Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality

MEL SLATER, MARTIN USOH, and ANTHONY STEED
University of London

This article presents an interactive technique for moving through an immersive virtual environment (or "virtual reality"). The technique is suitable for applications where locomotion is restricted to ground level. The technique is derived from the idea that presence in virtual environments may be enhanced the stronger the match between proprioceptive information from human body movements and sensory feedback from the computer-generated displays. The technique is an attempt to simulate body movements associated with walking. The participant "walks in place" to move through the virtual environment across distances greater than the physical limitations imposed by the electromagnetic tracking devices. A neural network is used to analyze the stream of coordinates from the head-mounted display, to determine whether or not the participant is walking on the spot. Whenever it determines the walking behavior, the participant is moved through virtual space in the direction of his or her gaze. We discuss two experimental studies to assess the impact on presence of this method in comparison to the usual hand-pointing method of navigation in virtual reality. The studies suggest that subjective rating of presence is enhanced by the walking method provided that participants associate subjectively with the virtual body provided in the environment. An application of the technique to climbing steps and ladders is also presented.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*artificial realities*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces; I.3.4 [**Computer Graphics**]: Graphics Utilities—*virtual device interfaces*; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*virtual reality*

General Terms: Experimentation, Human Factors

Additional Key Words and Phrases: Immersion, locomotion, navigation, neural networks, presence, virtual environments, virtual reality

This is a substantially revised and expanded version of Slater et al. [1994a]. This work is funded by the UK Engineering and Physical Sciences Research Council (ESPRC) and the Department of Trade and Industry, through grant CTA/2 of the London Parallel Applications Centre. Anthony Steed is supported by an EPSRC research studentship.
Authors' address: Department of Computer Science and London Parallel Applications Centre, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, U.K.; email: {mel; bigfoot; steed}@dcs.qmw.ac.uk.

1. INTRODUCTION

The ability to get from place to place is a fundamental requirement for action in both real and virtual environments. This requirement epitomizes what is very powerful yet what also may be flawed in virtual reality (VR) systems. These systems offer the possibility of perceptually immersing individuals into computer-generated environments, and yet the typical means for the most-basic form of interaction—locomotion—do not at all match the physical actions of walking in reality. Generally, the powerful illusion of immersion may be lost through naive interaction metaphors borrowed from nonimmersive forms of human-computer interaction.

This article describes an interactive technique for locomotion in an immersive virtual environment (or "virtual reality"). The technique is suitable in applications where the participants are constrained to ground level, for example, while exploring a virtual building, as in an architectural walk-through. The novelty of the technique is that participants carry out whole-body movements in a simulation of walking, without the necessity of hardware additional to the electromagnetic tracking devices on the head-mounted display (HMD) and glove (or 3D mouse). In brief, participants "walk in place" to move across virtual distances that are greater than the physical space determined by the range of the electromagnetic trackers. Pattern analysis of head movements as generated by the HMD predicts whether participants are walking in place or doing anything else at all. Whenever it is determined that they are walking in place, they are moved forward in the direction of gaze, so that the corresponding flow in the optical array gives the illusion of motion. Such illusory self-motion is usually called vection. Since the pattern analyzer (ideally) only detects head movements characteristic of walking in place, participants are able to take real physical steps, while remaining within an effective tracker range, without causing vection surplus to their actual movements.

In an earlier report [Slater et al. 1993] we presented the technique, called the Virtual Treadmill,[1] in the context of (at that time) a partially complete human factors evaluation. In this article we discuss the technique in the context of a model of presence in immersive virtual environments. We also present the implementation details and results of two empirical studies with users. The utility of this idea for climbing or descending steps and ladders is also discussed.

2. VIRTUAL ENVIRONMENTS

2.1 The Proprioceptive Sensory Data Loop

A VR system requires that the normal proprioceptive information we use unconsciously to form a mental model of the body be overlaid with sensory data that is supplied by computer-generated displays. Proprioception was

[1] The London Parallel Applications Centre had a holding patent covering the U.K. and other countries to protect aspects of this technology.

defined by Sacks [1985] as "that continuous but unconscious sensory flow from the movable parts of our body (muscles, tendons, joints), by which their position and tone and motion [are] continually monitored and adjusted, but in a way which is hidden from us because it is automatic and unconscious." Proprioception allows us to form a mental model that describes the dynamic spatial and relational disposition of our body and its parts. We know where our left foot is (without having to look) by tapping into this body model. We can clap our two hands together (with closed eyes) similarly by relying on this unconscious mental model formed from the proprioceptive data flow.

Tracking devices placed on the physical human body are required in order to map real body movements onto corresponding movements of the participant's self-representation in the virtual world. We call this self-representation a virtual body (VB). A fundamental requirement for an effective virtual reality is, therefore, that there is a consistency between proprioceptive information and sensory feedback, and in particular, between the mental body model and the VB.

Gibson's [1986] notion of the ambient optical array may be employed to elaborate these ideas. This is conceived as an arrangement consisting of a nested hierarchy of visual solid angles all with the same apex and completely surrounding the apex. The apex corresponds to a position in the environment, which may be occupied by an individual. Such an individual is not considered as a disembodied observer, taking up an abstract point in space, but as a live animal that moves continually through an all-surrounding environment, standing and moving on feet and with a head, eyes, ears, nose, mouth. This is not the abstract space of the mathematician.

Gibson argued that when an individual is immersed in an environment, perception of the self is inseparable from perception of the environment. When describing the occupation of a position in the ambient optical array by an individual he said that, "When the position becomes occupied, something very interesting happens to the ambient array: it contains information about the body of the observer" [Gibson 1986, p. 66]. Regarding the relationship between sensory information and self-perception he wrote: "The optical information to specify the self, including the head, body, arms and hands, *accompanies* the optical information to specify the environment. The two sources of information coexist" [Gibson 1986, p. 116].

This relationship between proprioceptive information and sensory data requires consistency, predictability, and completeness in order to function properly. For example, when proprioceptive information arises because we have moved a leg in such a way that it comes into contact with another object, the sensory data must correctly inform us, in all modalities, that this is indeed occurring. We see our leg move; we hear the "woosh" as it glides through the air; we feel it touch the object (and feel any expected level of pain); we hear the sound caused by our leg hitting the object; and we see the object itself react in accordance with our expectations. This loop is the crucial component of a convincing reality: the "reality" is virtual when the sensory data is computer generated.

2.2 Immersion

We call a computer system that supports such experience an "immersive virtual environment" (IVE). It is immersive since it immerses a representation of the person's body (the VB) in the computer-generated environment. It is a virtual environment in the sense defined by Ellis [1991]: consisting of content (objects and actors), geometry and dynamics, with an egocentric frame of reference, including perception of objects in depth, and giving rise to the normal ocular, auditory, vestibular, and other sensory cues and consequences. Whether or not a system can be classified as immersive depends crucially on the hardware, software, and peripherals (displays and body sensors) of that system. We use "immersion" as a description of a technology, rather than as a psychological characterization of what the system supplies to the human participant.

Immersion includes the extent to which the computer displays are extensive, surrounding, inclusive, vivid, and matching. The displays are more extensive the more sensory systems that they accommodate. They are surrounding to the extent that information can arrive at the person's sense organs from any (virtual) direction and the extent to which the individual can turn toward any direction and yet remain in the environment. They are inclusive to the extent that all external sensory data (from physical reality) are shut out. Their vividness is a function of the variety and richness of the sensory information they can generate [Steuer 1992].

In the context of visual displays, for example, color displays are more vivid than monochrome; high resolution is more vivid than low resolution; and displays depicting dynamically changing shadows are more vivid than those that do not. Vividness is concerned with the richness, information content, resolution, and quality of the displays. Finally, as we have argued above, immersion requires that there is a match between the participant's proprioceptive feedback about body movements and the information generated on the displays. The greater the degree of body mapping, the greater the extent to which the movements of the body can be accurately reproduced, and therefore the greater the potential match between proprioception and sensory data.

2.3 Presence

An IVE may lead to a sense of presence for a participant taking part in such an experience. Presence is the psychological sense of "being there" in the environment: it is an emergent property based on the immersive base given by the technology. However, any particular immersive system does not necessarily always lead to presence for all people: the factors that determine presence, given immersion, are an important area of study [Barfield and Weghorst 1993; Heeter 1993; Held and Durlach 1992; Loomis 1992; Sheridan 1992]. We concur with Steuer [1992] that presence is the central issue for virtual reality.

Our view concerning the relationship between immersion and presence is shown in Figure 1. The x-axis is the extent of the match between the

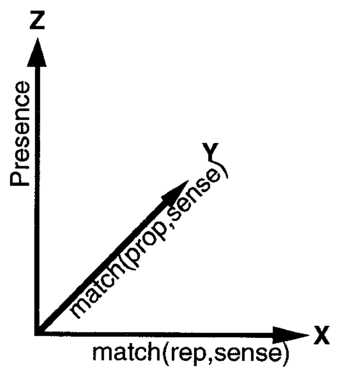

Fig. 1. Presence = (f(match(prop, sense), match(rep, sense; prop = proprioception; rep = internal representation; sense = sensory data.

displayed sensory data and the internal representation systems and subjective-world models typically employed by the participant. Although immersion is greater the greater the richness of the displays, as discussed above, we must also take into account the extent to which the information displayed allows the particular individuals to construct their own internal mental models of reality. For example, a vivid visual display system might afford some individuals a sense of "reality" but be unsuited for others in the absence of sound. Even though an excellent virtual body might exist in the VE, some individuals might reject it because it contradicts their personal self-model. We have explored the relationship between presence and this match between subjectivity and displayed data in earlier experiments [Slater et al. 1994b].

The y-axis is the extent of the match between proprioception and sensory data, as explained above. The changes to the display must be consistent with and match through time, without lag, changes caused by the individual's motility and locomotion—whether of individual limbs or the whole body, relative to the ground.

Our general hypothesis is that presence is a function of these two "matches"—that it positively increases with each of them. Note that the axes are orthogonal—a system might provide a superb degree of visual, auditory, and tactile display immersion, so that most individuals have sufficient data to construct their internal representations successfully but fail to provide a sufficient degree of match between the person's actions and the displayed results, thus breaking the link between sensory data and proprioception.

A further point about this hypothesis is that we would expect it to operate at many levels. At a very basic level, the displays should result in suitable parasympathetic responses in, for example, the ocular and vestibular systems. When an individual focuses visually on a near object the visual displays should likewise respond appropriately and immediately and again change immediately when the focus moves to a far object. Eye tracking should be enabled. At a much higher level, when a person moves, the shadow structure of the virtual body on nearby surfaces should change accordingly [Slater et al. 1995]. At a similarly high level, the interactive metaphors employed in the system should match the sensory data and proprioception. This brings us back to walking: if the optical flow indicates forward movement at ground level, then the proprioceptive information should correspond to this.

A specific hypothesis of this article is, therefore, that the degree of presence depends on the match between proprioceptive and sensory data. The greater the match, the greater the extent to which the participant can associate with the VB as a representation of self. Since the VB is perceived as being in the VE, this should give rise to a belief (or suspension of disbelief) in the presence of self in that environment. In particular, the closer that the action required for forward locomotion corresponds to really "walking" the greater the sense of presence.

3. LOCOMOTION

3.1 Other Methods

There is a tendency in VR research to use hand gestures to do everything, from grasping objects (a natural application), to scaling the world, and to navigation [Robinett and Holloway 1992; Vaananen and Bohm 1993]. This approach overloads greatly the hand gesture idea—the user has to learn a complete vocabulary of gestures in order to be effective in the virtual world. Small differences between gestures can be confusing, and in any case there is no guarantee of a correspondence among the gesture, the action to be performed, and the displayed outcome.

The standard VR metaphor for locomotion is a hand gesture, with the direction of navigation determined either by gaze or by the direction of pointing. The VPL method for navigation, as demonstrated at SIGGRAPH 90, for example, used the DataGlove to recognize a pointing hand gesture where the direction of movement was controlled by the pointing direction.

Song and Norman [1993] review a number of techniques, distinguishing between navigation based on eyepoint movement and that of object movement. Here we are interested in "naturalistic" navigation, appropriate for a walkthrough application, so we rule out navigation via manipulation of a root object in a scene hierarchy [Ware and Osborne 1990].

Fairchild et al. [1993] introduced a leaning metaphor for navigation, where the participant moves in the direction of body lean. The technique involves extending the apparent movement in virtual space in comparison with the real movement. In fact, this is an "ice skating" metaphor, which may not be appropriate, for example, to architects taking their clients on a virtual tour.

In the context of architectural walkthrough we require participants to experience a sense of moving through the virtual building interior in a manner that maximizes sensory data and proprioception. Brooks [1992] used a steerable treadmill for this purpose. However, the use of any such device as a treadmill, footpads, roller skates [Iwata and Matsuda 1992], or even a large area mat with sensing devices imposes constraints on the movements of participants. Moreover, there will always be an application where the virtual space to be covered is much larger than the physical space available—one of the major advantages of VR systems.

3.2 Walking

We require that participants be able to take advantage of the range available with an electromagnetic tracker, such as a Polhemus device, in order to cover small distances by moving their bodies and really walking. Beyond the range of the sensor though, they should still carry out movements reminiscent of walking, while staying within range. If this is possible, then proprioceptive information (associated with "walking") matches sensory data (flow in the optical array consistent with motion) to a much greater extent than motion based on hand gesture interfaces.

The new method for locomotion at ground level allows participants to move around in the space defined by the electromagnetic tracker as usual. To cover a virtual distance that is larger than the physical space afforded by the tracker, the participant *walks in place*. While carrying out this activity he or she will move forward in virtual space in the direction of gaze. It is almost walking, but no forward movement takes place in physical space. (We never have to explain to users that direction is determined by gaze; they just pick this up automatically.)

A major advantage of this technique is that the hand is not used at all for navigation. The hand may be entirely reserved for the purposes for which it is used in everyday reality, that is, the manipulation of objects and activation of controls.

3.3 Implementation

The implementation of this technique is very straightforward. We have used a feed-forward neural net [Hertz et al. 1991] to construct a pattern recognizer that detects whether participants are "walking in place" or doing something else. The HMD tracker delivers a stream of position values (x_i, y_i, z_i) from which we compute differences first $(\Delta x_i, \Delta y_i, \Delta z_i)$ $(i = 1, 2, \dots, n)$. We choose sequences of n data points, and the corresponding delta-coordinates are inputs to the bottom layer of the net so that there are $3n$ units at the bottom layer. There are two intermediate layers of m_1 and m_2 hidden units $(m_1 \leq m_2)$, and the top layer consists of a single unit, which outputs either 1 corresponding to "walking in place" or 0 for anything else.

We obtain training data from a person, which are used to compute the weights for the net using back-propagation. During the training phase the subject walks on the spot while immersed in an almost-featureless environment. He or she is asked to carry out a number of different activities, such as bending down, moving around, turning the head, and mixtures of these, interspersed with periods of walking in place. This continues for five to ten minutes. An operator records binary data into the computer, corresponding to whether or not the subject is walking in place. The data, together with the corresponding sequences of delta-coordinates, are then used to train the neural net. The resulting network equations are then implemented on the VR machine as part of the code of the process that deals with detection of events indicating forward movement.

After experimenting with a number of alternatives, we have found that a value of $n = 20$, $m_1 = 5$, and $m_2 = 10$ gives good results. We have never obtained 100% accuracy from any network, and this would not be expected. There are two possible kinds of error, equivalent to Type I and Type II errors of statistical testing, where the null hypothesis is taken as "the person is not walking on the spot." The net may predict that the person is walking when they are not (Type I error) or may predict that the person is not walking when they are (Type II error). The Type I error is the one that causes the most confusion for people and is also the one that is most difficult to rectify—in the sense that once they have been involuntarily moved from where they want to be, it is almost impossible to "undo" this. Hence our efforts have concentrated on reducing this kind of error. We do not use the output of the net directly but only change from not moving to moving if a sequence of p 1s is observed and from moving to not moving if a sequence of q 0s is observed ($q < p$). In practice we have used $p = 4$ and $q = 2$.

3.4 Results with the Neural Network

Among 16 people who took part in an evaluation, the mean success rate for their networks, that is, the proportion of time that the net correctly predicted their activity, was 91%. The minimum and maximum rates were 85 and 96%. The mean Type I error was 10%, with a minimum of 6% and a maximum of 15%. The corresponding figures for Type II error are 6, 2, and 16%. Given the simplicity of the pattern recognizer we were surprised at how well the system performed in practice. We also have an arbitrarily designated "standard" network that most casual visitors to the laboratory are able to use without the necessity of a net being trained for their personal style of walking.

The Polhemus Isotrak tracking device actually returns data to the application at a rate of about 30Hz. The overall error is largely caused by the actual output lagging behind the real output by typically five samples, at the end of each sequence of 1s or 0s. It is likely that, with further investigation of the neural net training method or the employment of alternative pattern recognition techniques, results will improve.

4. EXPERIMENTAL EVALUATION

In this section we consider the results of two studies: a pilot experiment and a main experiment—each to assess the influences of the walking metaphor on ease of navigation and presence. In each case there were a number of subjects, divided equally into two groups. The first study is partially reported in Slater et al. [1993], and the second is reported here. The control groups (the "pointers") navigated the environment using a 3D mouse, initiating movement by pressing a button, with direction of movement controlled by pointing. The experimental groups (the "walkers") used the walking technique. In each case the mouse was also used for grasping objects. The task was to pick up an object, take it into a room, and place it on a particular chair. The chair was placed in such a way that the subjects had to cross a chasm over another room about 20 feet below in order to reach it.

The experiments were implemented on a DIVISION ProVision200 system. The ProVision system includes a DIVISION 3D mouse and a Virtual Research Flight Helmet as the head-mounted display. Polhemus sensors are used for position tracking of the head and the mouse. Scene rendering is performed using an Intel i860 microprocessor (one per eye) to create an RGB RS-170 video signal which is fed to an internal NTSC video encoder and then to the displays of the Flight Helmet. These displays (for the left and right eye) are color LCDs with a 360×240 resolution, and the HMD provides a horizontal field of view of about 75 degrees. The frame update rate achieved during the experiments was about 15 frames per second.

All subjects saw a virtual body as self-representation. They would see a representation of their right hand, and their thumb and first-finger activation of the 3D pointer buttons would be reflected in movements of their corresponding virtual finger and thumb. The hand was attached to an arm, which could be bent and twisted in response to similar movements of the real arm and wrist. The arm was connected to an entire, but simple, block-like body representation, complete with legs and left arm. Forward movement was accompanied by walking motions of the virtual legs. If the subjects turned their real head around by more than 60 degrees, then the virtual body would be reoriented accordingly. So, for example, if they turned their real body around and then looked down at their virtual feet, their orientation would line up with their real body. However, turning only the head around by more than 60 degrees and looking down (an infrequent occurrence) would result in the real body being out of alignment with the virtual body.

4.1 Navigation

With respect to the ease of navigating the environment, subjects in both experiments marginally preferred to use the pointing technique. This result was not surprising: as Brooks et al. [1992] noted, with the real treadmill more energy certainly is required to use the whole body in a walking activity, compared to pressing a mouse button or making a hand gesture (or driving a car, with respect to the similar comparison in everyday reality). Moreover, the networks did not work with 100% accuracy, in contrast to the accuracy of the pointing method.

In the postexperiment questionnaire three questions were asked of all subjects, covering three aspects of navigation: *general movement*—that is, how simple or complicated it was to move around; *placement*—that is, the ease of getting from one place to another; and *how "natural" the movement was*. The questions are shown in Table I, with results given for both experiments (the results should not be combined since there were some differences between the two experiments). The differences between the answers given by the "pointers" and "walkers" are not statistically significant. However, Figure 2 shows scattergrams (for those in the experimental group) of the answers to the three questions against the Type I error for the pilot study only (such data were not available from the main study). The sample size involved is too small to carry out meaningful significance tests, but the graphs indicate that

Table I. Questions Relatıng to Ease of Navigation

General Movement	**Getting from Place to Place**	**Natural / Unnatural**
Did you find it relatively "simple" or relatively "complicated" to move through the computer-generated world?	How difficult or straight-forward was it for you to get from place to place?	The act of moving from place to place in the computer-generated world can seem to be relatively "natural" or relatively "unnatural." Please rate your experience of this.
To move through the world was...	*To get from place to place was...*	*The act of moving from place to place seemed to me to be performed...*
1. Very Complicated	1. Very Difficult	1. Very Unnatural
...	...	...
7. Very Simple	7. Very Straightforward	7. Very Natural
PILOT STUDY		
Mean Response	**Mean Response**	**Mean Response**
Control Group: 5.0, $n = 6$ Exper. Group: 5.1, $n = 8$	Control Group: 4.9, $n = 6$ Exper. Group: 5.5, $n = 8$	Control Group: 3.4, $n = 6$ Exper. Group: 3.9, $n = 8$
MAIN STUDY		
Control Group: 5.5, $n = 8$ Exper. Group: 4.9, $n = 8$	Control Group: 5.7, $n = 6$ Exper. Group: 4.7, $n = 8$	Control Group: 4.2, $n = 6$ Exper. Group: 4.2, $n = 8$

a decrease in Type I error generally leads to an improvement in ease of navigation. This suggests that a better pattern recognition technique could result in a superior performance for this method of navigation, compared to the pointing method. In other words, it is worthwhile improving the pattern recognition technique, because decrease in error is likely to result in a substantial improvement in subjective evaluation. (With the pointing technique there is no similar improvement that can be made.)

4.2 Presence

It is the sense of presence with which we are mainly concerned. Here we discuss the results of the main experiment that compared the two different techniques for navigation with respect to the effect on reported sense of presence.

There were 16 subjects, divided into two groups of eight. These were selected by asking for volunteers on the Queen Mary and Westfield College (QMW) campus, excluding people who had experienced our VR system before or who knew of the purposes of our research. The control groups (the "pointers") moved through the environment using the DIVISION 3D mouse, by pressing a button, with direction of movement controlled by pointing. The experimental groups (the "walkers") used the walking technique. All subjects

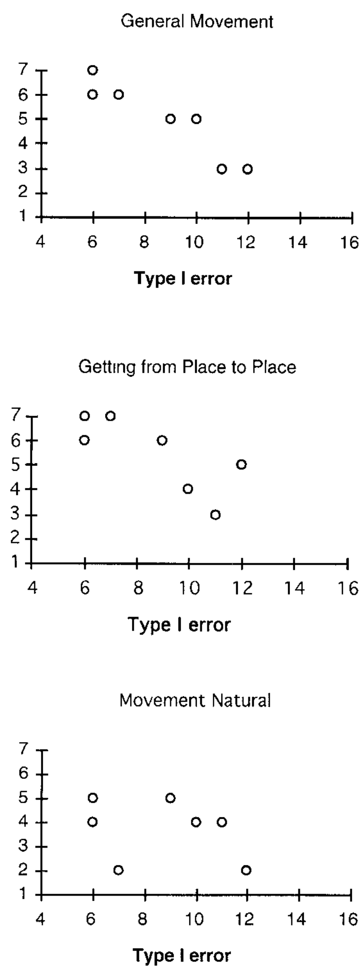

Fig. 2. Evaluation of navigation by Type I error.

used the same ("standard") network based on the walking-in-place behavior of one individual. Both walkers and pointers used the mouse for grasping objects. Intersecting the virtual hand with an object and pulling the first-finger (trigger) button resulted in the object being attached to the hand. The object would fall when the trigger button was released.

The task in the experiment was to pick up an object located in a corridor, take it into a room, and place it on a particular chair. The chair was placed in such a way that the subjects had to cross a chasm over another room about 20 feet below in order to reach it. The subjects could get to the chair either by going out of their way to walk around a wide ledge around the edges of the room or by moving directly across the chasm. This was a simple virtual version of the famous visual cliff experiment [Gibson and Walk 1960].

Subjective presence was assessed in three ways: the sense of "being there" in the VE, the extent to which the virtual world seemed more like the presenting reality than the real world, and the sense of visiting somewhere rather than seeing images of something. Each was rated by subjects on an ordinal seven-point scale, where 7 was the highest score, using a questionnaire given immediately after the experiment. These three scores were combined into one by counting the total number of six or seven responses from the three questions. Hence, the result was a value between 0 and 3.

Other questions relevant to the analysis concerned the degree of nausea experienced in the VR and the extent of association with the VB: "To what extent did you associate with the computer-generated limbs and body as being 'your body' while in the virtual reality?" They were also asked to rate the degree of vertigo, if any, induced by the virtual precipice and to compare their reaction to this in relation to how they would have reacted to a similar situation in real life: "To what extent was your reaction when looking down over the drop in the virtual reality the same as it would have been in a similar situation in real life?"

All subjects were watched by an observer, who, in particular, recorded whether or not they moved to the chair by walking around the ledge at the side of the room or by walking directly across the precipice. In the event, only four subjects out of the 16 (two from each group) walked across the precipice.

The main conclusion from the statistical analysis was that for the "walkers" the greater their association with the VB the higher the presence score, whereas for the "pointers" there was no correlation between VB association and the presence score. In other words, participants who identified strongly with the virtual body had a greater degree of reported presence if they were in the "walking" group than if they were in the "pointing" group. Association with the VB is important. This certainly belongs to the x-axis of Figure 1: indicating that it is not simply a question of whether a VB is provided by the system and how well it functions but also the individual's personal evaluation of this VB, the degree of "match" to his or her internal world models. It also belongs to the y-axis, as discussed in Section 7.

There were two other statistically significant factors: first, path taken to the chair. A path directly over the precipice was associated with lower presence. This is as would be expected and is useful in corroborating the veracity of the presence score. Second was the degree of nausea. A higher level of reported nausea was associated with a higher degree of presence. This same result has been found in each of our studies.

We speculate that the vection in VR is a cause of both simulator sickness and an influence on presence [McCauley and Sharkey 1993]. Finding nausea and presence to be associated would, therefore, not be surprising. There is the further point that presence is concerned with the effect of the environment on the individual. An increased sense of presence is likely to be correlated with the human brain paying more attention to the detailed operation of the environment and therefore to the discrepancy between the visual and vestibular systems. However, this may be a temporary phenomenon that will be overcome with greater exposure. This is speculation and would need to be examined by an independent study.

5. STATISTICAL ANALYSIS FOR PRESENCE

The dependent variable (p) was taken as the number of six or seven answers to the three questions as stated above. The independent variable was the group (experimental or control). The explanatory variables were VB (degree of association with the Virtual Body), S the reported nausea, and P for path (= 1 for a path around the sides of the room and 2 for a direct path across the precipice).

This situation may be treated by logistic regression [Cox 1970], where the dependent variable is binomially distributed, with expected value related by the logistic function to a linear predictor.

Let the independent and explanatory variables be denoted by $x_1, x_2, \ldots, x_k$. Then the linear predictor is an expression of the form:

$$\eta_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \; (i = 1, 2, \ldots, N) \tag{1}$$

where $N(= 16)$ is the number of observations. The logistic regression model links the expected value $E(p_i)$ to the linear predictor as:

$$E(p_i) = \frac{n}{1 + \exp(-\eta_i)} \tag{2}$$

where $n(= 3)$ is the number of binomial trials per observation.

Maximum-likelihood estimation is used to obtain estimates of the β coefficients. The deviance (minus twice the log-likelihood ratio of two models) may be used as a goodness-of-fit significance test, comparing the null model ($\beta_j = 0$, $j = 1, \ldots, k$) with any given model. The change in deviance for adding or deleting groups of variables may also be used to test for their significance. The (change in) deviance has an approximate χ^2 distribution with degrees of freedom dependent on the number of parameters (added or deleted).

Table II shows the results. The overall model is significant. For a good fit, the overall deviance should be small, so that a value of less than the tabulated value is significant. No term can be deleted from the model without increasing the deviance significantly (at the 5% level).

The analysis relies on the assumption that the dependent variable is binomially distributed. This assumption is made as a heuristic but cannot be justified in an obvious way. The presence-related questions were each separated by at least three others in the questionnaire, and for any respondent, not knowing the purposes of the study and not aware of the concept of presence, it would be reasonable to assume that their answers did not directly influence one another and therefore that the "trials" were independent.

An alternative analysis was carried out, where the three presence scores were combined into a single scale using principal-components analysis [Kendall 1975]. The first principal component is the linear combination of the original variables that maximizes the total variance. The second is orthogonal

Table II. Logistic Regression Equations

Group	Model	When $P = 2$ (path directly over precipice)
Walkers	$\hat{\eta} = -16.9 + 2.6^{*}\mathrm{VB} + 1.3^{*}\mathrm{S}$	-2.7
Pointers	$\hat{\eta} = -3.1 + \mathit{0.1}^{*}\mathrm{VB} + 1.3^{*}\mathrm{S}$	-2.7

Nonsignificant coefficients are shown in *italics*; $\hat{\eta}$ = fitted values for the presence scale; VB = VB association; S = nausea; P = path.

Deletion of Model Term	Change in Deviance	Change in d.f.	χ^2 at 5% level
S	6.624	1	3.841
P	3.867	1	3.841
Group. VB	10.922	2	5.991

Overall Deviance = 11.424; d.f. = 10; χ^2 at 5% on 10 d.f. = 18.307.

Table III. Regression Equations

Group	Model	When C = 2 "same as real life"
Walkers	$\hat{y} = -4.5 + 1.7^{*}\mathrm{VB} + 1.2^{*}\mathrm{S}$	+2.5
Pointers	$\hat{y} = 3.4 + \mathit{0.3}^{*}\mathrm{VB} + 1.2^{*}\mathrm{S}$	+2.5

Nonsignificant coefficients are shown in *italics*; $\hat{y}$ = fitted values for the presence scale based on principal components (coefficients are given to 1 d.p.); VB = VB association; S = nausea; C = vertigo comparison.

to the first and maximizes the total residual variance. The first two principal components accounted for 96% of the total variation in the original three variables (the first for 67% and the second for 29%). The single presence score was taken as the norm of the vector given by the first two principal components.

A regression analysis using this new presence score resulted in a model qualitatively similar to that described above. Here though, instead of P (path) being significant, the variable representing the comparison between vertigo experienced in the virtual world with what might have been experienced in the real world was significant instead. A higher degree of presence was associated with the comparison resulting in a "same as real life." The overall regression was significant at 5% with a multiple squared correlation coefficient of 0.81. This is summarized in Table III.

6. STEPS AND LADDERS

6.1 Walking on Steps and Ladders

In the previous sections we have made a case, together with supporting experimental evidence, that the walking-in-place technique tends to increase subjective presence, in comparison with the pointing technique based on a simple hand gesture, provided that there is an association with the VB.

The same idea can be applied to the problem of navigating steps and ladders. One alternative is to use the familiar pointing technique and to "fly." While in some applications there may be a place for such magical activity, the very fact that mundane objects such as steps and ladders are in the environment would indicate that a more-mundane method of locomotion be employed. The walking-in-place technique carries over in a straightforward manner to this problem.

When the collision detection process in the virtual reality system detects a collision with the bottom step of a staircase, continued walking will move the participant up the steps. Walking down the steps is achieved by turning around and continuing to walk. If at any moment the participant's virtual legs move off the steps (should this be possible in the application), then they would "fall" to the ground immediately below. Since walking backward down steps is something usually avoided, we do not provide any special means for doing this. However, it would be easy to support backward walking and walking backward down steps by taking into account the position of the hand in relation to body line: a hand behind the body would result in backward walking.

Ladders are slightly different; once the person has ascended part of the ladder, they might decide to descend at any moment. In the case of steps, the participant would naturally turn around to descend. Obviously this does not make sense for ladders. Also, when climbing ladders it is usual for the hands to be used. Therefore, in order to indicate ascent or descent of the ladder, hand position is taken into account. While carrying out the walking-in-place behavior on a ladder, if the hand is above the head then the participant will ascend the ladder and descend when below the head. Once again it is a whole-body gesture, rather than simply the use of the hand, that is required in order to achieve the required result in an intuitive manner. If at any time the virtual legs come off the rungs of the ladder, then the climber will "fall" to the ground below.

6.2 Evaluation for Usability

We have thus far only carried out a simple study to test for usability. A scenario was constructed consisting of steps leading up to the second story of a house. The steps led in through a doorway, which entered into a room consisting of a few everyday items such as a couch, television, and so on. There was a window and a ladder down to the ground outside propped up against the wall just below the window. There was a bucket on the ground outside, at the foot of the ladder. Examples are shown in the Figures 3–6.

The task was to walk up the steps, enter the room, climb onto the ladder and down to the ground, pick up the bucket, take it back up into the room, down the stairs, and back outside. The designer of this scene was taken as the "expert"—and completed the scenario in three minutes, including one fall from the ladder. Five other people, all of whom had used the VR system before, were invited to try out the scenario. One person also completed the task in three minutes, without any falls. Another took four minutes, also

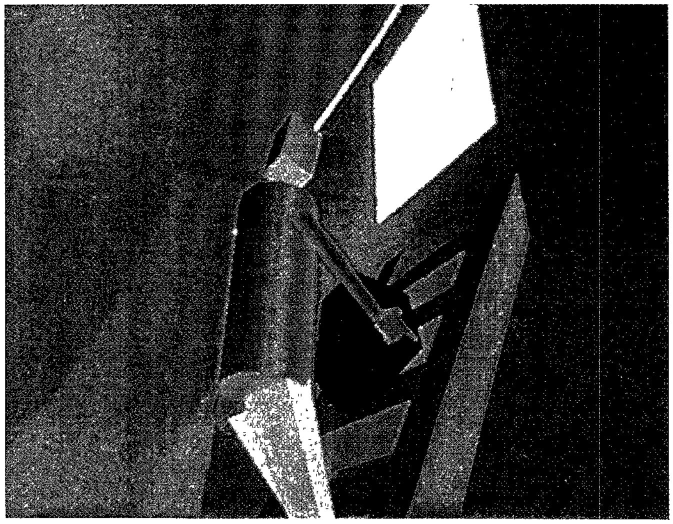

Fig. 3. An exocentric view of a participant ascending the ladder.

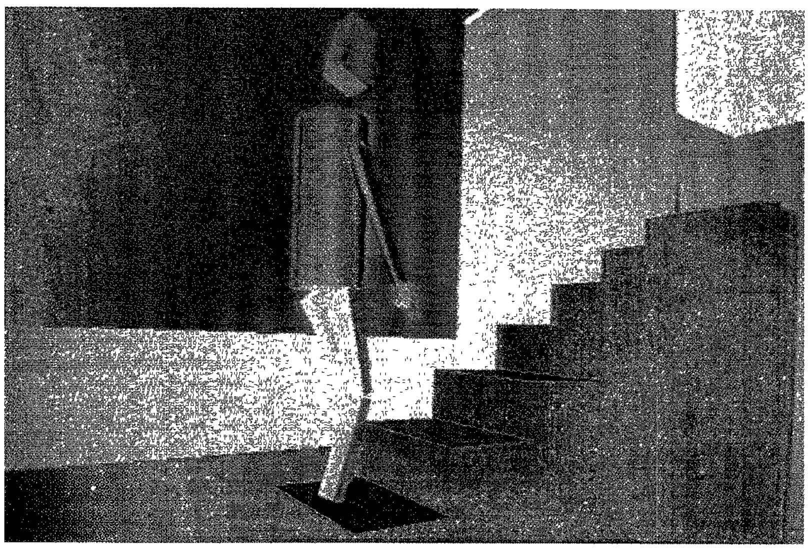

Fig. 4. An exocentric view of a participant ascending the steps.

without any falls. The third required five minutes, with two falls from the ladder. The remaining two each took eight minutes, with one and two falls from the ladder, respectively. The results of this simple experiment were encouraging enough for us to consider devising specific pattern recognizers for these types of activities.

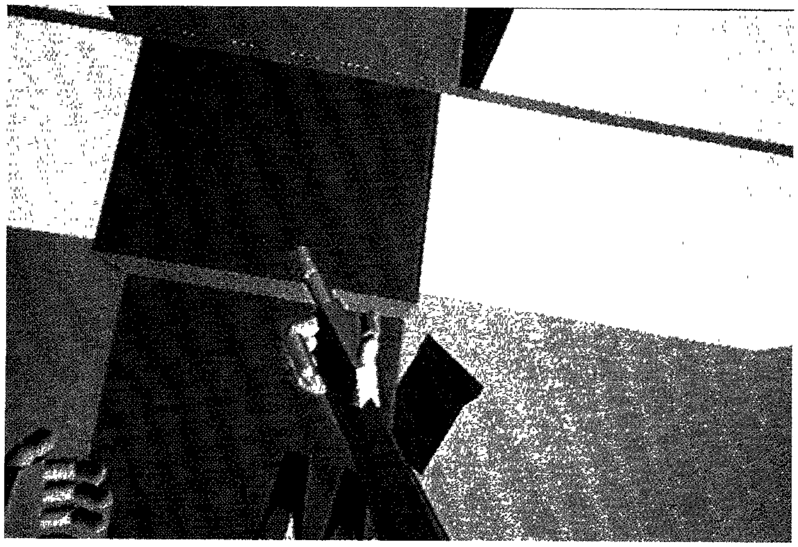

Fig. 5. An egocentric view looking downward at the virtual body while on the steps.

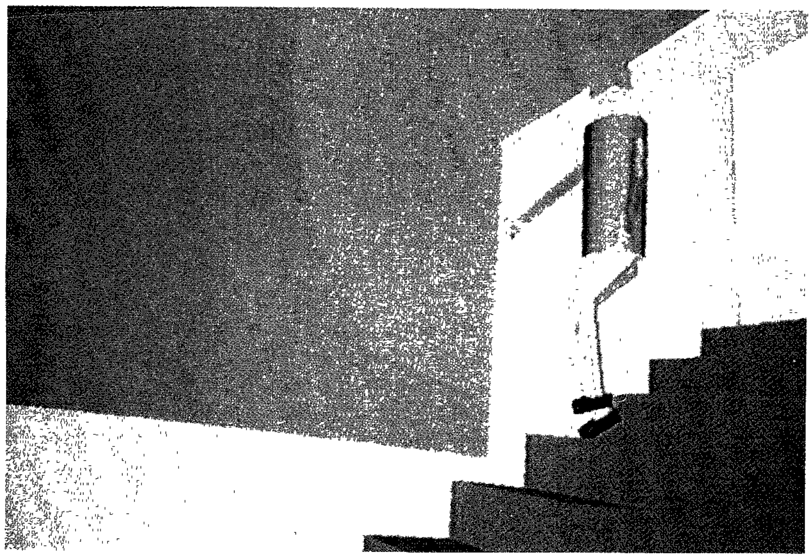

Fig. 6. An exocentric view of a participant descending the steps.

7. CONCLUSIONS

The rudimentary model for presence in virtual environments, illustrated in Figure 1, forms a context in which the walking-in-place technique for locomotion at ground level can be considered. We argue that the walking technique is a shift along the y-axis of Figure 1, compared to the pointing technique, and therefore other things being equal should result in a greater sense of presence. However, we found that this is modified by the degree of association of the individual with the virtual body. In fact this factor spans both x and y-axes: lack of association may be due to lag between real and displayed virtual movements (y-axis), or immobility of the virtual left hands and feet (y-axis), or to the rather simple visual body model (x-axis). In any case, the VB association is significantly positively correlated with a subjective presence for the walkers but not for the pointers, which is certainly consistent with the proposed model.

In earlier work [Slater and Usoh 1994] we used the term "body-centered interaction" for techniques that try to match proprioception and sensory data. The walking-in-place method is a clear example of this. When the method works well it feels like walking, and the corresponding flow in the optical array matches both head movements and the movements of the feet. Also, the technique is very easy to understand for there is little to learn as such; therefore, this is less of a *metaphor* than other techniques. In this case we walk by "almost walking," rather than doing some other activity that is completely different from walking and then having to make the mental association between cause and effect. The empirical evidence does not support the notion that people prefer this for navigation compared to pointing, but it does suggest that improved performance of the neural net-based pattern recognizer may lead to such a preference.

We have described the technique applied to climbing or descending steps and ladders. This may be useful in circumstances where the interaction style should be relatively mundane, rather than requiring magical effects such as "flying." Training for fire fighting, the application that inspired the extension to steps and ladders, clearly falls into this category.

REFERENCES

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

BROOKS, F. P. ET. AL. 1992. Final technical report: Walkthrough Project, Six generations of building walkthroughs. Tech. Rep. Dept. of Computer Science, Univ of North Carolina, Chapel Hill, N.C.

COX, D. R. 1970. *Analysis of Binary Data*. Menthuen, London.

ELLIS, S. R. 1991. Nature and origin of virtual environments: A bibliographic essay. *Comput Syst. Eng. 2,* 4, 321–347.

FAIRCHILD, K. M., LEE, B H., LOO, J., NG, H., AND SERRA, L. 1993. The heaven and earth virtual reality: Designing applications for novice users In *IEEE Virtual Reality Annual International Symposium (VRAIS)* (Seattle, Wash., Sept. 18–22). IEEE. New York, 47–53.

GIBSON, J. J. 1986. *The Ecological Approach to Visual Perception*. Lawrence Erlbaum, Hillsdale, N.J.

GIBSON, E. J. AND WALK, R. D. 1960. The visual cliff. *Sci. Am. 202,* 64–71.

HEETER, C. 1992. Being there: The subjective experience of presence, telepresence. *Presence: Teleoper. Virtual Env. 1,* 2 (Spring), 262–271.

HELD, R. M. AND DURLACH, N. I. 1992. Telepresence. *Presence: Teleoper. Virtual Env. 1* (Winter), 109–112.

HERTZ, J., KROGH, A., AND PALMER, R. G. 1991. *Introduction to the Theory of Neural Computation.* Addison-Wesley, Reading, Mass.

IWATA, H. AND MATSUDA, K. 1992. Haptic walkthrough simulator: Its design and application to studies on cognitive map. In the *2nd International Conference on Artificial Reality and Tele-existence, ICAT 92.* 185–192.

KENDALL, M. 1975. *Multivariate Analysis.* Charles Griffin & Co. Ltd., London.

LOOMIS, J. M. 1992. Presence and distal attribution: Phenomenology, determinants, and assessment. In *Human Vision, Visual Processing and Digital Display*, vol. 3. SPIE, 590–594.

MCCAULEY, M. E. AND SHARKEY, T. J. 1993. Cybersickness: Perception of self-motion in virtual environments. *Presence: Teleoper. Virtual Env. 1*, 3, 311–318.

ROBINETT, W. AND HOLLOWAY, R. 1992. Implementation of flying, scaling and grabbing in virtual worlds. In the *ACM Symposium on Interactive 3D Graphics* (Cambridge Mass.). ACM, New York.

SACKS, O. 1985. *The Man Who Mistook His Wife for a Hat.* Picador, London.

SHERIDAN, T. B. 1992. Musings on telepresence and virtual presence, telepresence. *Presence: Teleoper. Virtual Env. 1* (Winter), 120–126.

SLATER, M. AND USOH, M. 1994. Body centred interaction in immersive virtual environments. In *Artificial Life and Virtual Reality*, N. Magnenat Thalmann and D. Thalmann, Eds. John Wiley and Sons, New York, 125–148.

SLATER, M., STEED, A., AND USOH, M. 1993. The virtual treadmill: A naturalistic metaphor for navigation in immersive virtual environments. In the *1st Eurographics Workshop on Virtual Reality*, M. Goebel, Ed. Eurographics Assoc., 71–86.

SLATER, M., USOH, M., AND STEED, A. 1994a. Steps and ladders in virtual reality. In *ACM Virtual Reality Science and Technology (VRST)*, G. Singh and D. Thalmann, Eds. ACM, New York, 45–54.

SLATER, M., USOH, M., AND STEED, A. 1994b. Depth of presence in immersive virtual environments. *Presence: Teleoper. Virtual Env. 3,* 2, 130–144.

SLATER, M., USOH, M., AND CHRYSANTHOU, Y. 1995. The influence of dynamic shadows on presence in immersive virtual environments. In the *2nd Eurgraphics Workshop on Virtual Environments* (Monte Carlo, Jan. 31–Feb. 1). Eurographics Assoc.

SONG, D. AND NORMAN, M. 1993. Nonlinear interactive motion control techniques for virtual space navigation. In the *IEEE Virtual Reality Annual International Symposium (VRAIS)* (Seattle, Wash., Sept. 18–22). IEEE, New York, 111–117.

STEUER, J. 1992. Defining virtual reality: Dimensions determining telepresence. *J. Commun. 42,* 4, 73–93.

VAANANEN, K. AND BOHM, K. 1993. Gesture driven interaction as a human factor in virtual environments—an approach with neural networks. In *Virtual Reality Systems*, R. A. Earnshaw and M. A. Gigante, Eds. Academic Press, New York, 93–106.

WARE, C. AND OSBORNE, S. 1990. Exploration and virtual camera control in virtual three dimensional environments. In *Proceedings of the 1990 Symposium on Interactive 3D Graphics.* ACM, New York, 175–183.

Received November 1994; revised February 1995; accepted February 1995