) the real-world, and (b) the mapping into the virtual meeting room

3.5 Viewing control

VMX provides additional movement controls. It is possible to watch a presentation in seated mode, and this has the advantage of allowing a user to keep an eye on other participants. Figure 6 shows a presentation viewed from one participant's seat; three of the other participants are watching the presentation, while the fourth is distracted, and looking down at something in their real world space. The prototype implementation allows

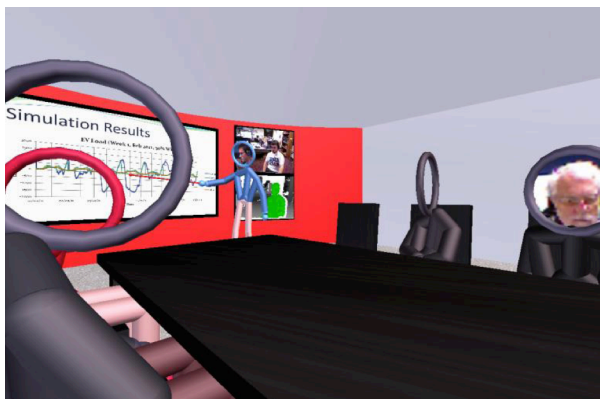
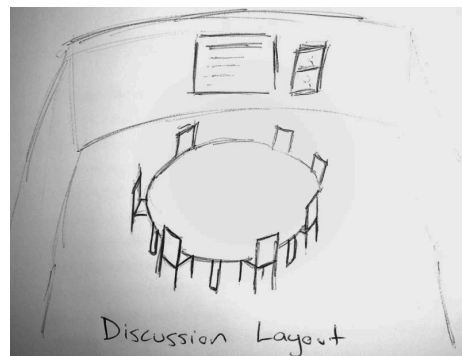


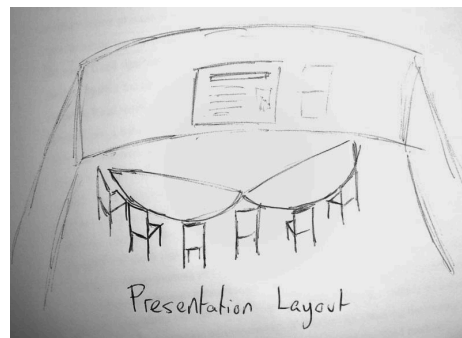
Figure 6: A presentation in progress, as viewed in first-person by a seated participant

participants to set their viewpoint to a third-person perspective (Salamin, Thalman & Vexo, 2006) from elsewhere than their seated position. The head-of-table view of Figure 5(b) demonstrates a better-chosen viewpoint for the presentation.

An experimental feature of VMX is table reshaping to accommodate presentations, and to obviate the need for a third-person perspective. The concept is shown in Figure 7. There is no reason to keep table layout fixed; the format suited to presentation may be different from that suited to discussion. As the mapping from a participant's real world space to the virtual world is an artifice, it can be changed dynamically, with the only constraint that the change should be animated to avoid participant confusion.



(a)



(b)

Figure 7: The table can be morphed from a discussion layout (a) to one more suited to a presentation (b)

4. VMX implementation details and issues

The following brief comments relate to pertinent features and issues with the current VMX implementation.

- (i) The prototype VMX system is implemented in C# using the XNA 3.1 graphics library and version 1.0 Kinect SDK. Networking is a client / server model, with clients streaming position and video to a server for distribution. Video is sent via UDP; position uses a TCP channel. The system is quite demanding of network bandwidth; careful setting of video resolution and size was necessary to get acceptable performance for 6 users in a local area network environment (only 35x35 pixels for facial features).
- (ii) The Kinect system cannot provide bone locations for legs and feet when a user is seated, because the participant's legs are not in the sensor view. The

VMX system calculates and imposes a seated posture for these non-tracked bones.

- (iii) Kinect does not resolve head direction; the head is just a single bone. VMX assumes that the head is oriented at right angles to the line between the shoulders. Shoulder orientation is thus taken as a proxy for head orientation.
- (iv) Kinect does not track finger positions. To compensate for this lack, VMX implements an image analysis system to identify hand gestures, in particular the finger point gesture. The image analysis takes advantage of Kinect camera depth information to separate the participant from the background.
- (v) Kinect also uses image analysis to identify the location of the large presentation screen, for the purpose of resolving the target of presenter pointing.
- (vi) The VMX prototype does not incorporate audio facilities. For testing purposes we used an independent program – Teamspeak (www.TeamSpeak.com) – to provide the audio channel. The only disadvantage is that a user must establish audio connection independently of their meeting login. However, from an experimental development viewpoint, it was convenient to have a reliable system to use, and therefore not to have the evaluation results complicated by any issues that might have arisen with audio.

5. Testing and evaluation

In order to evaluate the VMS software a trial meeting was conducted with five active participants plus an observer, all present in the virtual meeting room, but physically located in separate spaces. The agenda for this meeting included a research presentation followed by a round-table discussion. Figures 5 and 6 show scenes recorded from this trial. All participants received brief instructions on the use of the VMX software at the start of the meeting. Following the meeting participants completed a questionnaire divided into five sections, covering the features of the VMX software, the experience of delivering a presentation, the experience of the meeting in general, how VMX compared with other forms of meeting, and other and general issues.

Points of note raised during the meeting and in the questionnaires can be summarised as follows:

- Participants reported having difficulty in seeing past the heads of others, in spite of the transparent back-of-head feature. Some overcame this problem by moving the camera to a third-person perspective, but none used the head-shrinking facility for this purpose.
- Participants reported some difficulty in discerning facial expressions. In part this was due to the lower than ideal resolution (35x35 pixels for faces) dictated by network bandwidth issues, and in the case of the presenter, exacerbated by their distance from the display screen.
- Most participants reported being able to recognise when someone was applauding, pointing, raising their hand to ask a question, and other gestures involving hand/arm movement, although at least one

instance of a raised hand did go unnoticed by the presenter (this can happen in real life too).

- Some reported that body language was at times obscured by jitter in skeleton position data.
- In general, it was possible to detect in which direction participants were looking, and who was speaking at any given time. However, there were suggestions that a flag above the current speakers head would be a useful addition.
- There were some reports of activity outside of the meeting being observed, through head movement, body movement, and eye-gaze direction.
- There were some problems when the presenter unintentionally moved out of the camera frame, and the presentation view controls (scroll, zoom) required practice to be effectively used.
- At the suggestion of one participant, the presentation screen was used to record notes during the discussion phase of the meeting. Small modifications to the system could make this feature even more effective.
- Opinions on the relative merit of automatic (first-person) and manual (third-person) camera control were mixed, but there was general agreement that there was a place for both.
- When compared with other forms of remote meeting (video conference, teleconference, non-Kinect virtual world meeting) all participants reported that the VMX experience was as good or better. They found it more relaxed than a videoconference, and more immersive and engaging than other forms of remote meeting. None, however, suggested that the VMX meeting was as good as a face-to-face meeting.

In summary, we would suggest that this outcome from the initial evaluation of a prototype system is very positive, and shows significant promise for moving forward with the notion of meetings in virtual worlds.

6. Conclusion

6.1 Summary

This project set out to address a number of issues identified in existing virtual meeting software as described in the background discussion (Section 2). We believe that VMX successfully solves or at least considerably improves the situation in each case.

First, the issues surrounding fine grained avatar control. In VMX, user body movements are directly measured and replayed on the avatar. This successfully transmits explicit gestures, such as hand raising, in a way that requires no other action by a user except simply making the gesture. Because all movements are transmitted, there is a continuous sense of 'liveness'. The avatar orients wherever its user is facing. If the user thumps the table, mops their brow, or just wriggles in their seat, their avatar follows suit. Within the limitations of the avatar representation, reproduction of movement is faithful in timing and intensity. An immediate, enthusiastic raising of the hand is clearly different from a slow, incomplete gesture.

VMX displays live video of participants' faces on their avatars' faces (head tori). This personalises the

avatar and provides the primary means of identification. The face is not always in view, so identification may have to wait until a person turns, but that is not too different from what might happen in a real world meeting. A secondary means of identification is provided by colouring the avatar body. This makes the avatars look different from one another, and serves as a memory aid to retain identification when the avatar is facing away. A second effect of face video is the transmission of non-verbal cues, such as facial expression and eye gaze. Eye gaze reproduction is not very meaningful in the left/right dimension. VMX's head orientation is a more useful indication of gaze target. However, the vertical dimension of eye gaze shows clearly when a person's attention is focused on the table surface of their real world, rather than the meeting. Overall, the combination of avatar body position mapping, and face video provides an excellent rendering of a participant's level of engagement with the meeting.

Presentation mode enables a participant to address the meeting, supported by a 'projected' document. VMX's extension to Kinect's tracking capability provides a virtual pointer so that audience attention can be directed to items on that document. Face video has the curious effect of allowing any speaker to achieve what only very good speakers can achieve in the real world. By looking at the Kinect sensor, a presenter can appear to be looking directly at each and every audience member.

Finally, navigation. VMX does not provide large scale navigation because there seems to be no reason to simulate movement that is at best marginally relevant to the task of attending a meeting. In VMX, a person is at the meeting table, or giving a presentation, and they 'teleport' between the two states. Indeed, this 'limitation' provides a practical advantage. There is not the opportunity offered in the real world to stumble, drop papers, or knock over chairs. Other simulated circumstances might dictate a different approach. For example, at a virtual cocktail party, participants should be allowed to circulate around discussion groups.

VMX implements an innovative feature in small scale navigation. Using body rotation to orient a person's avatar might lead to them no longer facing their real world screen. To avoid this, the system amplifies their angle of rotation. Whilst this may seem unnatural, participants adapted to the effect very easily, and after a short time found that they were looking around the virtual scene without conscious effort – in a swivel office chair, this could be described as steering by the seat of the pants.

6.2 Further work

The VMX system described is a prototype, and is under on-going development. A range of refinements based on experience so far are being considered, including further gesture use for presentation control and the meeting table manipulation already described.

The possibility of extending VMX to a large conference or performance setting is a greater challenge, given the communications limitations experienced with the small-scale prototype. However, a possible solution would be to implement local groups in the audience, allowing people to attend in small groups (similar in size

to those that VMX currently supports), and within these groups having participants communicate directly with one another. Providing full audience feedback to the presenter could be achieved by an aggregation tree of video feeds, merging into a single video image stream of the audience. A similar aggregation of audio could allow audience members a general impression of overall sound – allowing an accurate reproduction of applause, for example.

7. References

- Al Qahtani, S.H. (2010): Virtual worlds as meeting places. Unpublished COMP591 Honours Dissertation, Department of Computer Science, University of Waikato.
- Arnfolk, P. and Kogg, B. (2003): Service transformation—managing a shift from business travel to virtual meetings. *Journal of Cleaner Production*, **11**(8), 859-872.
- Bekkering, E. & Shim, J.P. (2006): Trust in videoconferencing. *Commun. ACM*, **49**(7), 103-107.
- Chapanis, A., Ochsman, R.B., Parrish, R.N. and Weeks, G.D. (1972): Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem solving. *Human Factors*, **14**(6), 487-509.
- Dean, Jesse (2012): Using motion controllers in virtual conferencing. MSc Thesis, University of Waikato.
- DeVincenzi, A., Lining Yao, Ishii, H. & Raskar, R. (2011): Kinected conference: augmenting video imaging with calibrated depth and audio. *Proceedings of the ACM 2011 conference on Computer supported cooperative work (CSCW '11)*: 621-624, ACM, New York.
- Egido, C. (1988): Video conferencing as a technology to support group work: a review of its failures. *Proceedings of the 1988 ACM conference on Computer-supported cooperative work (CSCW '88)*: 13-24, ACM, New York.
- Erickson, T., Shami, N.S., Kellogg, W.A. and Levine, D.W. (2011): Synchronous interaction among hundreds: an evaluation of a conference in an avatar-based virtual environment. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*: 503-512, ACM, New York.
- Kinect: *Kinect Sensor*. Retrieved 19 August, 2013, <http://msdn.microsoft.com/en-us/library/hh438998.aspx>
- Lindeman, R.W., Reiners, D. and Steed, A. (2009): Practicing what we preach: IEEE VR 2009 virtual program committee meeting. *IEEE Computer Graphics and Applications*, **29**(2), 80-83.
- Mackay, D.J.C. (2008): *Sustainable energy – without the hot air*. UIT Cambridge. ISBN 978-0-9544529-3-3. Available free online from www.withouthotair.com.
- Purdey, J.M. & Nye, P. (2000): The impact of communication media on negotiation outcomes. *Intl. J. of Conflict Management*. **11**(2), 162-187.
- Salamín, P., Thalmann, D. & Vexo, F. (2006): The benefits of third-person perspective in virtual and augmented reality?. In *Proceedings of the ACM symposium on Virtual reality software and technology (VRST '06)*: 27-30, ACM, New York.

- Schuemie, M.J., van der Straaten, P., Krijn, M., & van der Mast, C.A.P.G. (2001): Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, **4**(2), 183-201.
- Steuer, J. (1992): Defining virtual reality: Dimensions determining telepresence. *Journal of Communication* **42**(4), 73-93.
- Taylor, T.L. (2002): Living digitally: Embodiment in virtual worlds. In R. Schroeder (Ed.) *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. Springer-Verlag, New York.
- Yong Li Lan (2003). Shakespeare as Virtual Event. *Theatre Research International*, **1**, 46-60. doi:10.1017/S0307883303000142.