The Making of a Presence Experiment: Responses to Virtual Fire

Bernhard Spanlang[1], Torsten Fröhlich[1], Vanessa F. Descalzo[2], Angus Antley[3] and Mel Slater[3,4]

[1]Departament de LSI, Universitat Politècnica de Catalunya, Barcelona, Spain

[2]Instituto de Neurociencias UMH-CSIC, San Juan de Alicante, Spain

[3]Department of Computer Science University College London London, UK

[4]ICREA-Universitat Politècnica de Catalunya Departament de LSI Barcelona, Spain

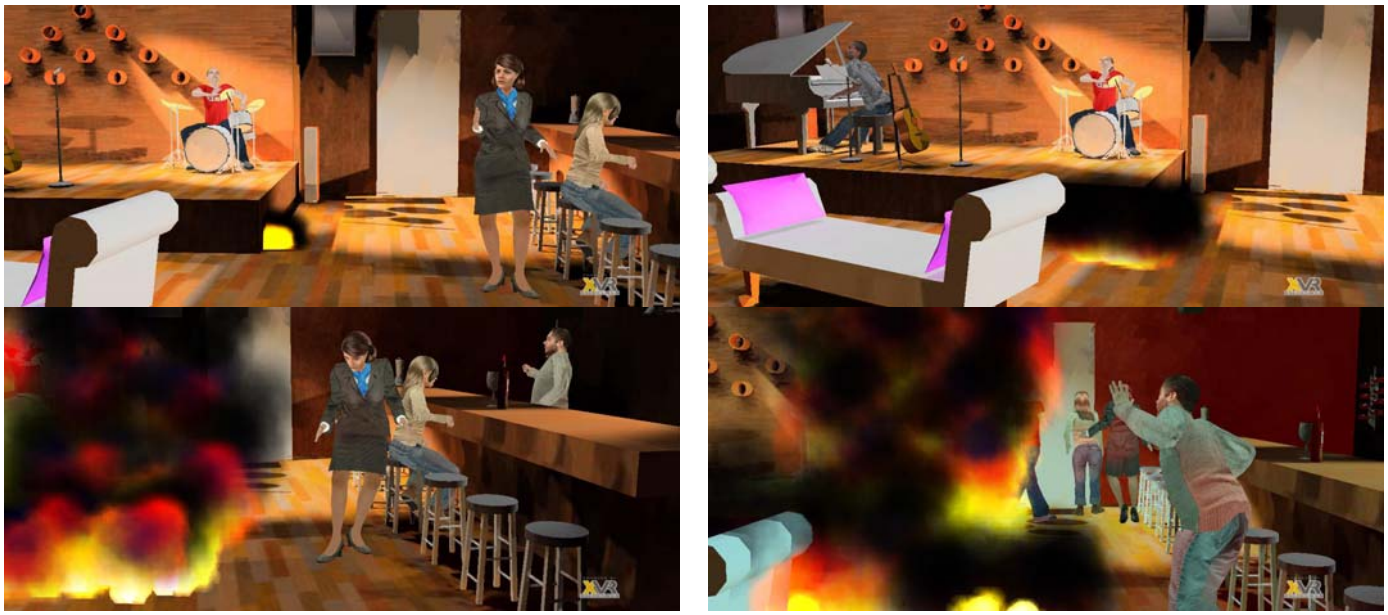

Figure 1 – Fire in the bar: The fire starts, as it spreads it is ignored, the characters occasionally glance at it, eventually they run.

Abstract

This paper describes the makings of an experiment to test various hypotheses relating to how and under what conditions people would respond to a virtual fire occurring within a social setting. An analogy is drawn with making a movie, and all the steps involved, from the specification of the hypotheses through to a full analysis of one participant are presented. Some conclusions are drawn about presence.

1. Introduction

Designing and implementing a large scale presence experiment is somewhat like making a movie. In this paper we will describe the process that led to a complex presence experiment, and also show the results in detail of one participant in the final pilot study. In this research although we follow the idea that presence is essentially a qualia associated with immersive media (the feeling of being there) we are more interested in the consequences of that feeling. We follow the idea in [1] that a sign of presence is that people respond realistically to virtual situations and events, where response is considered at multiple levels, including physiological, behavioural, emotional, and cognitive. Following this idea, we can avoid over-reliance on questionnaires in the assessment of presence, although of course they would still play a role. In this paper we show how starting from this definition, over the course of 18 months we designed and implemented an experiment, and provide detailed results from the first pilot study that employed the final version of the scenario.

2. Designing the Experiment

2.1 Basic Plot, Hypotheses

A movie starts from a short synopsis of the plot, that has to be 'sold' to the film studio and potential funder. In our case, however, instead of a simple plot, there is a research question: do people respond realistically to events and situations within virtual environments, and is this a function

of the level of immersion (type of display and interaction equipment used)? The question of the impact of the level of immersion on presence is an obvious one, but surprisingly there is very little experimental evidence about this in the literature, exceptions being with respect to multi-person environments as described for example in [2].

A particular specialisation of this question can be summarised as: Does virtual fire scare you? Does the extent to which it scares you depend on the degree of immersion of the presentation (e.g., desktop compared to Cave) and does it depend on how virtual people respond to it and to you? The idea of considering virtual fire as the basis of the scenario came from the discussion in Chapter 1 of [3] which hypothesises different levels of response of people when confronted with virtual fire in different immersive media. It was suggested there that e.g., in playing a computer game on a normal PC screen that players may manipulate their avatars out of the danger, but of course would not move their own bodies out of the way. In a more immersive environment, such as a Cave, there may be a tendency to physically get away from the source of the danger.

There has been previous work on people's responses to fire in an immersive virtual environment [4]. In that study people in a head-mounted display toured a virtual library and their spatial behaviours in response to various levels of fire were recorded. They were instructed that something would happen and that 'they were supposed to react as naturally as possible.' The natural response to fire is to move away from it, so we cannot tell whether participants moved away because of this instruction or if it was because they really wanted to get away. Also in [4] participants moved through use of a joystick rather than really moved their bodies. In our experiment we differ in these respects – we give the participants no warning that something may happen, and no instruction about how to behave. Moreover, we are interested in whether participants would really physically remove themselves from the scene rather than only move their virtual selves through a joystick.

Another area that we wished to consider was the approach proposed in [5], termed 'correlational presence'. Here the idea is that high presence is more likely the more that events in the environment correlate appropriately with the actions of the participant. For example, when virtual characters (avatars) respond to the presence of the participant in the environment this is more likely to induce corresponding responses in the participant, which should then induce responses in the avatars, setting up a natural feedback loop, as happens in real interactions in every day life. So our question here was how the response to fire would be a function of the response of other virtual characters both to the fire itself and to the participant.

2.2 Screen Play, Scenario

Having designed the basic plot this has to be turned into a scenario, the equivalent of a screen play. What happens at every single moment of the experiment? How do we transform the basic hypotheses into a fully scripted story? At this stage we have a fundamental choice whether to make this a 'between groups' or 'within groups' experiment, since there are two main factors. The first factor is degree of immersiveness of the platform used both from the display point of view (desktop, head-mounted display, one or more Cave screens, full Cave; mono, stereo, field of view), and the interaction point of view (head tracking or no head-tracking, interaction paradigms for locomotion, etc.). The second factor is the degree of interactivity of the virtual characters (none at all – they ignore the fire and the participant, they respond to the fire but not the participant, they respond to the fire and the participant). Hence there could be an $m \times n$ factorial design, in a between groups experiment, with $N \geq 5$ participants per cell, where both m and n could be quite large. This would not have been feasible within the resources of the experiment.

The design that we arrived at was a compromise, in fact both between groups and within groups. Instead of considering many different possibilities in terms of display and interaction, we would limit ourselves to the practical choices that application builders have to make: should we use a desktop, a HMD, a large screen display or a Cave for our application? Basically, we would not consider possibilities such as whether or not stereo in the Cave makes a difference, or whether or not head-tracking makes a difference, since in reality, if application builders have these capabilities they are not going to not use them! Their real choices are between types of system rather than variations within a system. Hence the experiment would be between groups for a number of different systems: a desktop system at each of two different sites in different countries, in order to use this to calibrate for the effect of cultural and language differences between these countries; two different Caves of higher and lower quality in each of these countries; a wide field of view light weight HMD in one of the countries. There would be 10 participants in each condition. The reason why this is between groups is that for this type of experiment it rules out bias effects. For example, imagine a participant who first does the experiment on a desktop, and then later repeats it in a Cave. They will know for sure what it is we are testing (it is obvious), and there is an asymmetry between the two conditions so that a randomized order design does not help.

Regarding the second factor, the responses of the virtual characters in the scenario, we decided to make this within groups, so that each participant would experience a range of virtual character behaviours. However, this had to be done in a subtle way – for example, not by running several different scenarios, in some of which the characters are non-responsive and others in which they were – for here again, the participants could easily guess what it is we are trying to find out, and therefore could result in biased responses. Instead, we would change the behaviours of the characters within the scenario itself, and examine how the responses of the participants changed accordingly.

By this time we have the structure, the basic design, but no actual story. In what context should there be a fire, how could we allow for baseline physiological readings, how could we have characters reactive and also non-reactive, and so on? Another constraint was that we were particularly interested in small group scenarios because of its connection with some of our other related research.

We decided on a social event, such as a bar, in which music would be playing and there would be a few virtual people present. Then a fire breaks out. The overall story would be as follows (Figure1):

1. The participant enters into a bar devoid of people. This lasts for just over 2 minutes and its purpose is to establish base line readings for the physiological measures. During this time music is playing.
2. Then a whiteout occurs (the screens turn white for 2s) after which the bar reappears, but this time with animated virtual characters, musicians, a drunken dancing woman, a male bar tender and another female customer. The characters pay no attention to the participant. This lasts for approximately 2.5 minutes.
3. After this time, there is a sudden loud cracking sound, like a fuse blowing, and a small fire starts at the corner of the stage. The fire gradually increases in intensity. Although one or two of the characters occasionally look towards it, there is no other reaction from them. The fire eventually becomes severe.
4. After about another 2 minutes the characters start shouting that there is a fire, and their voices show increasing signs of panic. However, they do not change their actual behaviour, so that there is an anomaly between what they are saying and what they are doing.
5. Finally, all of the characters do run out through an exit, and the participant is left alone with the fire for approximately another 30s.

We can therefore study the following situations, across the various types of display: the reactions of the participants to:

- the outbreak and spread of the fire
- the non-response of the characters to the fire
- the anomaly caused when the characters shout about leaving, but their behaviour is unchanged
- the characters running out of the bar
- standing alone in the bar with the fire raging.

None of the above includes interaction between the characters and the participant. This is because the experiment will be in two parts, where a sequel will see a modification of 2, 3 and 4 above involving the characters by the bar eventually acknowledging the presence of the participant, then issuing an invitation to join them, and finally encouraging the participant to run out with them.

2.3 Making the Scenery and Hiring the Actors

In the making of a movie a vital part of the process is to choose location and scenery. In the case of a virtual environment this requires the design of the scenario. This includes geometric design, design of the lighting, capturing and replay of sounds, and putting this all together in a consistent way. There are further complications, since we are taking physiological measures, and also recording tracking information in the Cave so that we know how much the participants moved, and therefore we also need some system to coordinate all of this, and provide data that is exactly time-stamped to events in the scenario itself, otherwise analysis is impossible. For the implementation we used the XVR system[1] which supports all the platforms that we used, and to coordinate all the different signals we used VRPN[2].

The movie producer has to hire actors suitable for the part. For a virtual scenario, we create the actors. This is a very complex and labour intensive task. For interactive real-time visualisation of the characters we have extended the 3D Character Animation Library Cal3D[3] and have integrated it with the XVR system. The motions that characters carry out are based on motion capture data which is blended, looped and scripted in Cal3D and XVR.

The visualisation system most challenging for viewing complex realistic looking characters is the Cave multi stereo-display system. All displays of the Cave system have to be synchronised to a very high degree. The Cave network renderer distributes graphics commands to each cluster machine that drives a display over a high speed Ethernet network. The aim of visualising large deformable meshes in multi display systems is to keep the network traffic as low as possible in order to permit high frame rates. Therefore we cache geometry data of characters in vertex buffers on the graphics hardware of each cluster machine during initialisation. For animation the mesh deformations are carried out on the programmable graphics hardware of all cluster machines. Only the changes of a character's skeleton are transmitted for each animation frame. This allows us to display several characters at 10K-100K polygons at very high frame rates. Of course the same strategy can be used for a single mono or stereo display or for a head-mounted display.

2.4 Rehearsals

The previous section has outlined how the overall experimental scenario was designed and implemented, which took approximately 2 person years to accomplish. What this process leads to is a kind of rough cut. The scenario and experimental design has to be tested in practice with participants in pilot studies until the researchers are 100% satisfied that there are no errors, and that the experiment can

[1] http://wiki.vrmedia.it/index.php?title=Main_Page
[2] http://www.cs.unc.edu/Research/vrpn/
[3] http://sourceforge.net/projects/cal3d

run smoothly, and it can answer the questions that were posed. However, during the course of the pilot experiments it is also likely that there will be surprises, which may change some of the goals of the experiment itself.

Altogether we carried out pilots over about 3 months, in both countries, mainly using the two Caves as the most complex platform. Here are a few highlights of what happened.

One of the most striking results was that almost universally the pilot participants reported that their level of stress was highest during the time that the avatars were ignoring the fire. This was because they were concerned about their safety. They expressed frustration that they wanted somehow to intervene, and warn the avatars about what was happening and tell them to go out, but since the characters were ignoring them, they did not feel able to take any action.

Although this was one of the most sophisticated virtual environments in every sense that the group involved had ever built, the presence scores on the 'SUS questionnaire' [6] suddenly became quite low compared to our past experience over many previous experiments, and even compared to earlier trials on this experiment, and on both Cave systems. This illustrates well the complexity of building these environments and using them for experiments. Some of the changes that we had made in the scenario as a result of the earlier pilots had led to an anomaly that was important, but that we as the creators of the system could not see. In particular, the virtual character sitting by the bar carried out some repetitive animations, but these were hidden from view by the avatar standing just in front of her. However, in a slightly modified scenario, the one at the bar could be seen in full. Her animations were not only repetitive but had a discontinuity at the end of each repetition. Attention of participants was repeatedly drawn to this, as if this one event out of everything else that was happening became dominant. Hence reported presence was very low. Once this anomaly was removed, the presence scores reverted to normal levels compatible with what we have seen in experiments over many years.

3. The TV Pilot

Just at the point that we had completed the pilot studies and were about to recruit for the experiments, we were approached by a TV company in Cataluyna, who had somehow heard about the experiment, and wished to include it in a science programme. It was decided that the best way to do this was for a person completely naïve to the goals of the study, but working for the TV company, to actually do the full experiment, including all the information sheets, questionnaires, physio recordings, ethical forms, and so on. This was agreed, and it was decided to do a study of her reactions. She was therefore the first experiment participant with the completely working and pilot-tested environment.

3.1 Questionnaires and Behaviour

Her average score on the presence questionnaire was 4.8 out of a maximum of 7, indicating quite a high level of reported presence, well in line with many studies over the years. During the scenario the participant was seen moving towards the virtual characters near the bar, and waving back at the dancing avatar whenever she waved. Once the fire started she at first stepped away from it, and then later attempted to douse the fire by shaking the hem of her dress near it, as if to blow air towards it.

3.2 ECG Analysis

We recorded ECG at 256Hz, respiration at 32Hz and skin conductance at 32Hz, using a Mind Media Nexus physiological recording device, integrated into the VRPN and XVR programs as mentioned above.

From the ECG signal heart rate (HR) and heart rate standard deviation (SDHR) were extracted. We divided the time into a number of intervals and show the heart rate, variability and the ratio of these in Table 1.

	HR bpm	SDHR bmp	HR/SDHR
(a) Baseline to Start of Peopled bar	88.6	7.4	11.9
(b) Start of Peopled bar to Start of Fire	90.1	6.8	13.3
(c) Start of fire to people shouting	92.2	6.1	15.1
(d) People shouting to runout	94.7	9.9	9.5
(e) People run out to the end	95.1	5.0	18.8

Table 1 – Heart Rate and Heart Rate Variability

Typically, high heart rate and low heart rate variability indicates stress. So as we would expect during the baseline the heart rate is relatively low, and the variability is relatively high. When the real scenario starts the heart rate increases a bit and the variability goes down. It is normal for this to occur at the start of any virtual reality experiment, since the participant is in a new situation that holds a number of unknowns for them. When the fire starts the heart rate increases still more, and the variability decreases again, indicating more stress. When the people start shouting although the heart rate goes up so does the variability. An interpretation was that there was some feeling of relief that finally the virtual characters were noticing the fire and were preparing to leave. However, after they finally run out and the participant is alone with the fire, this seems to be the most stressful situation of all – the ratio goes to the maximum, the heart rate is at the highest, and the variability at the lowest. This would suggest that although the situation was not real, in terms of the heart activity there were responses that fitted with the situation.

3.3 Skin Conductance

This is the electrical activity caused by changes in sweat on the skin. What is important is not the absolute level, but the sudden changes. These indicate moments of arousal. The number of such moments, called "Skin Conductance Responses" (SCR) indicates the overall level of arousal. The rates of SCRs per 10s in the same time periods as in Table 1 are 5.2, 4.7, 4.7, 4.5, and 6.0. The level of arousal during the baseline was high but then stabilised until after the people had run out of the bar, and the participant was left alone. Here it reached its highest level, with nearly one SCR per second. So in this period there was relatively high arousal and high stress.

3.4 Respiration

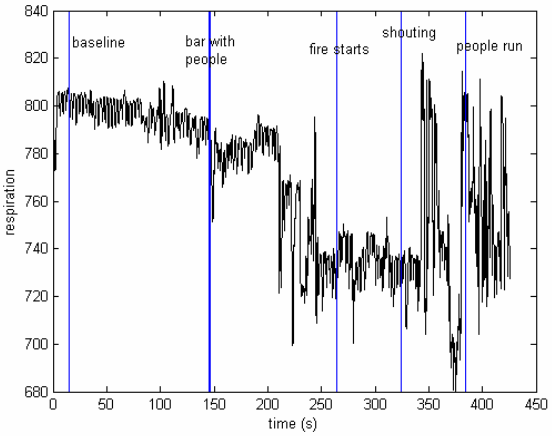

Figure 2 – Respiration throughout the experiment

Figure 2 shows the changes in respiration in the different segments of the experiment. Respiration was shallow and constant during the baseline period and the participant was holding her breath. It became more agitated during the initial period in the bar once the people were there and then once the fire started became relatively shallow and constant, and then became very agitated during the period after the characters were shouting and then left the bar.

3.5 Interview

In the interview that followed she said that she felt concern for the virtual characters when they ignored the fire, and felt relief when they finally reacted appropriately, which was in line with what almost all of the pilot subjects had previously told us. She also said that she felt that she needed to interact with the characters, and attempted to do so by waving back at the female avatar whenever she waved. She asked the experimenters what could be learned from such an experiment. The answer given was: Never attempt to put out a fire by waving your dress around to blow air over it.

This last point is an important one. We are interested in presence because in principle it informs us about how people might behave in a similar real situation. In a context such as people's response to danger this is a crucial aspect, since such lessons can be learned in a safe environment.

4. Conclusion

This paper has tried to give some insight into the complexities involved in making a presence experiment. Also it should be realised that there is significant financial cost, probably when all is taken into account in excess of €150,000 (salaries, equipment, software, consumables, and payments for participants). We have found from our pilots that people did tend to respond with some level of realism to the virtual fire (e.g., another pilot participant covered her eyes, and stepped back away from the fire, some reported feeling heat, and even smelling the smoke). So some level of presence occurs. However, when faced with a real fire of such ferocity, would people just stand there, and behave the way that our participants did? It is true that their hearts race, and they show other physiological signs of stress, but they really do not actually behave as if it were real!

One way to think about the goal of presence research is that to be successful it should be able to discover what would be necessary to make people actually and physically run away from a virtual fire.

Acknowledgements

This research is funded by the EU FET PRESENCCIA project Contract Number 27731. Thanks to Franco Tecchia, Marco Gillies and David Swapp for their excellent help and support.

References

[1] Sanchez-Vives, M.V. and M. Slater, *From Presence to Consciousness through Virtual Reality.* Nature Reviews Neuroscience, 2005. 6(4): p. 332-339.

[2] Schroeder, R., A. Steed, A.S. Axelsson, I. Heldal, A. Abelin, J. Widestrom, A. Nilsson, and M. Slater, *Collaborating in networked immersive spaces: as good as being there together?* Computers & Graphics-Uk, 2001. 25(5): p. 781-788.

[3] Slater, M., A. Steed, and Y. Chrysanthou, *Computer Graphics and Virtual Environments: From Realism to Real-TIme*. 2001: Addison-Wesley.

[4] Gamberini, L., P. Cottone, A. Spagnolli, D. Varotto, and G. Mantovani, *Responding to a fire emergency in a virtual environment: different patterns of action for different situations.* Ergonomics, 2003. 46(8): p. 842-858.

[5] Gillies, M. and M. Slater. *Non-Verbal Communication for Correlational Characters*. 2005. London: University College London - Department of Computer Science.103-106

[6] Usoh, M., E. Catena, S. Arman, and M. Slater, *Using presence questionnaires in reality.* Presence-Teleoperators and Virtual Environments, 2000. 9(5): p. 497-503.