Reinforcement Learning Utilizes Proxemics: An Avatar Learns to Manipulate the Position of People in Immersive Virtual Reality

IASON KASTANIS
Universitat de Barcelona
MEL SLATER
ICREA-Universitat de Barcelona

A Reinforcement Learning (RL) method was used to train a virtual character to move participants to a specified location. The virtual environment depicted an alleyway displayed through a wide field-of-view head-tracked stereo head-mounted display. Based on proxemics theory we predicted that when the character approached within personal or intimate distance to the participants, they would be inclined to move backwards out of the way. We carried out a between-groups experiment with 30 female participants, with 10 assigned arbitrarily to each of the following three groups: In the Intimate condition the character could approach within 0.38m and in the Social condition no nearer than 1.2m. In the Random condition the actions of the virtual character were chosen randomly from amongst the same set as in the RL method, and the virtual character could approach within 0.38m. The experiment continued in each case until the participant either reached the target or 7 minutes had elapsed. The distributions of the times taken to reach the target showed significant differences between the three groups, with 9 out of 10 in the Intimate condition reaching the target significantly faster than the 6 out of 10 who reached the target in the Social condition. Only 1 out of 10 in the Random condition reached the target. The experiment is an example of applied presence theory: we rely on the many findings that people tend to respond realistically in immersive virtual environments, and use this to get people to achieve a task of which they had been unaware. This method opens up the door for many such applications where the virtual environment adapts to the responses of the human participants with the aim of achieving particular goals.

Categories and Subject Descriptors: H5.1 [Information, Interfaces and Presentation]: – Multimedia Information Systems Artificial, Augmented and Virtual Realities. I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality
General Terms: Experimentation, Human Factors
Additional Key Words and Phrases: human-computer interaction, proxemics, virtual characters, avatars.

1. INTRODUCTION

There has been significant research over the past two decades indicating that people tend to respond realistically to situations and events portrayed within immersive virtual environments (IVE), see [Sanchez-Vives and Slater 2005] for a review. Much of this has been under the umbrella of 'presence research' which studies the extent to which participants experience a sense of being in the place depicted by the IVE displays [Held & Durlach 1992; Sheridan 1992; Barfield & Weghorst 1993; Slater & Wilbur 1997; Draper et al. 1998]. One particular example of such realistic response is 'proxemics'. Results that show that people tend to behave in IVEs according to the predictions of proxemics theory with respect to social encounters between people in reality - that is, they tend to maintain socially appropriate distances between themselves and virtual characters they encounter in the IVE, e.g., [Bailenson et al. 2003]. In this paper we exploit these earlier results to show that a machine learning (ML) agent represented by a virtual character (or 'avatar') in an IVE can learn to manipulate the behavior of people so that they achieve a task that they had not previously been instructed to carry out. In particular we show that proxemics may be exploited by the ML agent to control the avatar in such a way that it learns to drive participants to a certain pre-specified position in the virtual space, even though the participants were unaware that reaching this point in space was their task.

In this paper by 'agent' we mean the underlying ML process that learns to achieve its goal using real-time data based on the actions of the human participant. The particular ML method that we have used is Reinforcement Learning (RL), but we do not make any special claims that this is the best or ideal method for the problem, it is only illustrative of the general paradigm that we are proposing - that based on an assumed realistic response by people to situations and events in an IVE an agent can learn to manipulate their behavior. Finally, by 'avatar' we mean a virtual humanoid character that represents the state of the agent by its position in the virtual space relative to the participant's, and also the actions determined by the agent, represented by its walking and gestural behavior.

There have been numerous papers investigating the concept of presence, using questionnaires and qualitative methods, e.g. [Lessiter et al. 2001; Witmer et al. 2005; Witmer and Singer 1998; Garau et al. 2008; Schubert et al. 2001; Usoh et al. 2000], and behavioral studies including physiological and brain-activation responses, e.g. [Pan et al. 2008; Razzaque et al. 2002; Zimmons and Panter 2003; Meehan et al. 2003; Hoffman et al. 2003; Meehan et al. 2002; Hoffman et al. 2004; Rey et al. 2008]. These have studied whether people tend to respond to events and situations in virtual environments realistically. It has been repeatedly shown that this 'response-as-if-real' (RAIR) does occur in various ways during exposures to IVEs.

This research was supported by the EU's 7th Framework Program project MIMICS (Grant Number 215756).
Authors' addresses: Iason Kastanis, i.kastanis@ub.edu, Facultat de Psicologia, Campus de Mundet - Edifici Teatre, Passeig de la Vall d'Hebron 171, 08035 Barcelona – Spain; www.event-lab.org;

In the research presented in this paper we have used an inverse strategy in comparison with previous work. Instead of carrying out experiments that manipulate factors in order to examine people's responses, here we assumed that RAIR would occur, and consequently whether an agent could exploit this to learn the conditions under which the participant can be driven to realize a particular task. In this case we assumed that proxemics behavior would be activated, and that this could be the basis for an agent to learn how to get participants to move to a particular pre-specified location, by exploiting the fact that they would be likely to move away when approached too close by an avatar representing the agent. This is similar to applications that implicitly rely on the assumption that RAIR occurs – especially, for example, in psychotherapy, where it is required that patients do experience sufficient anxiety when confronting the feared situation in a virtual environment, e.g. [Wallach et al. 2009; Krijn et al. 2004; Baas et al. 2004; Harris et al. 2002; Rothbaum and Hodges 1999; Rizzo et al. 2005; Krijn et al. 2007; Rothbaum et al. 2000; Regenbrecht et al. 1998; Hodges et al. 1995] including the use of IVEs in the study of paranoia, e.g. [Freeman et al. 2008]. However, here we systematically exploit a particular RAIR - proxemics - in order to achieve a very specific target behavior.

Hall [1973] introduced the term proxemics which is concerned with implicit social rules of interpersonal distance that people maintain, each conveying a different social meaning. Hall proposed four such regions of interpersonal distance: public ($\geq$3.7m, <7.6), social ($\geq$1.2m, <3.7), personal ($\geq$0.46m, <1.2), and intimate (<0.46m) (Figure 1). A person might feel embarrassed or surprised when these rules are not adhered to and may try to back away if, for example, someone approaches them within the range of intimate distance, the inner most region in Figure 1, when this is not appropriate given the relationship between them. It has been shown that such proxemics rules tend to occur in virtual environments [Bailenson et al. 2003; Bailenson et al. 2001; Guye-Vuilleme et al. 1999; Friedman et al. 2007; Wilcox et al. 2006] in relationships between people and virtual human characters, and recently it has been shown that proxemics violations generate physiological responses in people with respect to skin conductance as predicted by the theory [Llobera et al. 2010].

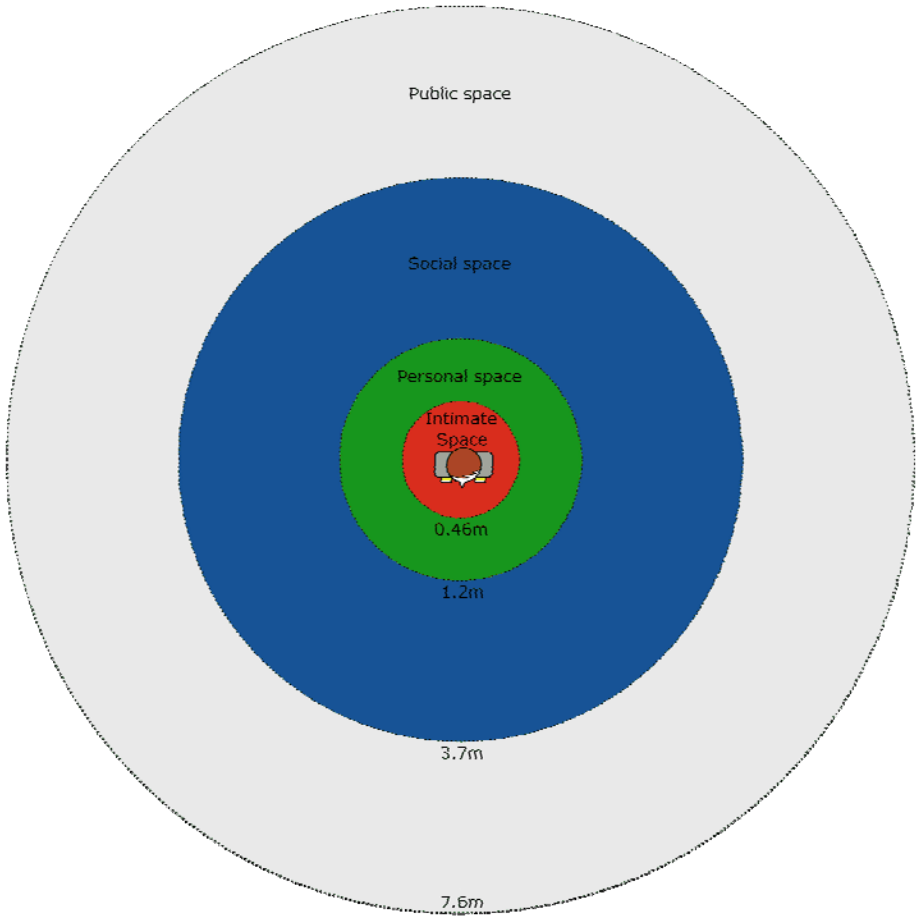

Figure 1 PROXEMICS zones according to the theory of Hall [1973].

2. HYPOTHESES AND EXPERIMENT OVERVIEW

The experiment reported in this paper places a participant in a virtual environment depicting an alleyway, seen through a wide field-of-view, lightweight, head-tracked and stereo head-mounted display. There a male avatar beckons to the participant to approach towards him, or moves itself towards or away from the participant. Our expectation is that people would tend to obey rules of proxemics and react by moving backwards when the avatar gets too close.

This expectation is based not only on previous experiment results discussed above, but also it follows from our theoretical framework which has deconstructed presence into two orthogonal concepts – place illusion (PI) and plausibility (Psi) [Slater et al. 2010]. In the experiment reported here the participant perceives the world through moving her body in a natural way (e.g., head turns result in appropriate visual update) and even navigating using a form of body-centered interaction. This employment of realistic sensorimotor contingencies for perception typically results in PI, the strong feeling or qualia of being in the place depicted by IVE. Moreover in the IVE there are events that clearly relate directly and personally to the participant (the avatar calls out to her and waves, or moves closer but stops in front) – one aspect that contributes to Psi, the illusion that the events happening in the IVE are real. When PI and Psi both occur the prediction is that people would be more likely to exhibit RAIR.

Our aim in this experiment here is not, however, to *investigate* presence or proxemics, but to use these as mechanism for making participants achieve a certain goal within the IVE. The focus therefore is on whether the system can learn from the participant responses and use these to achieve the goal.

30 people were recruited to an experiment designed with the purpose of testing this idea. Two conditions that used RL based on interpersonal distances were examined. One was where the avatar could approach the human up to 0.38m, within the intimate region, and another one where the avatar could only approach the human up to 1.2m, within the social region (Figure 1). The third condition was a control to test the performance of the RL method by comparing it to a random choice of the agent's actions, but where the avatar also could approach the participant up to a distance of 0.38m.

Our first hypothesis is that RL will perform better than the random condition in which significantly less people (if any) would reach the target, and which anyway would take a far longer time than those in the RL conditions. The second hypothesis was that in both the intimate and the social conditions the participant might be driven to the target, but based on proxemics theory this would be more efficiently achieved in terms of time taken by those in the intimate compared to the social distance conditions.

3. REINFORCEMENT LEARNING (RL)

RL is a class of machine learning algorithms that aims to learn response functions by interacting with the environment and receiving rewards. An agent receives the state of the environment and selects an action. As a consequence of this action it receives a reward. The goal of RL methods is to maximize the long-term reward based on past experience. Originally developed in the artificial intelligence community using ideas from animal learning theory, it now forms a common research language between biologists and engineers [Doya 2007] with both sides progressing the field further. The use of Reinforcement Learning methods is typically found in robotics applications. Walking robots are discussed in [Collins et al. 2005] where RL is used to obtain control policies for the articulated parts.

There has also been a growing interest in applying RL methods [Sutton and Barto 1998] in controlling characters in both the robotics [Collins et al. 2005; Vigorito 2007; Wu et al. 2009] and computer graphics communities [Ikemoto et al. 2005; McCann and Pollard 2007; Lo and Zwicker 2008; Shum et al. 2008; Lee and Popović 2010; Friedman and Gillies 2005]. Typically these applications aim to control mechanical issues such walking, facial animations and moving within a VE, their goal being a high level of realism of motor behavior. Wu et al in [Wu et al. 2009] used RL for controlling the parameters of the inverse kinematic model of the 31 servo motors in order to produce realistic facial expressions. The expressions were learned by comparing them to an image database with the use of facial recognition software. In [Vigorito 2007] RL was used to plan paths for robots in rough terrains. Similar applications of RL can also be found in the computer graphics community. Friedman and Gillies [2005] used RL to teach avatars to use body language. The user of their system assigned rewards manually in order to train the character to choose the appropriate body language for the given context. They used RL to calculate the optimal path, in which the avatar avoids obstacles and hostile entities while trying to keep the length of the path as short as possible.

Controlling character animation is a challenging problem in the computer graphics community. One of the difficulties arises from the problem of balancing realism with precision in the animation. A RL solution to this problem was proposed in [McCann and Pollard 2007] where data driven animation for game controllers was created. The aim was to exactly combine the ability to control the character precisely with the quality of the produced motion. Lo and Zwicker [2008] used data from motion capture, as well as parameterized motion and interpolations for similar purposes. In [Treuille et al. 2007] the aim was to produce high fidelity animations while performing complex navigation and obstacle avoidance tasks. Shum et al [2008] used RL to create avatar controllers while efficiently handling the high dimensional state space of animations. In a recent paper by Lee and Popović [2010] controllers with different styles were generated by an RL approach. The different styles, namely normal, playful and watchful defined how the avatar behaved towards obstacles in a navigation task. These controllers were calculated by providing examples of the desired style and estimating the reward function. The estimation of the reward function was performed by a modified apprenticeship algorithm [Abbeel and Ng 2004].

In contrast to previous applications we provide an example of the use of RL to make human participants in an IVE achieve particular goals through interaction with an avatar controlled by an RL agent. Here, therefore, RL is not intended to increase the realism of the behavior of the avatar, but for the avatar to learn to generate a particular behavior in the human participant.

4. MATERIALS AND METHODS

4.1 Participants

Thirty female participants were recruited by advertisement around the campus. Their mean age was 26 ± 8 (S.D.) years. After initial trials it was obvious that dyadic interactions in terms of interpersonal distances were dependent on the gender of both the participant and the avatar. Thus we decided to recruit only in one gender in order to remove one possible source of variation, and females for the practical reason that 85% of the students on our campus are female. They were paid €5 for their participation in the study. The experiment was approved by the ethics committee of IDIBAPS at the Hospital Clinic of University of Barcelona.

4.2 Materials

A Fakespace Labs Wide5 stereo head-mounted display was used which has field of view of 150º × 88º with an estimated 1600 × 1200 resolution displayed at 60Hz[1], and also

[1] http://www.fakespacelabs.com/files/Download/Wide5%20Data%20Sheet.pdf

Sennheiser HD215 stereo headphones. The head-tracking system was an Intersense IS900.

The virtual environment was programmed and displayed using XVR [Carrozzino et al. 2005]. For animating the avatar the HALCA library (Hardware Accelerated Library for Character Animation) was used [Gillies and Spanlang 2010; Mortensen et al. 2008]. The agent was controlled by BAREL (Basic Reinforcement Learning Toolbox) with the Q(λ) algorithm for discrete state spaces [Sutton and Barto 1998].

4.3 The Scenario

The scenario depicted an alleyway with the agent and the participant positioned as shown in Figure 2. The participant could see facing towards her a male avatar 10m away. In Figure 3 the environment can be seen from the perspective of the human participant.

The avatar could perform the following actions: be in idle mode (standing still but fidgeting slightly in order to give the impression of aliveness), walk towards the participant, look behind himself and walk backwards or wave towards the participant while calling ‘Come here!’. The avatar was not programmed to make direct gaze contact with the participant, although this could occur by chance. Idling animations were displayed between the end of an action caused by the agent and the start of a new one. They lasted for a mean time of 2.4±0.4s (S.D.) the exact time chosen randomly. Walking forward took 0.8s and 1.5s for 1m and 2m, respectively. Walking backwards was slightly slower taking 1.8s and 3.2s. The idle action took 4.2s to complete and the waving 3.7s. The waving was accompanied by a randomly selected voice recording of one of the following phrases, ‘Come come!’, ‘Come here’, ‘Hey come here’, ‘Hey here’ and ‘Over here’. These were said in Catalan, Spanish or English according to a prior language choice of the participant.

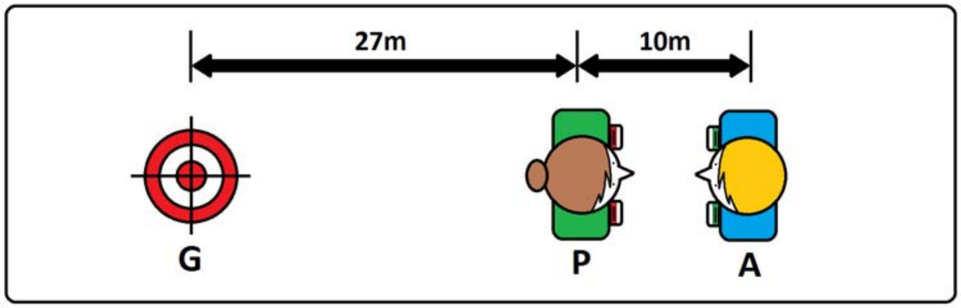

Figure 2 A top down diagram of the environment at the start of the experiment. G represents the goal, P the participant and A the agent. P can only move forward or backwards and cannot cross behind A.

4.4 The Reinforcement Learning Parameters

The aim of the RL agent was to find which actions of the avatar caused the participant to move towards the target. The state of the agent was calculated according to the distance of its avatar from the participant. The values of the agent's state were chosen from proxemics theory: the avatar was within personal distance (<1.2m), social distance (<3.7m), public distance (<7.6m) or not-engaged (>7.6) (Figure 1). In each state the agent could determine the avatar to take one of the following 6 actions: walk forward for 1m, walk forward 2m, walk backward 1m, walk backward 2m, idle and wave. After each action the avatar in any case stood in idle mode for 2.4s, to give time for the participant to respond. The avatar and the participant could not move past each other. If they tried to do so they simply remained in the same positions in the VE.

Rewards were calculated by considering the distance from the target in the VE travelled by the human when the avatar performed one of the six actions and then stood in idle mode for about 2.4 seconds, as previously explained. The reward function was based on the response of the human. If the person moved towards the target the RL agent received a positive reward (1 + response distance in m) and in all other cases a negative reward (-1). The target was positioned 27 meters behind the starting position of the human participant (Figure 2). The Q(λ) algorithm, used for RL, takes advantage of the concept of eligibility traces, which bridges between temporal differencing and Monte Carlo methods. The spectrum between these two families of methods is spanned by the λ parameter, which defines how far back in the history of state-actions a reward will be assigned.

The algorithm was initialized without any training data. The agent automatically adapted individually to each participant. This was possible due to the compact state-action space, 4 by 6 for the intimate condition and 3 by 6 for the social condition. The parameters of $Q(\lambda)$ were as follows: λ was set to 0.7, the rate of learning α was equal to 0.2 and the discount factor γ to 0.15. Finally an ε-greedy policy improvement was used with ε set to 0.15.

4.5 Procedures

Participants were first asked which language they preferred for the experiment: Catalan, Spanish or English, in order to accommodate the variety of native languages on the campus. The participant was given an information sheet to read and the experiment was also verbally explained to them. If they agreed to do the experiment (all did) then they completed the consent form and a short questionnaire that obtained demographic information such as age, occupation, and so on. The participant was fitted with the HMD, the tracker and the headphones. After they had donned the HMD, they were asked to stay still and during that time their head position was recorded. Participants were also asked to keep their legs in a fixed position on the ground.

Then a training period started where they learned how to navigate in the VE, and also acclimatized to the environment; during this time the avatar was not shown, and of course the RL process was not operating. It was explained and demonstrated to them that they could move forwards and backwards in the VE by leaning without moving their feet from the floor. A forward lean would move the participant forward in the VE, and a backward lean would move them backwards. The participant always faced in the same direction. This is similar to the method for navigation discussed in [Mine et al. 1997]. As part of the training the participant was then instructed to go to particular locations in the VE and observe the posters and paintings on the walls. Once the participant was capable of moving to the required locations, and was also able to remain stationary by not leaning forwards or backwards, the training session was terminated. Finally the participant was asked to close her eyes while the training environment was switched off and the new one loaded.

At this time participants were told: ‘There will be another person in the environment, please move freely in order to feel comfortable.’ Once the participant was inside the new environment, their position was calibrated and the main stage of the experiment began.

The scenario started with the avatar signaling to the participant. The RL agent controlled whether the avatar would idle, wave, or move forwards or backwards according to the participant's response, i.e. her change of position within the VE. The experiment terminated when the participant had moved to the target location (Figure 2), 27 meters behind her starting position in the VE. If the participant had not reached the target location by the end of 7 minutes the experiment was terminated, and the participant removed the HMD and earphones. The participant was then debriefed and paid the €5.

4.6 Experimental Design

This was a single factor between-groups design with three conditions, with 10 participants arbitrarily assigned to each condition. The Intimate condition (I) used RL and the avatar could move up to intimate distance to the participant. The Social condition (S) used RL and the avatar could move up to social distance to the participant. In the Random condition (R) the actions chosen by the agent were the same set as in the RL conditions except that they were selected uniformly randomly. In this condition also the avatar could move up to intimate distance. Specifically, in conditions I and R the avatar could invade the participant's personal space (up to 0.38m) therefore reaching inside the intimate distance, while in condition S it could only go as near as 1.2m (Figure 1).

The maximum of 7 minutes duration was chosen by taking into account the number of possible actions by the avatar, the probability of a step backwards, and the possible size of that step. This had been fine tuned after running several pilot studies in order to allow for the experiment to be completed successfully or otherwise to be terminated without keeping the participant in the virtual reality system for too long to be comfortable.

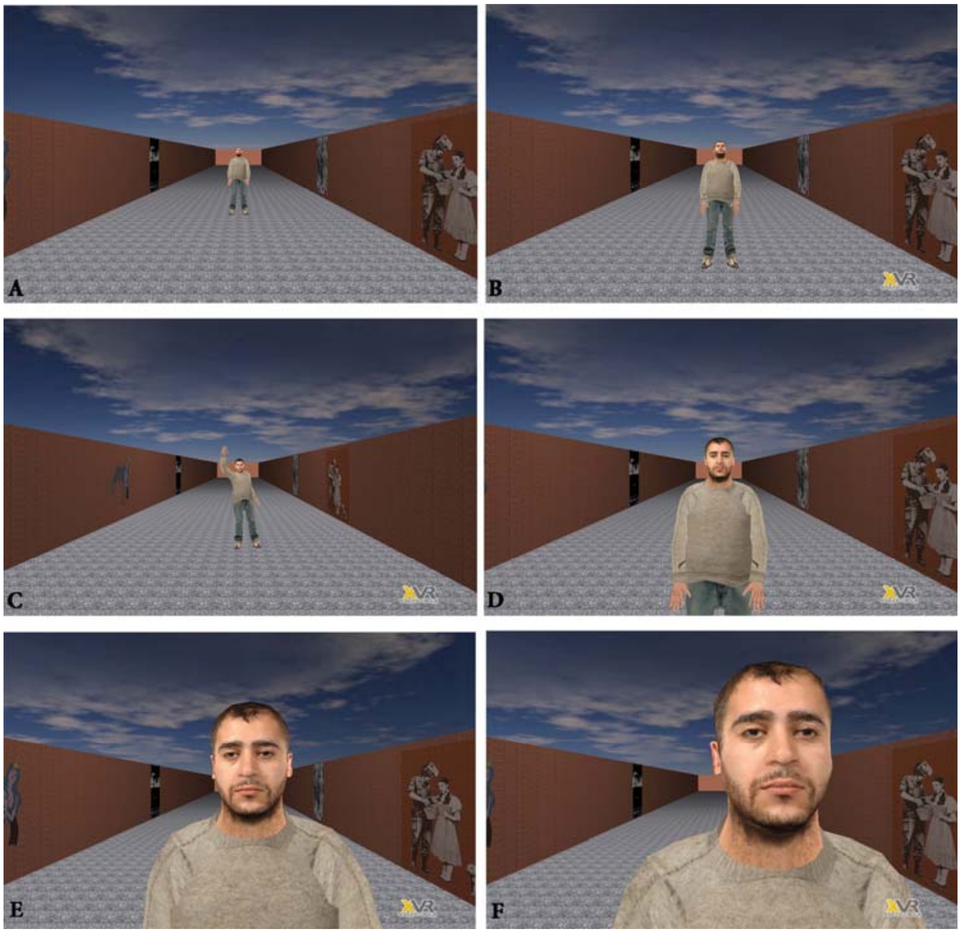

Figure 3 The scenario is an alleyway that contains a avatar seen from the participant's viewpoint: (A) the character is quite far away (B) he has walked closer (C) calls to the participant to 'come here!' while waving (D) has approached closer (E) is within personal distance (F) is within intimate distance.

4.8 Response Variables and Statistical Methods

There are two main response variables of interest. The first is the proportion of participants who reached the target. The second response variable is the overall time taken – either to reach the target or to terminate after 7 minutes without reaching the target.

The time distribution within each condition is an example of survival data, how long the participant 'survived' before being driven back out of the alleyway. Since there was a cutoff time of 420s this is *right-censored survival data* since we cannot know for any individual whether given enough time they might eventually have reached the target. This type of survival data can typically be modeled by the lognormal distribution, which arises when there is a system that is subject to many shocks, which eventually 'fails'. This is quite similar to the case of this experiment, where the participant is subject to many influences (the actions of the agent) and eventually the participant may be driven out of the alleyway (here of course this is a 'success' rather than a 'failure' – but this is only a matter of nomenclature). A random variable y has a log-normal distribution with parameters μ and σ when log(y) has a normal distribution with mean μ and standard deviation σ. Our method was to first fit the censored data by lognormal distributions and then use the Kolmogorov-Smirnov test to examine the goodness of fit. Then we provide confidence intervals for the parameters of the distribution.

5. RESULTS

It is clear from Figure 4 that in the RL conditions a far greater number of participants reached the target than in the random condition. In the R control condition only 1 out of 10 participants reached the target, while in the I condition there were 9 out of 10 and in the S condition 6 out of 10 participants who reached the target position within the 7 minutes.

Figure 5 shows the cumulative distribution functions (CDFs) for the I and S survival times. For example, it can be seen that at, say, 220 seconds 80% of those in the I condition had reached the target compared with 30% of those in the S condition. The first successful termination in the I condition occurred in just over 40s whereas in the S condition it was just over 100s. Already by just over 60s half of those in the I condition had reached the target, whereas in the S condition this occurred only after 340s. It is clear that the distributions of the time courses are quite different in the two cases and the Kolmogorov-Smirnov (KS) test rejects the hypothesis that the two samples are from the same distribution ($P = 0.03$). Note that the significance level here is an underestimate due to the censoring.

Moreover KS tests do not reject the hypotheses that the two censored survival time data sets (for I and S) are from lognormal distributions ($P = 0.49$ in the case of I and $P = 0.07$ for S).

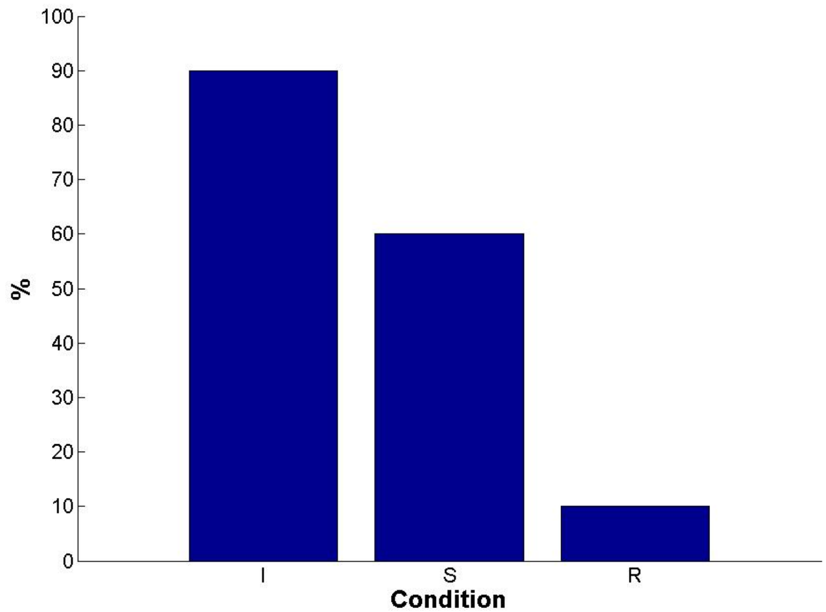

Figure 4 The percentage of participants reaching the target under the 3 conditions.

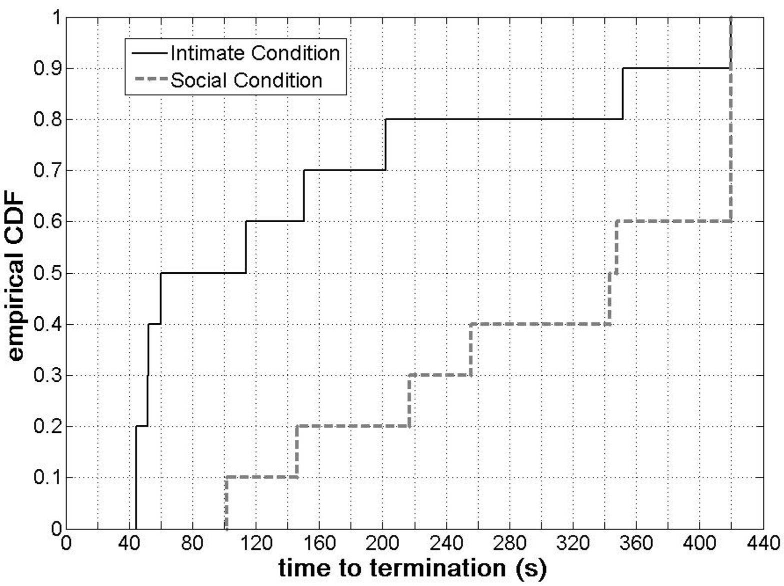

Figure 5 Empirical Cumulative Distribution Functions for the Intimate and Social Conditions.

Table 1 shows the maximum likelihood estimates for μ and σ and the corresponding 95% confidence intervals for each of the I and S conditions. While there is no significant difference between the I and S distributions with respect to σ, there is in the case of μ. To give an idea of the meaning of these parameters it is simple to use them to evaluate the mean and standard deviation of the log-normal distributions, shown in the last two columns. These are the estimates of the theoretical population values corresponding to what we would have observed for a very large data set. Note that since the method takes account of the censoring the estimated mean of the distribution for S is actually greater than the cutoff point of 420s. In other words although the variance between the two groups are not significantly different, the mean time to reach the target is far greater for the S group (444s) than the I group (164s) taking into account the censoring of the data. The difference between the two fitted probability density functions as a whole can also be seen in Figure 6.

Table 1 Parameter Estimates for the Lognormal Distribution Fits and their 95% Confidence Intervals.

Condition	$\hat{\mu}$	95% CI for $\hat{\mu}$	$\hat{\sigma}$	95% CI for $\hat{\sigma}$	Mean sec.	SD sec.
Intimate	4.69	4.13 - 5.26	0.90	0.56 - 1.45	164	183
Social	5.84	5.34 - 6.34	0.72	0.38 - 1.33	444	364

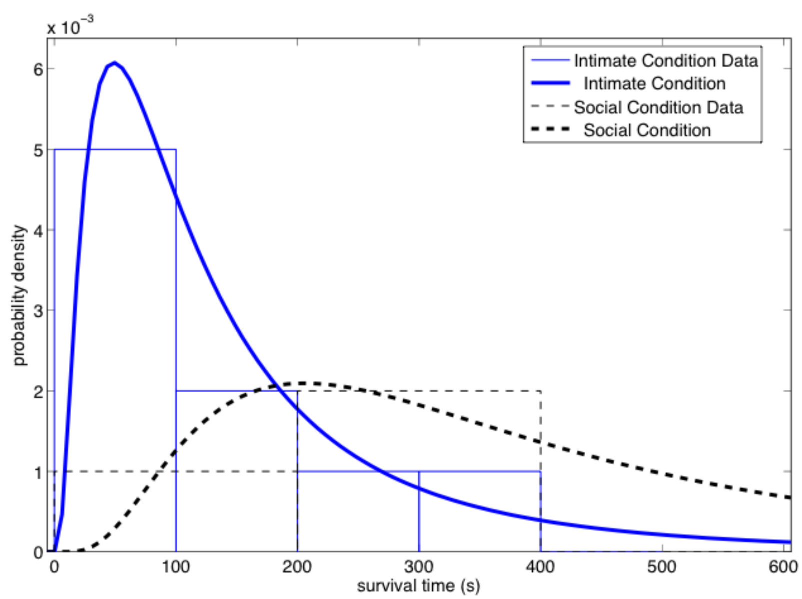

Figure 6 Fitted lognormal probability density functions to the I and S condition censored survival times.

Finally, it is worth considering the wave action in I and S conditions. After the initial piloting stage we found that when the agent was too far away the responses of the human became irrelevant to the agent's decisions. We therefore added the wave action, where the agent would wave his hand and call the participant, in order to get her attention. In Figure 7 it can be seen that the wave action caused the participant to change her position in the IVE the most while in the furthest away state (at least 7.6 meters away), by moving away from the target and towards the agent, which, of course, generates a large negative reward for the agent. However, this very well illustrates the operation of the RL algorithm that has the objective of maximizing its *long term reward*. It is interesting to note that the agent learned to choose actions that sacrificed short term reward for the ultimate maximization of its long term reward.

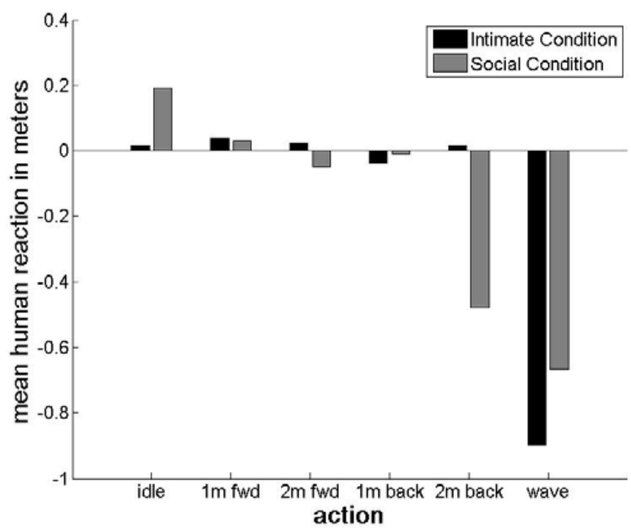

Figure 7 Mean human reaction, i.e. participant walks towards the target, for each of the possible actions of the agent, when the human participant was in the furthest away state (≥7.6m).

6. DISCUSSION

In this experiment the agent learned, using RL, the rule that when it moves so close to the participant that it breaks the convention of personal distance, the participant tended to move backwards. The RL agent also learned that generally the action of moving towards the participant only had an effect on her response when she and the avatar were already close enough. Hence the agent also discovered that it is possible to get the participant closer to the avatar when the distance between them is too great, and thereafter exploit proxemics to drive the participant backward to the far end of the alleyway. In a second condition, where the avatar never could enter within intimate distance, the proxemics rule was learned less efficiently and the agent's goal was less likely to be achieved.

The typical approach in presence research is to design an experiment that manipulates various factors, for example frame rate [Barfield and Hendrix 1995], navigation method [Usoh et al. 1999], rendering quality [Slater et al. 2009] etc., where the participants carry out a known task, and then the responses of the participants are examined with respect to how much presence or 'response-as-if-real' they exhibit. Here we have turned this on its head, and instead we have assumed the response-as-if-real, and the system learned to exploit this in order to get the participants to realize a task (in this case go to a certain location) that they were not previously informed about. Only three out of the twenty participants in the two RL conditions had worked out the goal of the agent after the experiment. The method is actually directed at the behavior of the human participant, and manipulates the behavior of the avatar that represents the agent only as a means to this end.

Studies of proxemics in IVE [Llobera et al. 2010; Bailenson et al. 2003] have added to the growing evidence that people respond realistically in IVEs and in particular tend to obey proxemics rules even in interaction with avatars. In prior work the actions of the avatars were determined in advance by the computer program. Here although the set of actions open to the avatar were fixed in advance, which sequence of actions it would take depended only on how the participants behaved. One interesting finding is that we learned that even in this simple case of a 1D walk, it was not possible to rely on only three types of action (our original intention) – idle, move forward or move back. A wave action was added that drew the attention of the participant to the avatar. This type of action was necessary to maintain the interaction or re-engage it and create the closed system needed for the successful application of this methodology.

Since the human and avatar could not walk past each other it might be thought that the method was bound to succeed. This is not the case, since the whole point of the approach was for the agent to learn that it had to get as close as possible to and therefore 'trap' the participant so that she could only go backwards or do nothing. The fact that some people did not reach the target (1 out of 10 in the I condition , 4 out of 10 in the S condition and 9 out 10 in the R condition) also shows that there was no inevitability built into the system.

Although we have argued that the results of the experiment are based on the fact that proxemics was operating, it is fair to note that there are alternative explanations. For example, the reason for the relative success of the method might not be because of the specific reason of *social proxemics* but simply because the female participants were afraid of the male avatar. In fact we have no way of knowing that, but even if it were the case it still does not affect our fundamental idea, since it would only mean that instead of exploiting proxemics the RL may for some participants have been exploiting some other type of response. However, the fact that the results were different between the Intimate and Social condition suggests that proxemics did operate. The most important point is that whatever the underlying reason, the agent did learn the appropriate behavior. Additionally proxemics was the starting point of this work, and is the simplest explanation for the result, especially given the weight of evidence from previous research.

Another possible problem with our results could be the fact that egocentric distance judgments are often mistaken in head-mounted displays. Studies show that usually distance estimation is typically compressed to between 70% and 90% of veridical distance when estimating distances of between 3m and 15m. For a recent review see [Kuhl et al 2009] and the references within. In the proxemics application, however, a compression of distance estimation can only make the response more likely, and also the personal space only starts at 1.2m, which is outside of the boundary of distances that have typically been studied. Moreover, this issue has not hitherto been reported as problematic in proxemics studies, probably because in this application the exact distance is not critical. However, it would be a further topic of research in this area to examine the influence of exact distance perceptions.

It is important to note in general the benefit of using an online learning approach since the response of each participant could be different and might even change during the duration of an experiment. The RL method used here was able to learn the responses of each individual participant in order for them to achieve the predetermined goal.

7. CONCLUSIONS

A novel approach in the design of IVEs has been presented in this paper by applying RL methods in order to drive a participant towards achieving a task – even though the task was unknown to them. This methodology opens up new approaches for the design of more robust virtual environment applications, ones that are explicitly ‘participant centered’, where it is recognized that *the goal is to affect the behavior of the participants* rather than to generate visual or behavioral effects on an avatar that are in themselves more realistic for their own sake. Here we were able to exploit the results of previous work that had shown proxemics rules operate in IVEs. The task was not to make the avatar look more realistic but to exploit these rules to affect the behavior of the participant. This is quite a different approach to the design of IVEs than is normally the case. This idea of an intelligent interaction system where the goal of the user is the primary objective of the system is applicable to many areas of Human Computer Interaction (HCI). Our approach has been supported by the results of the experiment. Allowing the avatar to invade personal space resulted in faster completion time than when only social distances were allowed. By comparing with a control condition the experiment has demonstrated that RL was capable of sampling the response functions of the participants, providing qualitative and quantitative measures of the participant’s perception of proxemics interactions. Our further work will investigate this paradigm in more complex scenarios, involving manipulation of other behavioral as well as physiological responses.

ACKNOWLEDGMENTS

We thank Anna Bellido for helping with the experiments, Joan Llobera for the voice recording and Doron Friedman for his comments and advice on RL.

REFERENCES

ABBEEL, P. AND NG, A. Y. 2004. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the Twenty-First international Conference on Machine Learning* (Banff, Alberta, Canada, July 04 - 08, 2004). ICML '04, vol. 69. ACM, New York, NY, 1

BAAS, J.M., NUGENT, M., LISSEK, S., PINE, D.S., AND GRILLON, C. 2004, Fear conditioning in virtual reality contexts: A new tool for the study of anxiety. *Biological Psychiatry*, vol. 55, no. 11, pp. 1056-1060.

BAILENSON, J.N., BLASCOVICH, J., BEALL, A.C., AND LOOMIS, J.M. 2003, Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, vol. 29, no. 7, pp. 819-833.

BAILENSON, J.N., BLASCOVICH, J., BEALL, A.C., AND LOOMIS, J.M. 2001, Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 6, pp. 583-598.

BARFIELD, W., AND HENDRIX, C. 1995, The Effect of update rate on the sense of presence within virtual environments. *Virtual Reality: The Journal of the Virtual Reality Society*, vol. 1, no. 1, pp. 3-16, 1995.

BARFIELD, W., AND WEGHORST, S. 1993, The sense of presence within virtual environments: A conceptual framework. In *Human-Computer Interaction:Software and Hardware Interfaces*, G. Salvendy and M. Smith, Eds, Elsevier Publisher, pp. 699-704.

CARROZZINO, M., TECCHIA, F., BACINELLI, S., CAPPELLETTI, C., AND BERGAMASCO, M. 2005. Lowering the development time of multimodal interactive application: the real-life experience of the XVR project. In *Proceedings of the 2005 ACM SIGCHI international Conference on Advances in Computer Entertainment Technology* (Valencia, Spain, June 15 - 17, 2005). ACE '05, vol. 265. ACM, New York, NY, pp. 270-273.

COLLINS, S., RUINA, A., TEDRAKE, R., AND WISSE, M. 2005, Efficient bipedal robots based on passive-dynamic walkers, *Science,* vol. 307, no. 5712, pp. 1082.

DOYA, K. 2007, Reinforcement learning: Computational theory and biological mechanisms. *HFSP Journal,* vol. 1, no. 1, pp. 30-40, May 2007, 2007.

DRAPER, J.V., KABER, D. B, AND USHER, J. M. 1998, Telepresence. *Human Factors,* vol. 40, no. 3, pp. 354-375.

FREEMAN, D., PUGH, K., ANTLEY, A., SLATER, M., BEBBINGTON, P., GITTINS, M., DUNN, G., KUIPERS, E., FOWLER, D., GARETY, P. 2008, Virtual reality study of paranoid thinking in the general population. *The British Journal of Psychiatry*, vol. 192, no. 4, pp. 258-263.

FRIEDMAN, D., and GILLIES, M. 2005. Teaching virtual characters how to use body language. In *Lecture Notes in Computer Science*, T. PANAYIOTOPOULOS, J. GRATCH, R. AYLETT, D. BALLIN, P. OLIVIER, and T. RIST, Eds. Springer-Verlag, London, pp. 205-214.

FRIEDMAN, D., STEED, A., AND SLATER, M. 2007, Spatial social behavior in second life. In *Proceedings of the 7th international Conference on intelligent Virtual Agents* (Paris, France, September 17 - 19, 2007). C. PELACHAUD, J. MARTIN, E. ANDRÉ, G. CHOLLET, K. KARPOUZIS, and D. PELÉ, Eds. Lecture Notes In Artificial Intelligence, vol. 4722. Springer-Verlag, Berlin, Heidelberg, pp.252-263.

GARAU, M., FRIEDMAN, D., WIDENFELD, H., ANTLEY, A., BROGNI, A., AND SLATER, M. 2008, Temporal and spatial variations in presence: Qualitative analysis of interviews from an experiment on breaks in presence. *Presence: Teleoperators and Virtual Environments*, vol. 17, no. 3, pp. 293-309.

GILLIES, M. AND SPANLANG, B. 2010, Comparing and evaluating real time character engines for virtual environments. *Presence: Teleoperators and Virtual Environments*, vol. 19, no. 2, pp. 95-117.

GUYE-VUILLEME, A., CAPIN, T.K., PANDZIC, S., MAGNENAT THALMANN N., AND THALMANN, D. 1999, Nonverbal communication interface for collaborative virtual environments. *Virtual Reality*, vol. 4, no. 1, pp. 49-59, 1999.

HALL, E.T. 1973, *The Hidden Dimension*, Leonardo.

HARRIS, S.R., KEMMERLING, R.L., AND NORTH, M.M. 2002, Brief virtual reality therapy for public speaking anxiety. *Cyberpsychology & Behavior*, vol. 5, no. 6, pp. 543-550.

HELD, R.M., AND DURLACH, N.I. 1992, Telepresence. *Presence: Teleoperators and Virtual Environments,* vol. 1, no. 1, pp. 109-112.

HODGES, L.F., KOOPER, R., MEYER, T.C., ROTHBAUM, B.O., OPDYKE, D., DE GRAAFF, J.J., WILLIFORD, J.S., AND NORTH, M.M. 1995, Virtual Environments for treating the fear of heights. *Computer*, vol. 28, no. 7, pp. 27-34.

HOFFMAN, H.G., RICHARDS, T., CODA, B., RICHARDS, A., AND SHARAR, S.R. 2003, The illusion of presence in immersive virtual reality during an fMRI brain scan. *Cyberpsychology & Behavior*, vol. 6, no. 2, pp. 127-131.

HOFFMAN, H.H., RICHARDS, T. L., CODA, B., BILLS, A.R., BLOUGH, D., RICHARDS, A.L., AND SHARAR, S.R. 2004, Modulation of thermal pain-related brain activity with virtual reality: evidence from fMRI.. *Neuroreport*, vol. 15, no. 8, pp. 1245-8.

IKEMOTO, L., ARIKAN, O., and FORSYTH, D. 2005. Learning to move autonomously in a hostile world. *In ACM SIGGRAPH 2005 Sketches* (Los Angeles, California, July 31 - August 04, 2005). J. Buhler, Ed. SIGGRAPH '05. ACM, New York, NY, p. 46.

KRIJN, M., EMMELKAMP, P.M.G., OLAFSSON, R.P., AND BIEMOND, R. 2004, Virtual reality exposure therapy of anxiety disorders: A review. *Clinical Psychology Review*, vol. 24, no. 3, pp. 259-281.

KRIJN, M., EMMELKAMP, P.M.G., OLAFSSON, R.P., BOUWMAN, M., VAN GERWEN, L.J., SPINHOVEN, P., SCHUEMIE, M.J., AND VAN DER MAST, C.A.P.G. 2007, Fear of flying treatment methods: Virtual reality exposure vs. cognitive behavioral therapy. *Aviation Space and Environmental Medicine*, vol. 78, no. 2, pp. 121-128.

KUHL, S.A., THOMPSON, W.B., AND CREEM-REGEHR, S.H., 2009, HMD Calibration and Its Effects on Distance Judgements, *ACM Transactions on Applied Perception (TAP)*, vol. 6, no. 3, pp. 1- 20.

LEE, S. J. AND POPOVIĆ, Z. 2010. Learning behavior styles with inverse reinforcement learning. *ACM Transactions on. Graphics, vol.* 29, no. 4, Paper: 122.

LESSITER. J., FREEMAN, J., KEOGH E., AND DAVIDOFF J. 2001, A cross-media presence questionnaire: The ITC-Sense of presence inventory. *Presence: Teleoperators and Virtual Environments,* vol. 10, no. 3, pp. 282-297.

LLOBERA, J., RUFFINI, G., AND SLATER, M. 2010, Proxemics with multiple dynamic characters in an immersive virtual environment . *ACM Transactions on Applied Perception (TAP)*, vol. 8, no. 1 (in press).

LO, W., AND ZWICKER, M. 2008. Real-time planning for parameterized human motion. In *Proceedings of the 2008 ACM Siggraph/Eurographics Symposium on Computer Animation* (Dublin, Ireland, July 07 - 09, 2008). Symposium on Computer Animation. Eurographics Association, Aire-la-Ville, Switzerland, pp.29-38.

MCCANN, J., AND POLLARD, N. 2007. Responsive characters from motion fragments. *ACM Transactions on Graphics*, vol. 26, no. 3, Article 6 (July 2007).

MEEHAN, M., INSKO, B., WHITTON, M., AND BROOKS, F.P. 2002. Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics*, vol. 21, no. 3 (July 2002), pp. 645-652.

MEEHAN, M., RAZZAQUE, S., WHITTON, M. C., and BROOKS, F. P. 2003. Effect of latency on presence in stressful virtual environments. In *Proceedings of the IEEE Virtual Reality 2003* (March 22 - 26, 2003). VR. IEEE Computer Society, Washington, DC, pp. 141-148.

MINE, M.R., BROOKS, F.P., AND SEQUIN, C.H. 1997. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and interactive Techniques International Conference on Computer Graphics and Interactive Techniques* (SIGGRAPH '97). ACM Press/Addison-Wesley Publishing Co., New York, NY, pp. 19-26.

MORTENSEN, J., YU, I., KHANNA, P., TECCHIA, F., SPANLANG, B., MARINO, G., AND SLATER, M. 2008. Real-time global illumination for virtual reality applications. *IEEE Computer Graphics and Applications*, vol. 28, no. 6, pp. 56-64.

PAN, X., GILLIES, M., AND SLATER, M. 2008, Male bodily responses during an interaction with a virtual woman. In *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 89-96.

RAZZAQUE, S., SWAPP, D., SLATER, M., AND WHITTON, M.C., 2002, Redirected walking in place. In *Proceedings of the Workshop on Virtual Environments 2002* (Barcelona, Spain, May 30 - 31, 2002). W. STÜRZLINGER AND S. MÜLLER, Eds. ACM International Conference Proceeding Series, vol. 23. Eurographics Association, Aire-la-Ville, Switzerland, pp. 123-130.

REGENBRECHT, H.T., SCHUBERT, T.W., AND FRIEDMANN, F. 1998, Measuring the sense of presence and its relations to fear of heights in virtual environments. *International Journal of Human-Computer Interaction*, vol. 10, no. 3, pp. 233-249.

REY, B., ALCAÒIZ, M., TEMBL, J., AND PARKHUTIK, V., 2008, Brain activity and presence: a preliminary study in different immersive conditions using transcranial Doppler monitoring. In *Proceedings of Presence 2008,* A. SPAGNOLLI, AND L. GAMBERINI, Eds., pp. 209-218.

RIZZO, A., PAIR, J., MCNERNEY, P. J., EASTLUND, E., MANSON, B., GRATCH, J., HILL, R., ROY, M.,SWARTOUT, B., GRAAP, K., and WIEDERHOLD, B. 2005, From training to toy to treatment: Preliminary results from virtual reality therapy: Application for Iraq war military personnel with combat-related post-traumatic stress disorder. *Cyberpsychology & Behavior*, vol. 8, no. 4, pp. 352-353.

ROTHBAUM, B.O., AND HODGES, L. F. 1999, The use of virtual reality exposure in the treatment of anxiety disorders. *Behavior Modification*, vol. 23, no. 4, pp. 507-525.

ROTHBAUM, B.O., HODGES, L., SMITH, S., LEE, J.H., AND PRICE, L. 2000, A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology*, vol. 68, no. 6, pp. 1020-1026.

SANCHEZ-VIVES, M.V., AND SLATER, M. 2005, From presence to consciousness through virtual reality. *Nature Reviews Neuroscience,* vol. 6, no. 4, pp. 332-339.

SCHUBERT, T., FRIEDMANN, F., AND REGENBRECHT, H. 2001, The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 266-281.

SHERIDAN, T.B. 1992, Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments,* vol. 1, no. 1, pp. 120-126.

SHUM, H. P., KOMURA, T., AND YAMAZAKI, S. 2008. Simulating interactions of avatars in high dimensional state space. In *Proceedings of the 2008 Symposium on interactive 3D Graphics and Games* (Redwood City, California, February 15 - 17, 2008). I3D '08. ACM, New York, NY, pp. 131-138.

SLATER M., SPANLANG B., COROMINAS, D. 2010. Simulating Virtual Environments within Virtual Environments as the Basis for a Psychophysics of Presence. *ACM Transactions on Graphics,* vol. 29, no. 4, Article 92 (July 2010).

SLATER, M., AND WILBUR, S. 1997, A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments,* vol. 6, no. 6, pp. 603-616.

SLATER, M., KHANNA, P., MORTENSEN, J., AND YU, I. 2009. Visual realism enhances realistic response in an immersive virtual environment. *IEEE Computer Graphics and Applications,* vol.29, no.3, pp. 76-84.

SUTTON, R., AND BARTO, A. 1998, *Reinforcement Learning: An Introduction*: MIT Press.

TREUILLE, A., LEE, Y., AND POPOVIĆ, Z. 2007. Near-optimal character animation with continuous control. *ACM Transactions on Graphics*, vol. 26, no. 3, Article 7 (July 2007).

USOH, M. CATENA, E., ARMAN, S. AND SLATER, M. 2000. Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 5, pp. 497-503.

USOH, M., ARTHUR, K., WHITTON, M.C., BASTOS, R., STEED, A., SLATER, M., AND BROOKS, F.P. 1999. Walking > walking-in-place > flying, in virtual environments. In *Proceedings of the 26th Annual Conference on Computer Graphics and interactive Techniques International Conference on Computer Graphics and Interactive Techniques* (SIGGRAPH '99). ACM Press/Addison-Wesley Publishing Co., New York, NY, pp. 359-364.

VIGORITO, C.M 2007. Distributed path planning for mobile robots using a swarm of interacting reinforcement learners. In *Proceedings of the 6th international Joint Conference on Autonomous Agents and Multiagent Systems* (Honolulu, Hawaii, May 14 - 18, 2007). AAMAS '07. ACM, New York, NY, pp.1-8.

WALLACH, H.S., SAFIR, M. AND M. BAR-ZVI, M. 2009, Virtual reality cognitive behavior therapy for public speaking anxiety: A randomized clinical trial. *Behavior Modification* 33, No. 3, pp. 314-338.

WILCOX, L. M., ALLISON, R. S., ELFASSY, S., AND GRELIK, C. 2006. Personal space in virtual reality. *ACM Transactions on Applied Perception,* vol. 3, no. 4, pp.412-428.

WITMER, B.G., AND SINGER, M.J. 1998, Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, pp. 225-240.

WITMER, B.G., JEROME, C.J., AND SINGER M.J. 2005, The factor structure of the presence questionnaire. *Presence: Teleoperators and Virtual Environments,* vol. 14, no. 3, pp. 298-312.

WU, T., BUTKO, N. J., RUVULO, P., BARTLETT, M. S., AND MOVELLAN, J. R. 2009. Learning to make facial expressions. In *Proceedings of the 2009 IEEE 8th international Conference on Development and Learning* (June 05 - 07, 2009). DEVLRN. IEEE Computer Society, Washington, DC, pp. 1-6.

ZIMMONS, P., AND PANTER, A. 2003, The influence of rendering quality on presence and task performance in a virtual environment. In *Proceedings of the IEEE Virtual Reality 2003* (March 22 - 26, 2003). VR. IEEE Computer Society, Washington, DC, pp. 293-294.