CHAPTER II

INTERACTION IN VIRTUAL ENVIRONMENTS

The research presented here has roots in several diverse fields, and builds on many previous results. In this chapter, we will briefly discuss prior work in related disciplines that has an overall bearing on this work. This includes concepts from the field of human-computer interaction, types of user interface evaluation, work in three-dimensional UIs and interaction, related work in the areas of perceptual and cognitive psychology, and previous efforts to evaluate components of immersive virtual environments. This general background will be presented here, but we will reserve discussion of research related to the particular tasks of viewpoint motion control, selection, and manipulation to the appropriate chapters devoted to those subjects.

2.1 Human-Computer Interaction Concepts

As we have noted, there exists a large body of work in the field of human-computer interaction that informs the current research. Many of the specific results and guidelines that are offered by HCI practitioners (e.g. Hix and Hartson, 1993) do not apply directly to immersive VEs, because of the difficulties of interaction in three dimensions, the difference in input and output devices, slower system responsiveness, and so on. However, these specific recommendations can often be generalized to principles that apply in any type of human-computer interface.

One set of general principles, or heuristics, were given by Nielsen (1993). He claimed that a small set of heuristics could account for a large percentage of the usability problems in any interactive system, given a sufficient number of experts to study the system. These heuristics are quite general (e.g. "speak the user's language"), and so they apply to any human-computer interface. However, this generality also causes the heuristics to be difficult to apply practically. In our research, we are searching for specific recommendations for virtual environment interfaces.

Some of the best known principles were described by Norman in the classic work entitled *The Design of Everyday Things* (Norman, 1990). These principles which apply to user interfaces are taken from a discussion of everyday artifacts that we use in our homes, schools, and offices. Since many virtual environments purport to represent a semi-realistic world, it is perhaps even more important that interaction in VEs follow these guidelines (Bowman and Hodges, 1995). Norman identifies four characteristics of usable artifacts: affordances, constraints, good mappings, and feedback. Affordances refer to the properties of an object that inform the user of its purpose and the way it can be used. Constraints are

limitations on the use of an object that guide users into proper actions. Good mappings mean that the conceptual model, or metaphor, on which an object is based is easily understood in the specific task domain of the object. Finally, feedback is the indication given by an artifact of the state of its operation or usage, to help the user understand what has happened so that the next action can be planned and carried out.

The idea of mappings proposed by Norman is related to previous work on the use of mental models (a user's understanding of the operation of an artifact) and metaphor (using the understanding of a known concept or object to explain the workings of an artifact) (Gentner and Stevens, 1983). The use of metaphor is an important strategy for UIs since we can explain to someone how to use a software application in terms of something he already understands. The risk is that an inappropriate metaphor will mislead or confuse the user, or that a forced metaphor, while understandable, may degrade user performance.

In traditional 2D user interfaces, there are two major categories of general metaphors. The *conversational metaphor* proposes a dialogue between the user and the computer in the form of a conversation. That is, the user issues a command and the system responds. This metaphor was largely used in command-line interfaces, such as a UNIX shell, but still exists in today's graphical UIs in the form of menu commands, dialog boxes, and so on. The other dominant metaphor is the *simulated world metaphor*, which represents the constructs of a computer application as objects with predictable behaviors in a mini-world. A common example is the desktop metaphor for personal computers, in which programs and data are represented as files which can be placed in folders, file cabinets, trash cans, and so forth, similar to the way paper documents are organized in an office. Since VEs are seen as virtual worlds, most use a very strong simulated world metaphor for almost all tasks. However, conversational elements may also have a place for certain actions in VEs.

Another important HCI concept relevant to this research is the notion of *task analysis*. Task analysis breaks down a task into its component parts, and formalizes the steps that must be taken to complete a task. This explicit characterization leads to a more detailed understanding of the task, and also to a more structured method for understanding various strategies applied to the task. When applied to UIs, task analysis can provide a framework for the design of ITs, as well as reveal reasons for the successes and failures of current approaches. We will use task analysis heavily in the design and evaluation of ITs for VE tasks.

There is a strong tradition in HCI and Human Factors research of formalizing models of human performance. Methods such as GOMS (Card, Moran, and Newell, 1983) and the Keystroke-level model (Card, Moran, and Newell, 1980) attempt to model human performance for a certain computer task by counting the numbers of low-level actions that must take place for the task to be completed. These low-level parts (based on a task analysis) may be explicit user motions, such as key presses, or cognitive processes that the user must carry out. By assigning time values to each of these low-level components, these models may also predict human speed for interfaces that have not yet been implemented or tested. Although our analysis will not attempt to model user action in such a fine-grained manner, we will follow the spirit of this earlier work.

Finally, the HCI literature has provided us with a number of techniques for UI evaluation. These methods represent a wide range in terms of cost, numbers of users needed, formality, and types of results. One of the most simple techniques is guideline-based evaluation (Nielsen and Molich, 1992), which is an informal analysis based on known principles such as those discussed above. This requires only that an expert or group of experts think about and/or use the interface briefly, and can often identify serious problems at an early development stage. The cognitive walkthrough technique (Polson et

al, 1992) is similar in that only UI experts are needed to carry it out, but it attempts to be slightly more formal by requiring the evaluators to follow a strict process and answer specific questions about each task within the interface. One of the most common evaluation methods is the usability study (Williges, 1984). Here, several users perform prescribed tasks with the UI, and are observed for task completion time, errors, and other issues. This method is slightly more time-consuming and expensive, but can identify important problems because of the fact that actual users participate. To obtain results that are even more applicable to the real world, some have also performed observations of users in the field (Holtzblatt and Jones, 1993), although it is questionable whether users work in a normal way while being observed. Finally, UI researchers can perform formal experiments in the scientific tradition (Eberts, 1994), which have specific hypotheses, are tightly controlled, and use statistical analysis to obtain results. These are the most expensive and time-consuming studies, and are usually used to obtain basic knowledge about an interface or technique that is quite different from that which has gone before it. Since our research falls into this category, we will make use of formal experimentation in our evaluation of ITs for virtual environments.

2.2 Three-Dimensional User Interfaces

User interface research has only recently begun to seriously consider truly three-dimensional applications and the added difficulties that they present. Common personal computer software is still almost exclusively 2D, except in a few niche applications. However, it is becoming increasingly important that 3D UIs are analyzed, understood, and designed well, as more 3D applications become mainstream. These applications include 3D CAD, architectural design, animation, visualization, and even entertainment. In all of these cases, the fact that information is displayed and manipulated in three dimensions provides a new challenge for UI designers. Some of the problems have been identified and categorized (Herndon, van Dam, and Gleicher, 1994, Hinckley et al, 1994), but there are few general principles or solutions for these difficulties.

For desktop 3D applications, the limitations and inherent 2D nature of common input devices, such as the mouse, pose a major challenge. In these cases, the two degrees of freedom (DOFs) of the input device must be mapped onto three, or in some cases six (three translational and three rotational), dimensions. For this reason, a good deal of research has gone into the design and analysis of input devices specifically for 3D applications (MacKenzie, 1995). One of the most common devices is the tracker (Meyer and Applewhite, 1992), which is a 6 DOF digitizer – that is, it is a sampling device which continuously outputs six scalar values representing position and orientation. Other devices like the Spaceball (TM) (Spacetec IMC, 1998) and the Sidewinder (TM) (Microsoft, 1998) are self-centering devices which sense displacement and rotation in all six dimensions. Other designs have focused on modifying the mouse, such as the "Rockin' Mouse" (Balakrishnan et al, 1997).

Analysis of input devices has been an important research topic in recent years. Card, Mackinlay, and Robertson provided a formal framework for design and evaluation of both 2D and 3D devices (Card, Mackinlay, and Robertson, 1990). Other studies have focused on the experimental comparison of two or more of these devices. For example, Zhai and Milgram (1993) compared the tracker to the spaceball for an object placement task. One problem with many of these studies is their implicit assumption that the input device and the interaction technique are inextricably linked. That is, an input device determines the IT that

must be used with it. We recognize the importance of a well-designed input device in a usable system, but claim that ITs can be evaluated separately. All of our experiments, then, will use a tracker for 6D input, but a multitude of different ITs will be studied.

Another area that has seen many research efforts is the standardization of 3D interfaces, analogous to the ubiquitous desktop metaphor in 2D. It has been argued that such standardization is necessary before VEs can become accepted tools in the real world. Several research efforts have attempted to provide a single interface metaphor that can allow usability and productivity for a wide range of VEs (e.g. Wloka and Greenfield, 1995, Rygol et al, 1995). We would claim, however, that standardization is not necessarily beneficial for VE interfaces at their current level of maturity. Rather, we will focus on optimizing the interaction for specific tasks in particular domains.

Recently, the field of *two-handed interaction* in three dimensions has been researched extensively. Two-handed interfaces are a new paradigm for 3D input that attempt to take advantage of the human ability to use both hands simultaneously to provide more intuitive, comfortable, and productive applications. Hinckley's work in this area is quite instructive (Hinckley et al, 1997). Using previous work in the analysis of two-handed tasks such as handwriting, he showed the validity of several principles for two-handed interfaces. These include the ideas that the hands should work complementarily, not necessarily in parallel, that the non-dominant hand provides a frame of reference within which the dominant hand works, and that the non-dominant hand is good at large, coarse-grained manipulation, while the dominant hand excels at fine-grained work. These principles have been applied to several non-immersive 3D applications (e.g. Goble et al, 1995, Mapes and Moshell, 1995), with encouraging results. We feel that the use of two-handed interfaces in immersive VEs is quite promising, and therefore will include two-handed techniques in our design and evaluation.

2.3 Perceptual and Cognitive Psychology Concepts

Since our research focuses on human performance when interacting with VEs, it is only natural that we should use the results of prior work investigating human capabilities in general. Much of this information comes from the fields of perceptual and cognitive psychology. Perceptual psychology studies the ways humans perceive their environment through the senses, while cognitive psychology focuses on the mental aspects – how humans reason, learn, remember, etc.

Since most immersive VEs are highly visual, it is quite important that we understand human visual perception. In particular, depth perception is crucial, since we are attempting to represent a 3D environment on 2D displays. Research has identified many visual cues that humans use to determine depth, and divide them into monocular vs. binocular (using one or two eyes), and static vs. dynamic (available from a single image or requiring motion) cues (Bruce and Green, 1990). Most depth cues are static and monocular, including linear perspective, texture gradient, relative height, and aerial perspective. Motion parallax, referring to the understanding of depth gained from head or eye motion, is a dynamic monocular cue. Stereopsis – the depth effect due to the fact that our two eyes receive two slightly different images of the world – is characterized as static binocular. Finally, there are oculomotor depth cues, which, unlike the others, do not depend on the images received at the retinae. These cues rely on information from the muscles which cause the eyes to focus (accommodation) and rotate (convergence). We cannot achieve a perfect representation of depth in current VEs, because the actual images all appear on a

screen at a single depth, and therefore the oculomotor cues – cues based on the convergence angle and accommodation of the eyes – are in conflict with other depth cues.

Stereo in particular is widely believed to be a very important depth cue that enhances immersive VEs. However, it is quite difficult to achieve a proper stereo effect, as it requires care in calibration, measurement, and rendering of the stereo pair (Davis and Hodges, 1995). Many studies have been performed comparing human performance in stereoscopic, monocular, and biocular (the same image presented to both eyes) viewing situations, and the general consensus is that stereo improves presence and can improve performance (Barfield, Hendrix, and Bystrom, 1997, Hendrix and Barfield, 1996). On the other hand, some studies have found that the addition of other cues to a non-stereo display may produce performance that is as good or better than performance with a stereoscopic display (e.g. Nemire, 1996). Because of technological limitations, our studies will use biocular displays, but will include many additional depth and feedback cues to aid performance.

Wickens has presented a good summary of the application of cognitive psychology to VEs (Wickens and Baker, 1995). An important concept from cognitive psychology that relates to the current work is the model human processor (Card, Moran, and Newell, 1986). This describes the cognitive process that people go through between perception and action. It is important to the study of interaction techniques because cognitive processing can have a significant effect on performance, including task completion time, number of errors, and ease of use. A major goal of IT designers is the creation of techniques which use few cognitive resources, and may become automatic in some sense, so that cognitive power may remain focused on the actual task at hand. One particularly important concept is the limitation on working memory described by Miller (1956). He reported that working memory can hold only seven plus or minus two "chunks" of information at a time. If more information needs to be recalled, previous chunks may be displaced or interfered with. Interfaces should be designed so that this limited space can be used for domain-specific information.

Finally, perceptual and cognitive psychology have shed light on individual differences in the ability of humans. One such line of work that relates to the current discussion is the study of spatial ability (McGee, 1979). Humans vary in their ability to reason spatially, especially in three dimensions. Studies such as the classic mental rotation experiments (Cooper and Shepherd, 1978) have demonstrated these differences. A user's spatial ability can have a significant effect on their performance in 3D interaction tasks. Therefore, we must be sure to consider individual differences when designing and evaluating ITs. Designers should attempt to create techniques which perform robustly for users with a wide range of spatial abilities. In evaluation, we must take care not to attribute a performance difference to the difference in techniques when it is actually caused by a user-specific characteristic, such as spatial ability.

2.4 Evaluation of Immersive Virtual Environments

Although virtual environments have been in existence for over thirty years, it has only been in the last few years that researchers have really begun to perform analysis and evaluation of technology, techniques, and applications of VEs. As stated earlier, this type of research is necessary if VEs are to become useful in the real world. In this section, then, we will review some of the work that has been done to quantify the effectiveness of VEs and human performance within them.

One question that should be asked at the outset is, "What evidence is there that immersive VEs are better than other types of computer applications for ANY tasks?" Researchers have addressed this issue in both a general sense and in specific applications. For example, Pausch et al (1993) performed a study comparing human performance using a head-mounted display with and without head tracking, and reported that head-tracking had a significant effect in improving results. In the application domain, Hodges et al (1995) have shown an immersive VE can produce results for psychological therapy which are similar or equivalent to those achieved when a physical environment is used.

Another issue that has intrigued researchers is the measurement of presence, or immersion. Barfield has attempted in several studies (Barfield et al, 1995) to relate the level of presence to task performance. Slater's work (Slater, Usoh, and Steed, 1994, 1995) has examined the effects of various display modalities, interaction techniques, and system algorithms on the reported level of presence. One problem with this type of research is the lack of a standard definition of presence and an appropriate measurement technique. Most studies have used qualitative measures (e.g. interviews or questionnaires), although some have attempted to relate other values to the sense of presence.

Another area of current research is the effect of various low-level system characteristics on performance in immersive VEs. Besides the studies addressing display type mentioned earlier, there have been experiments on the effect of mean frame rate (Richard et al, 1995), variance of frame rate (Watson et al, 1997), and level of visual detail (Watson, Walker, and Hodges, 1995). These experiments have generally used a standard task, such as visual search or pick and place, and compared users' speed and accuracy under the various experimental conditions. Such studies are similar in format to those we will present in this work, although our main independent variables are higher-level entities (interaction techniques). Based on this body of work, our studies will attempt to provide a "near best case" system environment, with a high average frame rate, low frame rate variance, and high visual detail in the entire display, so that our results will not be confounded by these variables.

Finally, recent research has attempted to apply common HCI design and assessment techniques to VEs. The most common example of this is the summative usability study, in which users do a structured set of tasks within a complete system or prototype system in order to reveal usability problems that can be solved in the next design iteration. It is becoming more common for VE developers to perform usability studies to verify the effectiveness of their designs (e.g. Bowman, Hodges, & Bolter, 1998, Arns, Cook, & Cruz-Neira, 1999). The concept of usability engineering includes guidelines and evaluation throughout the design cycle of a system, and this model has begun to see use for VEs as well (Hix et al, 1999).

There has been little work in the evaluation of specific interaction techniques for immersive VEs, although this may be changing. We will forgo a discussion of this body of work for now. Rather, in each of the chapters on a specific interaction task, we will review the relevant research on IT design and evaluation for that task.