VEMap: A Visualization Tool for Evaluating Emotional Responses in Virtual Environments

by

Hong Zhu

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Systems Design Engineering

Waterloo, Ontario, Canada, 2009

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

VEMap (virtual emotion map) can be seen as an advanced application of virtual environment (VE) technology to aid with design activities in architecture and urban planning, which can assist designers to understand users' opinions. The aim of this research and development work is to create a software application that allows designers to evaluate a user's emotional response to virtual representations of architectural or urban planning environments. In this project, a galvanic skin response (GSR) test is adopted as an objective measurement for collecting skin conductance data representing emotional arousal. At the same time, the user's self-reports are used as a form of subjective measurement for identifying emotional valence (i.e. positive, neutral, and negative). Finally, all of the information collected from both GSR readings (objective measurement) and self-reports (subjective measurement) are converted into coloured dots on the base map of the corresponding virtual environment (VE). According to the results of the VEmap evaluation and validation procedure, the beta-testing and evaluation of this project has been confirmed that VEmap may interpret users' emotional changes as evoked by VE mostly. From a usability perspective, there is no obvious difficulty present for participants on all the controls. Moreover, according to participants' comments, VEmap may increase users' interests and promote their involvement if it is applied in architectural design and urban planning. However, gender might have influence on self-report part, and virtual reality usage or 3D game experiences might affect navigation in VE.

Acknowledgements

Table of Contents

List of Figures

List of Tables

1 Introduction

Virtual environment (VE) technology, as a medium for creating and exploring design ideas, has been researched and applied gradually in the fields of architectural design and urban planning (Drettakis, 2007). Through the aid of VE technology, designers can use computers to generate two-dimensional (2D) and three-dimensional (3D) visualizations of their design concepts, allowing them to save the time and money that would normally be spent on building scaled physical models or sample buildings (Al-kodmany, 2002). As 3D computer-aided design software becomes more advanced allowing for higher resolution models, clients or potential residents of buildings can observe more details about the models, and understand those conceptual models first-hand without the confusion caused by trying to imagine 2D sketches as 3D models. With VE technology, accessories such as immersive head-sets that track head movements, designers and their clients can literally look around in the proposed design as well as zoom in and out of these models to see more detail or gain different visual perspectives of the 3D designs.

To date, the most common way of gathering feedback from clients and residents on their response to architectural designs is through subjective measurements, such as face-to-face interviews, and questionnaires (Bishop, 2003). Yet subjective measures can be limited in urban planning research relating to human response and potential interaction with proposed designs. For instance, designers can obtain feedback on whether potential clients report that they like a particular design. However, designers might not know the level of preference should their clients provide positive feedback, and whether that preference would be strong enough for the client to purchase the product or service. Sometimes users may have difficulties on deciding because of the subtle differences between products or because of complex comparison. Or users may modify their opinions deliberately if they do not want to express their real feelings (Laparran-Hernández *et al*, 2009). Alternatively, behaviour is subconscious and what people may actually do might be different from what they think they would do or what they report to the designer (Norman, 2004). For this reason, collecting

objective measures which have less bias will help overcome some of the limitations of subjective measures collected as part of design reviews.

One physiological response that is commonly used to objectively measure emotional arousal level is Galvanic Skin Response (GSR) (Dindo and Fowles, 2008; Meehan, 2001). GSR data is considered to be valid and reliable for marketing research (Wang and Minor, 2008), and has been used to effectively differentiate emotional responses to subtle design alternatives, such as changes in floor coverings (Lapparan-Hernandez et al, 2009).

The research and design of the software application － Virtual Emotional Map (VEMap), described in this thesis was inspired, in part, by an innovation called Bio Mapping by artist Christian Nold (2004). Nold (2004) collected GSR data while participants were navigating around the city of London, England using a handheld Global Positioning System (GPS). He then "mapped" the GSR data at locations corresponding to the GPS route using coloured dots to produce an "arousal" map of the participant's experience. However, a limitation of Nold's (2004) work is that he did not ask participants what they were feeling as they walked through the city, so he did not know whether the arousal levels represented positive or negative emotions.

Building upon Nold's Bio Mapping idea, the VEMap is designed to capture the emotional responses of users evoked by their real-time experiences within a VE. The VEMap application collects galvanic skin response (GSR) as an objective measure, and self-reports of the participant's general emotional states which are collected through on-line prompted questions that are encountered at key places of interest to the designers in the VE. The data collected through the VEMap application is then used to produce a "colour-coded" map of the individual's emotional response to the environment. As a result, the final VEMap displays the information about places that users have visited within the VE, and the corresponding emotional responses along the route. This information includes emotional arousal, which is described as a process that happens as a sequence of responses over time, indicated by some physiological measures such as heart rate and GSR. In addition, VEMap can report the information of emotional valence as well, which means the status of emotion,

whether positive or negative (Mandryk, 2005). With the help of VEMap, both designers and clients may better understand emotional responses to various design aspects of proposed real environments, and be able to make design or purchase decision accordingly.

1.1 Overview of Thesis

Chapter 2 provides background on the role of virtual environment technology as a design medium; the potential for using galvanic skin response (GSR) as a means of evaluating 3D designs; and the inspiration for the design concept of the Virtual Emotion Map (VEMap) application. Chapter 3 presents the research and development of the VEMap in terms of software and hardware implementation as the design component of this research and development project.

The research and design work on the VEMap described in this document has been carried out to a "proof-of-concept" stage. In order to validate the design of VEMap as a potential architectural and urban design evaluation tool, user tests were conducted with 14 participants. A description of the experimental setup and the results of the exploratory beta-testing study are covered in Chapters 4 and 5 respectively. A discussion of the results of the validation study is presented in Chapter 6. Chapter 7 presents recommendations for improvements to the VEMap application, as well as areas for further research into the use of emotional response measures for design assessment.

It must be noted that the objectives of the data collection described in Chapter 4 and 5 were to consider the "emotional maps" generated as well as participant comments made during the experimental trials, in order to conduct a first-pass review of the VEMap design. As described in chapter 7, the main goal of this project is validating that the design of VEMap can be useful in architectural design and urban planning field. It must also be noted that while males and females of varying levels of computer-gaming familiarity were included, the user performance data collected as part of this work are considered to be exploratory and do not represent a rigorously controlled experiment designed to test specific hypotheses relating to emotional responses to different types of virtual environments. The use of the VEMap as a platform for controlled experiments into emotional response to VEs is for future work.

2 Background

2.1 Virtual Environments as Design Media

The use of virtual environment technologies to learn about human behaviour and assess design usability has received attention in recent years (Bishop, 2003; Drettakis, 2007). Technically, virtual reality (VR) technology or more appropriately virtual environment (VE) technology (as it is difficult to truly achieve simulated reality) is used for real-time interaction between human beings, computer-generated interfaces and extremely complex data (Schroeder, 1993). It is a connection between the computer graphics technology, and various multimedia peripherals, such as transducers, sensors and digital input devices. VEs involve 3D simulated environments where users can perceive and interact with virtual objects which can be moved with six degrees of freedom in order to make them comparable to their counterparts in the real world. By navigating through a VE, manipulating the virtual objects, and performing other related actions, users can have immersive feelings in this simulated world according to feedback they get from output devices, such as visual and acoustic displays, and haptic devices that provide a sense of touch. The displays for VEs may provide total visual immersive effects where the VE graphics fill the user's field of vision (e.g. display wall, or head-mounted display), or may be partially immersive where the VE graphics still provide a sense of navigating in or around 3D worlds as can be commonly achieved while viewing the graphics for 3D video games on a home computer display monitor (Mazuryk and Gervautz, 1996).

Thanks to the improvements in computer hardware in the way of sensors and tracking devices and the software that can produce real-time 3D graphics, VEs can be used to produce 3D models that users can virtually "walk" through in order to carry out design reviews of products and designed architectural spaces (Howard and Gaborit, 2007). For example, the Urban Simulation Team at the University of California Los Angeles (UCLA) has been conducting research on the application of VEs to urban planning since the 1990s.The team has built a virtual model of almost the entire Los Angeles basin, which allows users to interactively fly, drive or walk through the city in real time. As an advanced visualization tool, VE techniques can be used through the whole design process, such as assisting on

prototyping during the design phase; and assisting on communication and evaluation between stakeholders, designers, customers and other related experts (Harrold, 2000; Drettakis, *et al*. 2007). However, while the fields of computer science and computer engineering have made advances on the technology side of creating and displaying virtual environments, there is still much to learn from a human factors perspective on how to best make practical use of VEs as a tool for improving the design of real world devices and spaces.

2.2 Using VEs for Architectural Design and Urban Planning

Virtual environment technology applications began with flight simulators for training pilots that were built by the aviation industry and U.S. Air Force during World War II (U.S. Congress, 1994). VE technology is applied not only in aviation and military training. Other areas including industry, medicine, education, entertainment, commerce, architecture and urban planning, are starting to take advantage of this technology for various purposes such as visualization, education, training and collaboration (e.g. Basdogan *et al*, 2007; Bell *et al*., 2004; Brown, 2006; Johnson *et al*, 2002; Sebok *et al*, 2002; and Thacker, 2003).

Besides applications for education and training, VEs can be vivid and interesting tools to encourage creative interaction and imagination. Because of their lively animations, colourful graphics and challenging environments for playing alone or with others, VEs have been used in the video game industry for entertainment purposes for many years. Beyond 3D video games, VEs are starting to be considered as an attractive tool for creating and promoting other application areas. For instance, VEs are now being used to demonstrate and "pre-sell" architectural designs by allowing potential clients and buyers to take a "virtual" walk through the 3D model to get a sense of the space, rather than just looking at 2D plans or a small scaled physical model of the building. For example, a journal article said that store sales in New Sports Square in Shanghai increased, and the mall was in fact sold out ahead of time because of a VE walk-through system (Chen, 2007).

The use of VEs is being gradually applied in the ecommerce field as well. By interacting with the virtual products such as houses or cell phones on a website, clients may get an increased feeling of ownership. With the help of designers or salesmen, users can customize some

product details according to their desires and make a unique product which they are more likely to want to purchase (Soo *et.al*, 2006). For instance, the 3D virtual shopping mall designed by Zhao and colleagues of Ningbo University and Zhejiang University, and the AdapTIVE Project designed by Dos Santos and colleagues of Unisinos University are examples of shopping applications of VE technology used in ecommerce areas (Zhao *et al*, 2004; Dos Santos *et al,* 2004). Unfortunately, true customization of designs for clients is still in a research phase due to the limitation of internet bandwidth and other computer-related technologies. Users seldom want to spend time waiting to download software or plug-ins, or learning how to manipulate confusing systems on an ecommerce website. As a result, human factors should be considered carefully while designing VE systems to be used to help market products or designs.

Obviously one important advantage of VE technology is that it can be used as a visualization tool for communication and representation of design ideas. Those objects which are hard to view such as atoms in a micro-world or very large skyscrapers can be transferred into detailed virtual models that can then be viewed from different perspectives (e.g. zoom in, birds-eye view). Those abstract concepts, which are often only imaginable and understandable by experts, can be converted to detailed models that can be shared with others. Therefore, VEs can help designers more clearly communicate their ideas and concepts to their customers (Mazuryk and Geruatz, 1996).

Designers, especially in architecture and urban planing areas, can experience difficulty in expressing conceptual models to partners and clients (Al-Kodmany, 2002; Ye, 2006). Sometimes because of limited specialized knowledge or the limits of models, it is difficult for the clients to understand completely how the space is to "look" or how the space will "feel" when they are in it. Misunderstanding and miscommunication on both sides can lead to expensive and repeated modification of work for designers, even a failed bid or contract if the client and designer cannot agree on the concept (Harrold, 2000). Take architectural design as an example: with a virtual building, clients can walk through and get a clear idea of the interior layout, decorations and details (e.g. wall and floor colours, furniture layouts). At the same time, suggestions or even some changes can be recorded automatically into the

system, such as a reminding assistant. In this case, designers can focus on expressing their ideas through the computer-generated models instead of being distracted by trying to remember clients' comments or advice. Furthermore, VEs make it easier for users to observe buildings of any size and from any perspective (both from inside and outside) if the model is in 3D instead of having to look at 2D graphics and then trying to imagine the same model in 3D space. Xu (2004) provides an example of how a 2D sketch of a floor plan and 3D virtual model can make it easier to understand the whole structure (See Figure 1).

Figure 1:3D walk through environment versus 2D architectural sketch, Original Source: Xu (2004)

Designing with VE tools can lead to a better understanding of building concepts that involve large data sets with respect to scale, context and structure. The 3D visualizations can aid brainstorming, which is the first stage in architecture design. VE tools can also allow alternative designs to be compared in order to choose the most balanced option compatible with the target environment, construction requirements and stakeholder demands (Al-Kodmany, 2002). Computer-aided design (CAD) software, such as AutoCAD® and 3ds Max®, are commonly used by designers for realizing 3D models with details such as high-quality geometry and textures for buildings, accurate and consistent lighting and shadows, traffic patterns, vegetation, and even individual users to demonstrate how the space could be used.

Typical CAD drawings can be limited in terms of immersive experience for the user due to limitations in their graphic renderings and desk-top based interactions. To improve upon these options, new VE design tools have been developed and studied which might be applied

to real-world design situations in the future. For example, Leetau (2004) describes a project, named "Shadowlight-Mirage" developed by Emma Smith. Shadowlight-Mirage allows designers to stand in front of a large computer display and draw "full-body" models that can then be viewed and enjoyed later as part of an immersive environment instead of just watching those models on a desktop computer (See Figure 2).

Figure 2: Shadowlight design project (Original source: Leetau, 2004)

The benefits of more interactive VEs can lead to a more satisfactory planning process or design according to the research of the Urban Simulation Team at UCLA (Howard, 2007). Furthermore, with the development of broadband internet techniques and the proliferation of the world-wide-web, designers and clients from different locations can view the same VE world or model, which can be watched, designed and modified online collaboratively (Harrold, 2000; Simpson, 2001; Wilson, 2000; Bell, 2004). A desired marketing outcome of such VE applications is to provide customers and clients with an immersive emotional experience so that they can provide accurate feedback on how they truly feel about a proposed design, product, or service (Laparran-Hernandez et al, 2009). This feeling is seemed as emotional response evoked by VEs.

So how does one measure emotional response to design? There are two main approaches of interest for this research and development work: 1) subjective measures based on participant self-reports of their emotional state; and 2) objective measures based on the physiological responses of the participant to the stimuli provided in the VE as gauged through galvanic skin response (GSR)

2.3 Evaluating Emotional Response in VEs Using Self-Reports

Emotional arousal, also called "activation", is a process within the sympathetic nervous system which indicates that the body is "preparing for fight or fighting" (Mandryk, 2005). During emotional arousal, the body typically experiences physiological responses of arousal (e.g. increased heart rate, increased sweating) even before the person is able to identify what emotional valence (positive or negative) is being felt.

From a usability perspective, the most commonly used method for evaluating users' feelings regarding a design is to simply ask them through questionnaires or interviews. However, while collecting subjective measures may be relatively easy from the researcher's point of view, the data collected can often be questionable. For instance, participants who are asked to report on their emotions after using a product may be only able to give a global assessment and may forget what their initial emotional response was, or what their emotional responses to specific parts of a test session were. As well, classifying specific emotions beyond "positive, neutral, or negative" may be difficult for the participant as emotional responses can be subtle (Laparran-Hernandez *et al*, 2009). In addition, participants may have compelled to report emotional states that they think the designer expected them to have experienced – rather than reporting what they actually experienced.

Similarly, researchers and designers may be influenced by their own expectations, and inadvertently interpret ambiguous self-reports of emotions in favour of their personal biases. For this reason, including objective measurements can increase the validity of the interpretation of emotional responses. Studies have shown that both GSR and heart rate are increased when participants performed poorly in a VE, or when they felt scared by some virtual stimuli (Meehan, 2001).

2.4 Evaluation of Emotional Arousal by Testing Galvanic Skin Response (GSR)

Testing galvanic skin response (GSR) has been researched for over one century and is generally accepted as an efficient and reliable way to measure emotional arousal level (Wang and Minor, 2008). Recently, GSR data has been collected as a means of providing insight into the emotional reaction of people to computer-generated designs and real-world urban spaces (Meehan, 2001, Laparran-Hernandez et al, 2009).

2.4.1 Overview of GSR as a Measure of Emotional Arousal

Electrodermal activity (EDA) refers to those active and passive electrical properties which occur in the skin. Galvanic Skin Response (GSR), also known as Skin Conductance Response (SCR), is one of the most common EDAs measured and is defined simply as "a change in the ability of the skin to conduct electricity" (Boucsein, 1992). For over a century, psychologists starting with Carl Jung have used GSR as a means of measuring arousal levels within the human, and then drawing conclusions as to the human's emotional state (positive or negative) corresponding with the level of arousal. For example, the trigger could be variable, such as auditory stimuli, images, 3D computer-generated environment and even real environment, depending on the purpose of the study (Boucsein, 1992; Wang, 2008; Meehan, 2001; Nold, 2004).

Boucsein (1992) provides a detailed explanation of the role that the sweat glands in the skin play in conducting electrodermal activity (EDA) which can be measured as an indicator of emotional arousal. A GSR measurement system is basically a very simple electrical model, in which the skin surface and corresponding dermal tissue can be treated as a simple resistor. By applying direct current to the skin in this model, typically from a simple source such as a dry cell or a battery cell, the instantaneous current flow can be detected and measured by applying Ohm's Law ($R=V/I$) (Figure 3).
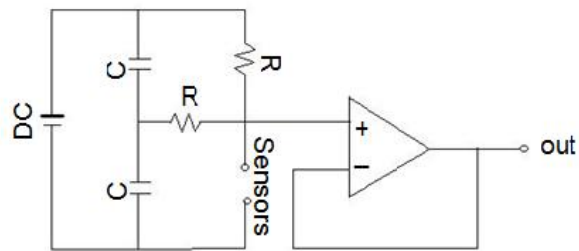
Figure 3: Basic schematic diagram of GSR meter

According to Prokasy (1973), the first research on the psychological significance of EDA and its relationship with emotional arousal was done by Féré at the end of 19th century. He measured skin resistance response during his study, and found that the sensation of emotional stimulations was accompanied by increases in skin conductance. That was the first attempt to use EDA as an index of emotional arousal. This theory was developed further by a number of studies in mid-20th century (Meehan, 2001). Consequently GSR is thought to be one of the most effective indexes to interpret emotional arousal (Wang and Minor, 2008).

From a physiological perspective, the sweat gland activity can be increased when there is a high level of emotional arousal. This phenomenon is called emotional sweating, and can be observed and measured most easily for research purposes on palmar (hands) and plantar (feet) sites (Boucsein, 1992). Studies have separated emotional arousal into two primary classes—a) the "peaceful" class which is characterized by very low arousal and is often associated with meditative states relating to reverence or blissful love; and b) the "aroused" class which is characterized by high levels of arousal often associated with anger, grief, romantic love, and joy. Statistical analyses of GSR signals can easily distinguish between high (aroused) and low (peaceful) arousal levels. However, there is no strong evidence to clearly indicate differences between positive and negative emotions by just analysing GSR data (Vervliet, 2004; Meehan, 2001).

In addition, two main other factors — environmental conditions and individual differences can also impact on GSR readings, and consequently on the interpretation of those readings in terms of emotional responses. Environmental conditions relating to temperature and humidity can influence skin moisture and sweating levels, and as a result confound GSR readings.

Setting up a stable environmental situation especially the room temperature at 23ºC can decrease such influences, and minimize the experimental error (Boucsein, 1992).

A propensity for sweating can also be influenced by gender, age, and race, thereby contributing to individual differences in GSR readings among participants in the same experiments. This is why it is important to establish baseline GSR readings to accommodate individual differences among a group of participants.

The most common sites for detecting GSR used by researchers are palmar or plantar surfaces, more specifically the medial phalanges of the index and middle fingers for bipolar recordings (Figure 4).
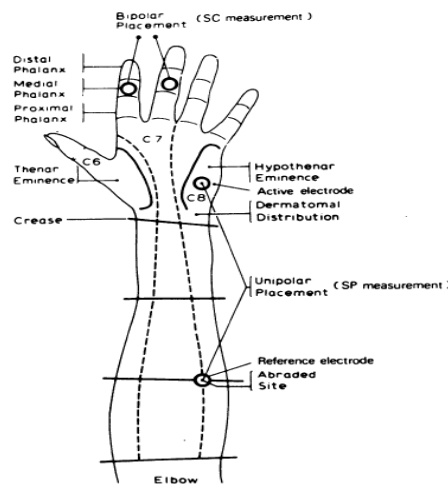
Figure 4: Recommended palmar electrodes sites for measuring skin conductance response (SCR) (Original Source: Christie, 1981)

Electrodes can be fixed easily on the medial phalanges of index and middle fingers of the participant's non-dominant hand during measurement procedure. The specific fingers to be used can be chosen so that they do not interfere with the main activity that the participant has to perform while GSR data collection is under way (Christie, 1981).

Pretreatment of these sites is typically not necessary except for extremely oily skin where the skin surface can be treated by alcohol. Because common pretreatment, such as washing with soap, can decrease NaCl concentration, and reduce skin conductance further. Boucsein (1992)

points out that the most important benefit of this measurement is simplicity and therefore there is no need to clean skin surface in a complicated way (Boucsein, 1992).

Sometimes what the subject does during a short period of time before his or her arrival can affect the measurement slightly. GSR data obtained from a subject who performs in a quiet and non-active way before an experiment may differ from someone who rushed upstairs and arrived at the last moment. In this case, keeping the previous activities of subjects under control for a set period of time before doing the measurement can minimize this kind of error (Boucsein, 1992). For example, asking subjects to fill out some pre-questionnaires followed by introducing them to the whole experiment or training them to use some equipment will help to keep their pre-experimental activity similar. It is recommended that the baseline measurement should be taken 20-30 minutes after arriving at the test section (Boucsein, 1992).

2.4.2 Record and Analyze Data

There are two types of skin conductance which are recorded as part of the GSR measures: tonic and phasic. Tonic skin conductance is the baseline level of skin conductance without any particular environment condition, and is often referred to as skin conductance level (SCL). Individual factors such as age, gender, and race may have an impact on the reading, which means that each person has a particular SCL (Prokasy, 1973).

Phasic skin conductance changes in response to environmental stimuli (visions, sounds etc.), and is correlated with emotional sweating which is generally referred to as skin conductance response (SCR) or galvanic skin response (GSR).

Depending on the goals of the experiment or research, data collection involves just GSR, or both skin conductance level (SCL) and GSR. The most basic calculation for plotting an arousal effect to a stimulus is to subtract the baseline measured as the SLC from the reactive response conveyed by the GSR value (Prokasy, 1973). This calculation is appropriate if one is interested in general emotional responses that occur at specific target points within a VE.

2.5 Understanding the role of emotion in VEs

Emotions are implicated in our phenomenological understanding of the VEs, the same as understanding the physical world around us. So information on emotions evoked by VEs in users, the feeling of presence in particular, is valuable for researchers or designers when evaluating the usability of VE applications. For example, whether game players can feel excited in the game world; whether trainees can use the knowledge and skills they gain in a 3D simulation fluently; whether potential customers can feel at home in the virtual condo – all contribute to the user's emotional experiences in the VEs. Subject's attention and reaction in VEs can be checked by monitoring their emotional responses from which researchers or designers can gain further insight into the effectiveness of the design so as to modify them accordingly. How to best evaluate emotions evoked by VEs using subjective methods and objective methods separately or combined still needs to be more fully researched (Morie, 2005).

2.5.1 Motivation: Emotional Mapping in the Real World

The concept for the design of the VEMap was motivated by the technology dubbed "Bio Mapping" that was first developed in England by Christian Nold (2004). Nold's Bio Mapping system is used to map a person's emotional response to a walk through a real world environment by collecting data using a hand held global position system (GPS) and a portable device to record galvanic skin response (GSR) data. Since GSR data is also collected in basic "lie detector" systems, Nold (2004) referred to his system as

"Bio Mapping = Lie Detector (GSR) + GPS" (See Figure 5).

Figure 5: Device (GSR tester + GPS) used in Bio Mapping project
(Original source: Nold, 2004)

In this study, the only thing the participant was asked to do was to walk in a prescribed area of a city and to return to the starting point within one hour. Data including global position and GSR were recorded simultaneously while the participant was navigating. After the experiment, the data was uploaded into custom-built mapping software in which the GPS data was converted into a "dot" on a geographical map and the GSR data collected at each GPS point of interest was used to "colour" the dot. Map dots with high GSR values were coloured red to indicate high emotional arousal; while map dots associated with low GSR values were coloured green to indicate lower emotional arousal. Figure 6 (a) shows an example of Bio Map for one participant. At the end of each Bio Mapping workshop in each city, the Bio Maps of all participants were collected together to make one comprehensive Emotion Map of that city, See Figure 6 (b).
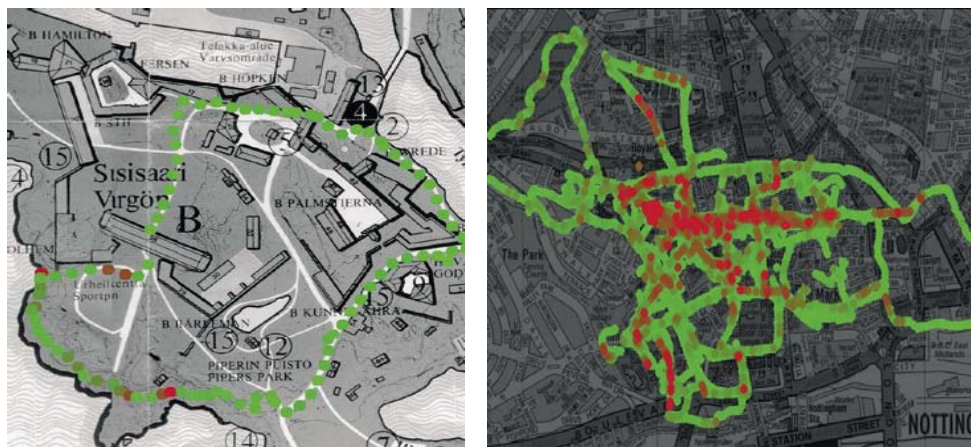
Figure 6: Emotion map of one participant (a); Overall emotion map of one city (b)
(Original Source: Nold, 2004)

Nold then changed the visual mapping strategy of emotional response to an environment from using 2-D colour dots into raised (3D) triangles. In the second version of Bio Map, the peak of each triangle represents the level of emotional arousal (Figure 7). The higher the peak displayed; the higher the emotional arousal level of that participant. So far, emotional maps have been created for large urban cities such as London, Brussels and San Francisco (Nold, 2004).
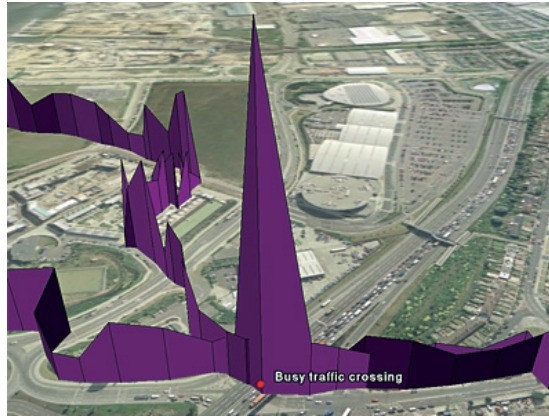
Figure 7: Bio Mapping using 3D triangles (Original Source: Nold, 2004)

From the perspective of truly mapping emotional responses to real or virtual worlds, a limitation of Nold's (2004) work is that he did not ask participants what they were feeling as they walked through the city, so he did not know whether the arousal levels represented positive or negative emotions. Thus it is difficult to tell whether the participants actually felt very positive about the spots that evoked high arousal level (e.g. a beautiful garden), or whether the high arousal levels were generated by anxiety or fear provoked by the urban setting (e.g. a dark alley way). In a complex urban environment, it is likely that some high arousal was evoked by positive aspects of the urban environment, and some high arousal was evoked by negative aspects. While the map of the city provides a context for the arousal levels, it is limited in terms of the feedback it can give designers. This research and development project looks to find a practical way to combine GSR with participant self-reports of their emotional responses to VEs in real-time.

3 Design of VEMap

3.1 Research and Design Objectives

Nold's (2004) Bio Mapping technique can be used to evaluate general emotional arousal while navigating in the real world environments. It was not clear whether the same technique could be reasonably applied to evaluating the emotional responses of participants while experiencing VEs. Thus, the main objectives for this research and development project were to explore the possibility of applying the basic Bio Mapping technique to the collecting of emotional responses to VEs; and to develop a "proof-of-concept" application, to be called Virtual Emotional Map (VEMap) that could be beta-tested with human participants. As mentioned earlier, a design tool that allows planners and architects to better understand the emotional responses of users to virtual models of buildings and public spaces could be a cost effective way of evaluating potential designs from the perspective of potential customers or users.

3.2 Design Challenges

Adapting Nold's (2004) Bio Mapping concept as a practical tool to evaluate emotional responses to VEs requires combining (understanding) of psychophysiology with engineering design and visualization techniques. One drawback of Nold's (2004) Bio Mapping technique is that it reflects raw arousal level, as reflected by GSR data, rather than indicating the valence (type or direction) of emotion (i.e. positive or negative) experienced by the user at the time of data collection. From a designer's perspective, in order to evaluate participant reactions to an environment in a meaningful way it will be insufficient to use only the GSR data because of the difficulty in distinguishing between negative or positive emotions based on arousal level alone. Consequently, the design of the VEMap had to incorporate a way of capturing a participant's self-report of emotional state at any given time during the walk-through of a VE. In summary, VEMap combines both objective measurement (GSR testing) and subjective evaluation (self-report) in order to provide designers with a more comprehensive picture of a participant's emotional responses to a virtual design.

3.3 VEMap: Proof of Concept

Nold (2004) conceptualized Bio Mapping in the physical world as:

Bio Mapping = Arousal Level + Physical Location,

Where arousal level is measured by GSR; and physical location is measured by GPS coordinates.

Using this as a starting point, the VEMap concept is conceptualized as:

VEMap = Arousal Level + Emotion Valence + Virtual Location

Where arousal level is measured by GSR; emotional valence is measured by self-report ratings; virtual location is measured by VE coordinates in x- and y-space.

There are some notable differences between the approach to mapping emotional responses used in the new design for VEMap and Nold's Bio Mapping technique:

- VEMap prompts the user to report general emotional state (i.e. positive, neutral, or negative) at key points in the VE. Nold's Bio Mapping relies on GSR data to determine arousal level and does not specifically ask participants about their perceived emotional state while GSR data is being collected.

- VEMap emotional state or valence (positive, neutral, negative) is represented by three (3) different dot colours. Participants can select which colour(s) should represent positive, neutral, and negative emotions on their VEMap. Nold's Bio Mapping only uses two colours of dots to indicate high and low levels of arousal. A decision was made to applying only three categories of emotions instead of five (more positive, positive, neutral, negative, more negative) or more as sometimes making a choice between positive and negative is much easier than between more negative and positive. As this is a new technique it was decided to start beta testing with the basic three colour approach first.

- VEMap presents level of arousal using gradients of a colour reflecting up to seven (7) levels of arousal based on equal sectioning of the differential between a participant's MAX GSR and MIN GSR (which is assumed to be the participant's low or neutral

level of arousal in order to decrease the individual differences of GSR readings). Nold's Bio Mapping uses two levels of arousal for each participant – high and low based on the mid-point of the participant's total GSR readings.

- VEMap location is calculated based on mapping the appropriate 3D space coordinates from the VE. Nold's Bio Mapping used GPS data to identify the participant's location in the real-world 3D space

Illustrations of Nold's Bio Mapping were shown in Figures 5 and 6. An example of a VEMap is shown in Figure 8. The VEMap is designed to present a detailed picture of the user's emotional experience in a specific VE.
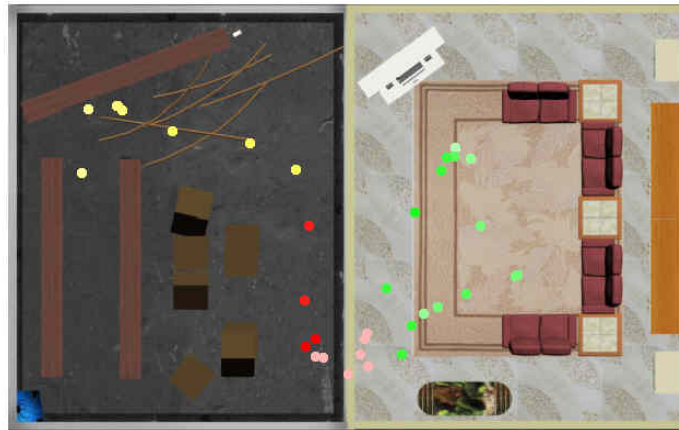
Figure 8: An example of VEMap where green dots = positive emotion; yellow dots = neutral emotion; and red dots = negative emotion. The intensity of dot colour corresponds to the intensity of arousal level generated within the participant at that location.

Figure 8 shows a "bird's-eye", or overhead map, view of a VE used for beta testing of the VEMap system. The overhead map view is used so that the designer can view the participant's emotional responses that occur along the route that the participant took through the VE. Figure 9 (a) (next page) shows the first-person perspective that the participant sees when travelling through the 3D virtual world. The "display" that appears in the centre of Figure 9 (b) is the "pop-up" screen that appears at set locations throughout the environment to cue the participant to provide a self-report on his or her emotional state at that moment.

There are three main components to the VEMap software: a) the Development Platform ─ Virtual Navigation and Collaboration Experimentation Platform(VNCEP) ─ a specialized VE development platform created for DRDC-Toronto, b)3D models created using *3ds Max®*); and then imported into VNCEP , c) and the VEMap (created by combining *Excel®* and *Surfer®* (Geography Information System Mapping Software)).

(a)   (b)

Figure 9: a) First-person perspective view while travelling through 3D virtual world;
b) Pop-up screen asking for user's subjective feedback

3.3.1 VEMap Development Platform and Hardware

The virtual worlds for VEMap are generated using *Virtual Navigation and Collaboration Experimentation Platform* (VNCEP) provided by DRDC-Toronto. VNCEP runs on a Pentium™ 4 PC desktop computer with a 3.4 GHz processor and 3 GB of RAM. The computer uses Windows™ XP operating system with a Quadro™ FX 3450/4000 SDI (256 MB) from NVIDIA™.

Objects to be placed in a VE are modeled using *3ds Max®,* and object behaviours are controlled using the programming language Lua. For example, since it is well known that participants can easily become disoriented in VEs, a set of 3D markers ─ rainbow coloured balls which can be distinguished easily from the surrounding environment, were created to help participants navigate (See Figure 10). From a usability perspective, the appearance of a 3D marker helps to automatically direct the participant's attention towards the next part of the route through the VE. The inclusion of markers was done to help reduce the frustration (and consequent negative emotion) that participants might experience if they become lost in the VE.
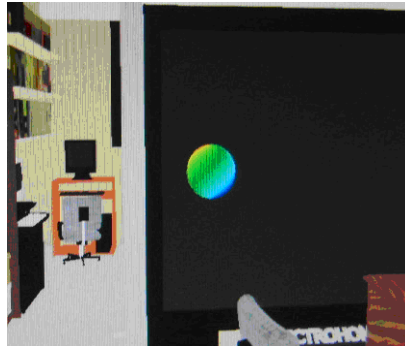
Figure 10: One sample of 3D markers for assisting navigation in VE.

The parameters for the behaviours of the markers were set so that when the participant comes in close proximity of the virtual marker, a popup screen is triggered. The pop-up screen asks the participant to report current emotional state at that time (Figure 9 (b)). The number of markers (and consequently pop-up screens) included in any given VE is up to the designer or researcher.

The VE images are projected onto an 81" Fakespace® ImmersaDesk display with 1280 by 1024 resolution at a 75Hz refresh rate. Non-stereoscopic images of 3D objects are used to reduce possibility of simulator sickness (See Figure 11).

Figure 11: 81" CRT display and equipment set-up used for experimental evaluation of VEMap

Movement through the environment uses keyboard controls similar to the controls used in popular first-person perspective computer games. Participants use the "W", "A", S", and "D" keys on the keyboard to navigate through the environment. The participant's "walking speed" is set to simulate walking through an environment at a pace equivalent to 1.5 m/s.

Figure 12 presents an overview of the information flow within the VEMap application.
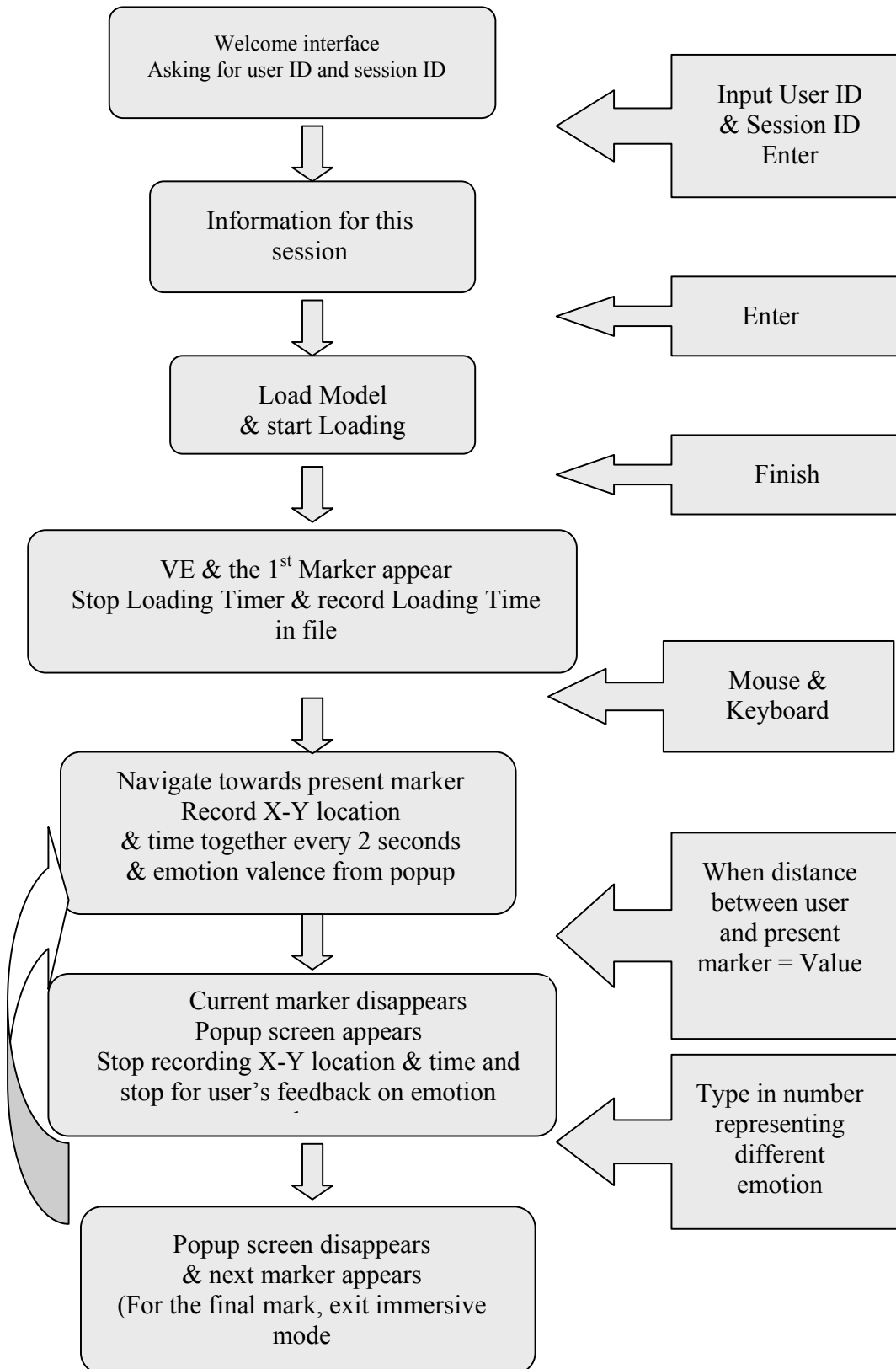
Figure 12: Flow of information within VEMap

3.3.2 Collecting GSR Data

GSR data are collected using a Biomonitor ME6000 designed and manufactured by Mega Electronics (See Figure 13). GSR signals are sampled 1000 times per second and recorded using *MegaWin® 3.0*. Once the data collection for a participant was completed within a particular VE, the GSR data was grouped in 2 second intervals in order to get appropriate number of GSR readings. And the RMS (Root Mean Square) average for that interval was used as the GSR reading for the X-Y location within the VE associated with that particular time. This transformation has been computed by *MegaWin® 3.0* as well.

$$\textbf{RMS Average [i]}_{\text{ in 2 seconds interval}} = \sqrt{\frac{\sum\limits_{i=n}^{n+2000-1} |Data_{Raw}[i]|^2}{2000}} \qquad [1]$$

Figure 13: ME 6000 Biomonitor and its sensors for collecting GSR data

In order to collect the GSR data, two Velcro-wrapped sensors are connected to the middle part of the thumb, and the medial phalange of the fifth finger of the participant's left hand. The Velcro-wraps were checked to make sure that the sensors did not interfere with fingers movement while controlling the keyboard (Figure 14). It is important that all connecting wires be adjusted so that the user feels comfortable as any feelings of discomfort are likely to have a negative impact on the GSR readings.

Figure 14: Left hand for controlling keyboard, collecting GSR data; right hand for controlling mouse.

According to Boucsein (1992), the sites most recommended for measuring GSR are the medial phalanges of the index and middle fingers. However, due to the need for participants to operate a keyboard with their left hand, the middle part of the thumb, and medial phalange of fifth finger of the left hand are used for collecting GSR data in the evaluation procedure of VEMap. The movements of fingers can influence the accuracy of GSR measurement. Following a pre-investigation study done among the Use-It Lab group members before the evaluation procedure of VEMap, it made sense to use the thumb and fifth finger for collecting GSR data as they were the digits used the least for operating the keyboard.

The characteristics of GSR are such that each person has his/her own skin conductance level. As a result, it is necessary to transform the raw GSR data to reflect the change (delta) in GSR output compared to the individual's SCL (or minimum GSR).

The basic transformation of GSR data used in VEMap is:

$$\text{VEMap Arousal Level} = \text{GSR}_{recorded} - \text{GSR}_{minimum} \qquad [2]$$

3.3.3 VEMap Mapping

Colour-based mapping of emotional response combines the GSR data collected by the ME6000, participant self-reports of emotional states, and the X-Y location data recorded by the VEMap software.

In order to transfer GSR data into coloured dots, the classification of a GSR value into corresponding colour with a particular RGB value was adopted for the mapping process. Participants and/or researchers may choose among seven pure colours to represent emotional states. These colours include red, green, blue, black, yellow, purple, cyan (see Table 1). Each of these colours has one or two values among three RGB values set to 255, with the rest being 0. For additional colours, the method keeps the values of 255 constant, and then sets the rest of the values from 0 to 180 with an interval of 30. As a result, there are seven distinct colours which can be used for indicating emotional valences.

Table 1: Seven colours used in VEMap (R-red, G-green, B-blue)

Red	(R-255, G-0, B-0)	
Green	(R-0, G-255, B-0)	
Blue	(R-0, G-0, B-255)	
Black	(R-0, G-0, B-0)	
Yellow	(R-255, G-255, B-0)	
Purple	(R-255, G-0, B-255)	
Cyan	(R-0, G-255, B-255)	

To be applicable across cultures, VEMap colour selections must be customizable by users. As mentioned, one of the features of the VEMap is that users can choose three colours to represent three kinds of emotions (i.e. positive, negative and neutral). Allowing participants to have the ability to choose the colours to represent emotional state accommodates for cultural effects of colour perception. For example, according to Chinese culture and tradition, red represents positive emotions, green for negative emotions and yellow for neutral emotions. Red is the most popular holiday colour in Chinese culture and is associated with happiness and prosperity, especially in the Chinese New Year. By contrast, in North American culture, red signals "warning" or "danger" and elicits more negative emotions than positive ones. The consequences for cultural variation in perception of colour affect the broader applications of VEMap. Additionally, personal preferences may also have an effect on colour interpretation. As a result, it could be easier for designers to understand their

clients' emotional responses if the designers themselves can select the colours to use, so that the resulting VEMap corresponds to their personal colour-choices to represent emotions.

Once the three emotional valence colours have been selected, each of those colours can be presented at seven different levels of intensity to represent level of arousal within a given emotional state. For each participant, GSR data are separated into seven arousal levels based on the individual's maximum and minimum GSR readings which are used to set the range for arousal. A colour intensity is then associated with each of the seven arousal levels. For example, the subgroup with the lowest GSR value is associated with the lightest colour. An example of this classification method is shown in Table 2. For the purposes of illustration only three colours are used. If a participant chose red to represent negative emotion, and yellow to represent neutral, and green to represent positive; and her/his GSR data ranges from 10,000.1 nS (nanosiemens) to 24,000.0 nS, he/she chooses red as negative emotion, yellow as neutral and green as positive emotion. The GSR data can be classified into 7 groups, each corresponding to a different colour intensity corresponding to increasing/decreasing arousal levels.

Table 2: An example of GSR classification and relative colour classification

Red Negative							
RGB of Red	255, 0, 0	255, 30,30	255, 60,60	255, 90,90	255,120,120	255, 150, 150	255, 180,180
Yellow Neutral							
RGB of Yellow	255, 255,0	255,255,30	255, 255,60	255, 255,90	255, 255, 120	255, 255, 150	255, 255,180
Green Positive							
RGB of Green	0, 255,0	30,255,30	60, 255,60	90, 255,90	120, 255, 120	150, 255, 150	180, 255,180
GSR(nS)	22000.1-24000	20000.1-22000	18000.1-20000	16000.1-18000	14000.1-16000	12000.1-14000	10000.1-12000
GSR Range 14000(nS)	12000-13999.9	10000-11999.9	8000-9999.9	6000-7999.9	4000-5999.9	2000-3999.9	0-1999.9

After data transformation and classification, colour dots with information about emotional responses can be mapped. Because GSR data and X-Y location data are recorded together with the same time parameter, it is convenient to connect the GSR and corresponding X-Y location of users' position data together if the time parameters are the same. As a result, the transformed GSR data – the coloured dots - can be mapped onto the base map of the VE along the user's route according to the X-Y location data.

4 Beta-Testing and Evaluation of VEMap

The purpose of the beta-testing of VEMap was to do a first-pass assessment of the face validity of the VEMap as a design tool for architecture and urban planning; as well as to carry out a design review of the user-interface and the procedures for generating emotional response maps in order to provide recommendations for further development and refinement of emotional mapping within the VEMap application. The beta-testing sessions were held in the Use-It Lab of University of Waterloo during July and August of 2008.

4.1 Participants

Fourteen participants were recruited to take part in the VEMap beta-testing sessions. There were 10 males and 4 females. All participants were undergraduate or graduate students of the University of Waterloo, age from 18 to 28 years old. Thirteen participants were right handed and only one was left handed.

In order to understand how the computer experience, especially with VE technology might influence emotional responses in VEs participants were classified into 4 groups based on their experience using VE technology: Very Frequent (more than once a week), Frequent (more than once a month), Occasional (less than once a month), and Never.

One (male) was classified into the Very Frequent group (more than once a week); four (one female and three males) in the Frequent group (more than once a month); seven (one female and six males) in the Occasional group (less than once a month); and two females in the Never group. Because of the unbalanced number of participants and the gender biases (there were no females in the Very Frequent group, and no males in the Never group), the Very Frequent group and the Frequent group were combined together as VF&FR group, the Occasional group and the Never group were combined together as OC&NE group for analysis purposes. Under this situation, there were five participants (one female and four male) in VF&FR (Very Frequent & Frequent) group, and nine participants (three female and six male) in OC&NE (Occasional & Never) group.

4.2 Equipment Setup

The main VEMap equipment used for the beta testing sessions is described in detail in Section 3.3 (VEMap: Proof of Concept). In addition, a digital video recorder was set up to follow what was happening on the computer screen while simultaneously recording the voiced comments of the participant, as all participants were encouraged to "think aloud" about what they were experiencing as a means of helping to further understand their basic emotional valence ratings entered on the pop-up screens.

4.3 Virtual Test Environments

In order to evaluate whether VEMap could represent users' emotional responses including emotional valence and emotional arousal level, three VEs were created to represent architectural and urban planning examples. They were: 1) a simple meeting room and storage room for the training session; 2) a simple virtual city with 9 blocks as an outdoor session to represent a hypothetical application in urban planning; and 3) a virtual lab based on the actual research lab at the University of Waterloo in which the VEMap evaluations were actually being carried out as this represented both an indoor session and a hypothetical application in architecture design related to interior decoration. All of the 3D models included in the VEs were made by using *3ds Max®.*

4.3.1 Training VE

Figure 15 shows the base map of the training rooms. The right panel illustrates a tidy and clean meeting room with bright and warm colours, simple decorations and furnishings, including sofas, coffee tables, cabinets, a fish tank and a TV set. The expectation is that this room will evoke positive emotions. The left panel is a dark and messy storage room, with dark grey and brown colours, random boxes on the floor, broken bookshelves and rusty metal pipes. The expectation here is that, negative emotions will be evoked in this room. The main goal of creating this model is for training participants in order to make them familiar with all the input and output devices for navigating through the VE, responding to the self-report screens, in order to get familiar with the feeling of having the GSR sensors attached to their fingers. In addition, it allowed the researcher and participant to determine if the participant

was experiencing any simulator sickness as a result of moving through the VE. This is important as such adverse responses to the VE equipment could have a confounding impact on the GSR data. It should be noted that some confusion in participants regarding navigating and responding was very common during the training procedure of experimental evaluation period. Therefore, the training session was helpful in familiarizing participants with the VEMap procedures.

Figure 15: Base map of training rooms

4.3.2 Virtual Outdoor VE

Figure 16 shows a birds-eye view of the outdoor VE. There are nine blocks in this outdoor model. There are residential area, a park, office buildings, and a construction site. This model could be seen as a simplified urban plan example. A mock urban planning evaluation was taken in this session. Participants were asked to imagine that they were potential residents of the urban area and should give feedback about each area. The reason for having blocks of varying urban functions is for understanding whether VEs with distinguishable characteristics can evoke different feelings in users, and whether these feelings are similar to what one might expect in the real world.
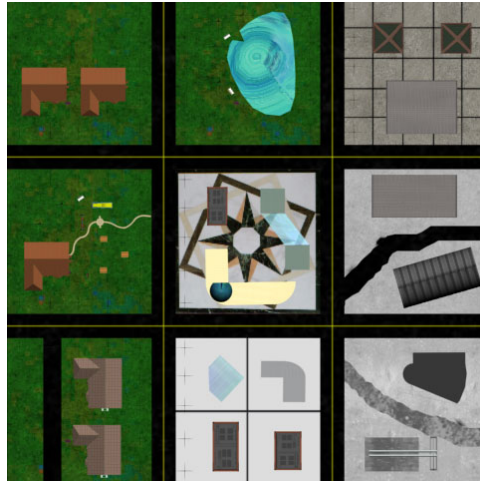
Figure 16: Base map of outdoor environment.

4.3.3 Virtual Indoor VE

Figure 17 shows the virtual research lab. This model is made to replicate the Use-It Lab and part of the corridor on the same floor located on the third floor of the Engineer 2 building at the University of Waterloo. Most parts of the virtual corridor and lab are similar to the real ones. For example, the colours of the doors along the corridor, size and colours of couch and the 81" Fakespace® ImmersaDesk CRT display in the lab were modeled according to the real objects. In order to simplify the model, small details such as some posters on the boards hanging on the walls of the corridor and tools on the tables which might not be noticeable were not included. The goal of this model is for determining the degree of similarity participants can feel between real world and corresponding VEs. If this degree of similarity is high, then it makes sense to use VE technology for evaluating architecture designs and urban plans.
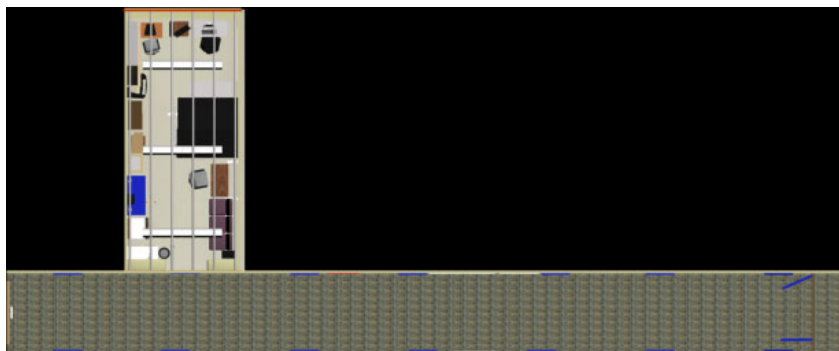
Figure 17: Base map of indoor environment (virtual Use-It Lab)

To ensure that participants can navigate through and see all the parts of each model, there is a pre-set route in each model by rainbow coloured marks (Figure 10). Because it would take long to explore the whole outdoor environment, the two pre-set routes of outdoor environment would lead participants going through some parts including all the styles. Figure 18 shows each model with setup route.

(a)   (b)

(c)   (d)
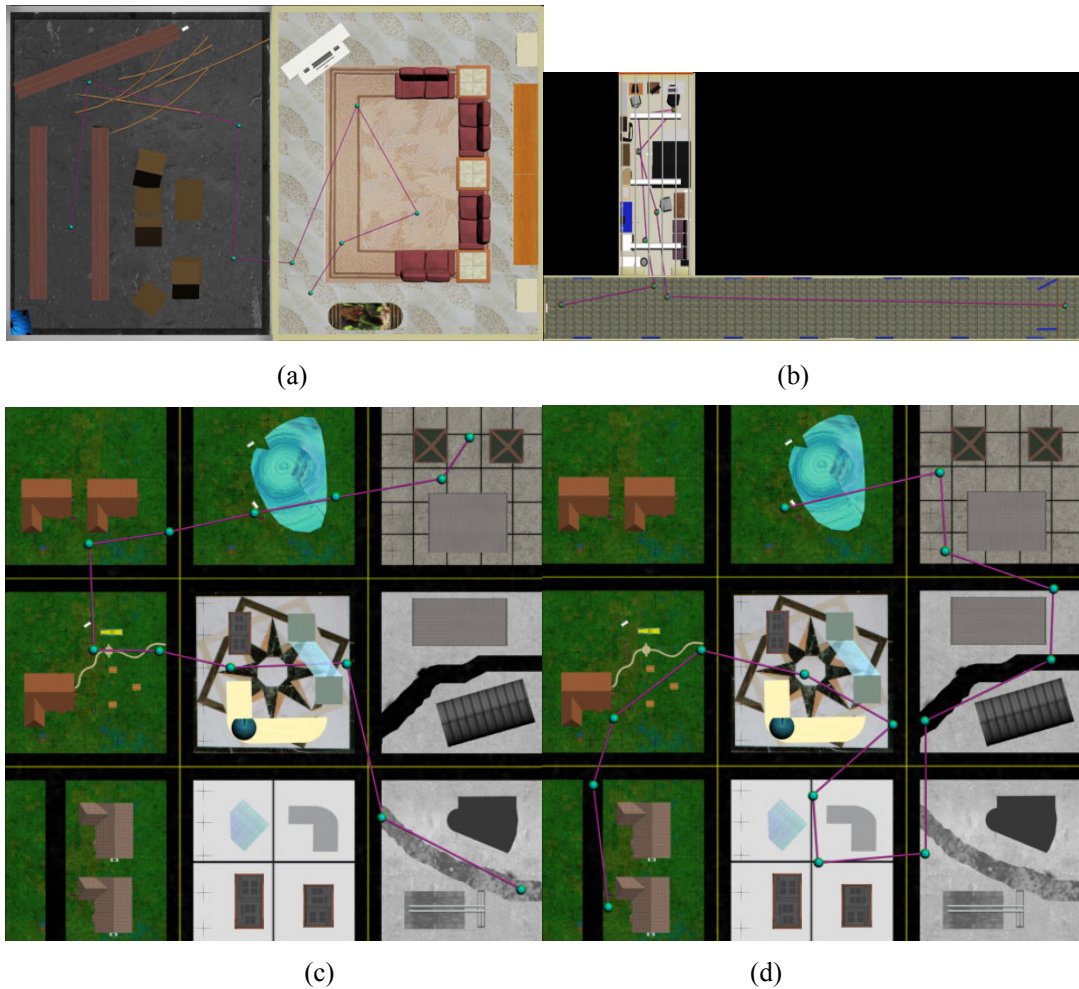
Figure 18: a) Setup route of train session; b) Setup route of indoor session; c) Setup route 1 of outdoor session; d) Setup route 2 of outdoor session

In summary, models can vary according to designers' requirements. The three models described here were designed for the purpose of assessing the efficacy of the VEMap.

4.4 Procedures

The 25 minutes duration of the entire beta-testing session included the following procedures:

a). Participants were provided with the Study Information Sheet (Appendix A) and Informed Consent Letter (Appendix B) to sign before the experimental procedure.

b). Participants were asked to complete a demographics questionnaire (Appendix C, Questions 1 and 2) for personal information such as age, gender, and VE technology experience. This introductory period lasted approximately 15 minutes, and also served to bring the user into a calm state as recommended by Boucsein (1992). The desired result is to minimize the influence of the user's pre-study activities on the GSR data.

c). Two GSR sensors embedded in a Velcro strap were wrapped to the thumb and fifth finger of left hand. This procedure was the same for that left-handed dominate participant.

d). To gain familiarity with the experimental interface, participants were asked to complete a training session. Each participant started the session near the door in the meeting room facing a painting on the wall and with the first marker in view (Figure 19). The participant needed to operate the keyboard and mouse for moving and turning around, and "walk" virtually towards the only present marker. When the participant was close enough to the marker, a popup screen would appear, allowing the user to give a response of feeling either a neutral (0), positive (1) or negative (2) emotion. After typing in the corresponding number and the ENTER key, the participant could go on navigating through other parts of the environment until reaching the final marker. All the controls in the three sessions were the same. There were a total of 8 markers in the training session (Table 3).

Figure 19: Training session from user's point of view

Upon completing the training session, participants began the experimental trials. There were two experimental trials consisting of an indoor and outdoor environment, in a random order. For the outdoor session, there were two routes and four start (end) points. Participants were randomly assigned to one start point of one route. In the indoor session, all the participants startd at the end of the corridor (furthest from the lab) and moved towards the virtual Use-It Lab where a marker was placed at the door (Figure 20). Then, once again by following markers, participants would explore the area and exit after finding the final marker. There were also 8 markers in this session (Table 3).

Figure 20: Indoor session from user's point of view

There was a slight modification to the allowed routes in the outdoor session (Figure 21). For reducing the time spent on exploring the whole complicated outdoor VE, and for randomization purposes, there were 2 routes and 4 starting points (4 end points). All 14 participants were randomly assigned to one route and one start point. Four of them were in route 1 starting point 1, four were in route 1 starting point 2, three were in route 2 starting point 3, and three were in route 2 starting point 4. The difference from the previous two sessions was that this session had 11 markers for route 1 and 14 markers for route 2 (Table 3).
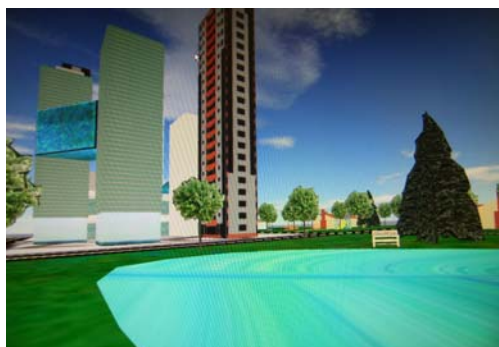
Figure 21: Outdoor session from the user's point of view

e). Participants were asked to follow paths through the experimental environments while providing verbal feedback by the "think aloud" method, which was captured by a digital video recorder positioned behind the participant.

f). At pre-selected points indicated by markers (14 points in route 1; 11 points in route 2 for the outdoor environment, 8 points for the indoor environment) along the path, the participants were asked to classify their emotional valence as positive, neutral or negative on a popup screen.

g). Upon arrival at the finish point, participants answered a question about their emotional responses experienced while exploring the environments.

h). Upon finishing all the experimental trials, participants finished the last two questions on the questionnaire (Appendix C, Questions 3 and 4) about the similarity rating between Virtual Lab and real representation and their comments.

i). At the end, participants were provided with a feedback letter (Appendix D).

Table 3: Number of marks for each route

Session		Number of Points
Training		8
Outdoor	Route 1	11
	Route 2	14
Indoor		8

5 Results

The results of the beta-testing include: participants' ratings of perceived similarity between the Virtual Use-IT Lab and the Real Use-IT Lab; GSR data recorded by the ME 6000 and *MegaWin®* which has been transformed for mapping, and corresponding VEMap samples which displayed the primary output of the beta-testing; and the matching of the participant's VEMap with the participant's verbal feedback recorded for considering the face validity of the generated VEMap. In addition, some preliminary analyses of factors contributing to exploratory performance such as gender and VE experience effects are presented.

5.1 Similarity of VE Lab and Real Lab

The most important step in validating VEMap as a useful tool is to determine if emotional responses captured within a VE match emotional responses experienced in the real environment. In this case, the experiences users have obtained from both VE and real world can be transferred to their counterpart environments. As part of the beta-testing of the VEMap application, participants were asked to indicate the similarity between their experiences in the real Use-IT lab and the virtual representation of the lab. Participant's response was captured using a 5-point Likert Scale, where "1" represents completely different and "5" represents completely the same (Appendix C). All fourteen participants perceived the virtual representation to be somewhat like the real-world. Seven participants chose a rating of "4", two chose "3", and five participants chose "2". Based on the questionnaires, the mean similarity score is 3.14 out of 5 (Std. Deviation = 0.95); which is significantly different from what one might randomly expect by chance (χ^2=13.86, p<0.01).

5.2 Transformation of GSR Data to Create VEMap

The overall GSR data of the 14 participants were processed by the transformation method mentioned in 3.3.2 Collecting GSR Data. The overall averaged RMS GSR data for each participant are shown in Table 4, including the minimum GSR value, maximum GSR value, and the range of GSR values for each session, which are used for coding the GSR data into coloured dots.

Table 4: Overall averaged RMS GSR data for each participant

Participant	Minimal GSR value (nS)			Maximal GSR value (nS)			Difference (nS)		
	Training	Outdoor	Indoor	Training	Outdoor	Indoor	Training	Outdoor	Indoor
1	30042	21434	21033	32602	32663	30620	2560	11229	9587
2	20926	23635	16062	32459	32645	32575	11534	9010	16513
3	19913	1556	11268	32473	28353	29539	12560	26796	18271
4	8061	16782	13180	12885	20748	16303	4824	3966	3123
5	7197	676	1164	22741	15325	17667	15544	14649	16503
6	29118	18735	24777	32494	30533	32565	3377	11798	7788
7	11976	18359	18702	22575	32301	27238	10599	13942	8535
8	18307	19485	17698	19415	22849	21287	1108	3364	3589
9	16888	13494	22783	30186	32405	32412	13298	18911	9630
10	6306	6981	8790	12874	7741	13746	6568	760	4956
11	5866	5925	6354	6011	6253	6657	145	329	303
12	26849	18783	16686	32637	22380	20392	5788	3598	3706
13	19569	21080	16665	32660	32202	29627	13091	11122	12962
14	21195	801	11575	26877	17810	20842	5682	17009	9267

Table 4 shows that the individual GSR data across the 14 participants are quite varied. For example, one participant's GSR data was from 5866 nS to 6011 nS with a range of 145 nS; but another participant's was from 1556 nS to 28353 nS with a range of approximately over 20000 nS. In order to simplify the mapping process, change in arousal level was calculated by subtracting the minimum GSR for the session from the sampled GSR, so that relative GSR data ranged from 0nS to a maximum level for that participant (See Section 3.3.2). Then the adjusted GSR data, ranging from 0nS to the participant's maximum level was separated into seven groups with equal intervals associated with the seven colour intensities.

Not only did the GSR range differ across the 14 participants, the change in arousal level differed while the participants travelled through the environments. For example, in the first block of the indoor session, 5 participants had increasing GSR; 1 had the opposite – decreasing data; and 8 of them had fluctuating data. Given what participants were recorded to have said while they were travelling through the VE, it appears that the reason for this behaviour is that participants experienced different emotional responses to the same location. Furthermore, even though participants may express similar emotions, there are differences in the GSR data between participants. Therefore, transforming GSR data independently was necessary because participant experienced within each VE is independent from one another;

and the arousal level of each participant is independent of the arousal level of any of the other participants. In other words, VEMap is a personal emotion map of one user that expressed his or her emotional responses to a VE in real-time.

In Figure 22, there are two samples of VEMap from the training session. In Figure 22a, the participant started with positive emotion but began to feel more negative emotion especially while facing the TV set. According to the verbal feedback from the participant, the colour of TV set was too white for her. Although this uncomfortable feeling was weaker when she turned back, it became stronger again when she entered the storage room. This behaviour can be explained by the perceived difficulty in passing through a narrow doorway. On the contrary, in Figure 22b the next participant always had positive emotion while in the meeting room, and even was excited about moving towards the fish tank, which was the favourite piece of furniture based on verbal comments. Similar to the participant's response in Figure 22a, this participant's emotional status changed upon entering the storage room. The participant expressed being uncomfortable due to the darkness and clutter in this environment. These two samples illustrate the similarities and differences in emotional response of the participants to the same environment.

(a)   (b)

Figure 22: Two samples of VEMap generated from the training session

Figure 23 displays an overall composite map for all 14 participants which expressed the common feelings for the training environment. From this map, the emotions in the meeting room appear to be more positive compared to those for the storage room. Most of the participants had higher arousal level at the entry to the meeting room, and especially in front

of the TV set. On the contrary, almost every one had strong negative emotions while crossing the door to the storage room, which according to their verbal descriptions they felt was dark and messy with a narrow entrance.

Figure 23: A composite VEMap from the training session generated from the data for all 14 participants

5.3 VEMap Compared to Verbal Protocols

During the beta-testing of the VEMap application, participants were required to follow a think-aloud protocol. In this way, participants' action and selection could be better understood for comparative analysis with the GSR data. By aid of a digital video recording of the computer display shot from the place participants were going visiting and their verbal feedback could be recorded at the same time. In this way, any comments or verbal feedback made by the participant, which could indicate emotional responses to the VE were captured. For example, some participants would point out their favourite features in the environment like "Oh, a Pool!" in the outdoor environment. From the audio recording of the trials, the comments were transcribed onto a map of the corresponding environment with the participant's path illustrated as shown in Figure 24. Every block of each route was marked as: No Comments, Comment and Match, or Comment and Non-Match by comparing the verbal feedback and corresponding VEMap data.
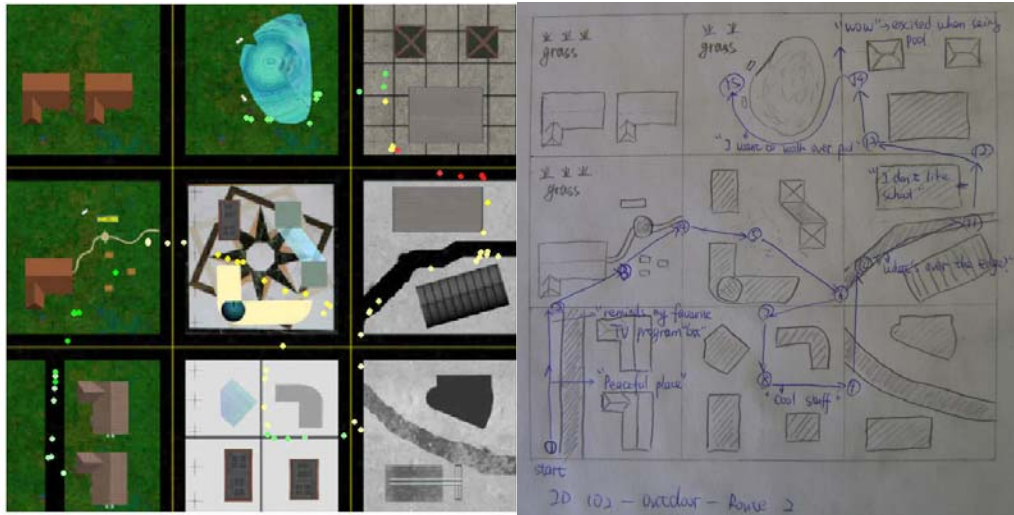
Figure 24: Comparison between VEMap and verbal feedback translation

In total, there were 366 points across the three sessions and 14 participants, where 8 points of training session for 13 participants; 11 points of outdoor session route 1 for 8 participants; 14 points of outdoor session route 2 for 5 participants; and 8 points of indoor session for 13 participants (one participant didn't provide verbal feedback during the whole procedure). According to the comparison between VEMap and corresponding verbal feedback, there were 129 "Comment and Match", 17 "Comment and Non-Match", and 220 "No Comments" ($\chi^2(1) = 85.92$, P < 0.05).

5.4 Exploratory Analysis

Participants seemed to have no trouble manipulating keyboard and mouse for navigating and providing feedback on the popup screens no matter which hand was dominant. The performance of participants was much better after the training session. The guide markers in the VEs appear to have been helpful in navigation as no participant became lost in the VEs during the trials.

Since navigation and decision making are completely different performances, it is necessary to analyze those data separately. For this reason, not only was overall time analyzed, the time parameter was also separated into two groups: one is the time participants spent on

navigating through the VE; the other one is the time participants spent on making decisions while the popup screen was showing.

5.4.1 Overall Time

The time participants spent on each session ranged from 99.9 seconds to 422.1 seconds. Table 5 shows the overall time spent on each session averaged over the 14 participants.

Table 5: Time data summary for each session

	Minimum (s)	Maximum (s)	Mean (s)	Std. (s)
Training	119.7	370.7	191.2	68.8
Outdoor	122.8	422.1	257.3	98.0
Indoor	99.9	235.2	149.3	40.6

This overall time can be broken down into two groups, navigation, and decision-making. Navigation time encapsulates all the time spent on exploring the environment (Table 6). Navigation time ranged from 74.4 seconds to 371.9 seconds displayed in Table 6. Decision-making time encapsulates all the time while popup screen was displayed to capture the participant's time to identify emotional valence for the previous VE area (Table 7). Decision-making time ranged from 8.9 seconds to 210.9 seconds, which was shown in Table 7.

Table 6: Summary time spent on navigating through VE

	Minimum (s)	Maximum (s)	Mean (s)	Std. (s)
Training	74.4	285.7	135.5	54.1
Outdoor	100.5	371.9	212.6	79.9
Indoor	88.3	188.8	117.1	28.4

Table 7: Summary time spent on decision making

	Minimum (s)	Maximum (s)	Mean (s)	Std. (s)
Training	15.0	101.0	55.8	23.2
Outdoor	13.3	210.9	44.9	49.7
Indoor	8.9	106.7	32.2	23.5

Table 5 - 7 indicate large individual differences between participants for the minimal and maximal values. For example, the time spent on making a decision for the indoor session

ranged from 8.9 seconds to 106.7 seconds. Potential sources of the variability can be due to gender and experience with VEs.

5.4.2 Gender

Keeping in mind that there were not an equal number of males and females for the beta-testing, the followed analyses are considered to be exploratory only. On average males spent 42.1 seconds less than that spent by females without consideration of session difference (Table 8). From the statistical analysis (Appendix E), there appears to be a slight gender effect, but the evidence is not strong enough to be considered significantly to support gender effect (t-test (40) = -1.486, Df = 40, p = 0.145 > 0.05).

Table 8: Summary of overall time spent by male and female participants

Gender	N	Mean (s)	Std. Deviation (s)
Male	10	187.2	77.4
Female	4	229.3	96.2

Across the three VEs, the average time spent by females is more than males (Table 9). This trend happened for all the three sessions. However, similar to the finding for overall time without session influences, gender effect was not found to be significant, where $F_{train}(1) = 0.801$ (p > 0.05); $F_{outdoor}(1) = 0.675$ (p > 0.05); $F_{indoor}(1) = 3.576$ (p > 0.05) (Appendix E & Table 9).

Table 9: Summary of overall time spent by male and female participants separated by sessions

	Gender	N	Mean (s)	Std. Deviation (s)	F	Sig.
Training	Male	10	180.7	73.1	0.801	0.388
	Female	4	217.4	56.6		
Outdoor	Male	10	243.5	82.1	0.675	0.427
	Female	4	291.7	138.4		
Indoor	Male	10	137.5	29.9	3.576	0.083
	Female	4	178.9	53.0		

5.4.2.1 Gender by Navigation Time

From the statistical analysis in Table 10 about gender effect without consideration of sessions effects on navigation time, although the average time spent by females is higher than that spent by males, the difference is not significant (t-test(40) = -0.506, Df = 40, p = 0.616 > 0.05) (Appendix E).

Table 10: Summary of navigation time spent by male and female
without consideration of session influence

Gender	N	Mean (s)	Std. Deviation (s)
Male	30	151.5	69.0
Female	12	163.8	76.3

From Table 11, the same result was obtained while taking the separate VE sessions into account. While females on average took more time to navigate through each VE than males, these differences were not found to be significant ($F_{train}(1) = 0.156$ (p > 0.05); $F_{outdoor}(1) = 0.021$ (p > 0.05); $F_{indoor}(1) = 0.991$ (p > 0.05)) (See Table 11 & Appendix E).

Table 11: Summary of navigation time spent by male and female participants separated by sessions

	Gender	N	Mean (s)	Std. Deviation (s)	F	Sig.
Training	Male	10	131.7	59.4	0.156	0.700
	Female	4	144.8	43.8		
Outdoor	Male	10	210.6	72.9	0.021	0.888
	Female	4	217.6	108.2		
Indoor	Male	10	112.3	22.0	0.991	0.339
	Female	4	129.0	42.2		

5.4.2.2 Gender by Emotional Response Time

The statistical analysis of popup time without consideration of session influence is opposite to that of navigation time. As shown in Table 12, females spent significantly more time than males in reporting their emotional valences (t-test (40) = -2.680, p = 0.011 < 0.05) (Appendix E). Females took longer (mean = 65.5 seconds, Std. Deviation = 55.18) compared to male (mean = 35.7 seconds, Std. Deviation = 17.65) in evaluating their emotion responses. This is an interesting finding given that the unbalanced numbers would make it more difficult to statistically detect significant differences between males and females.

Table 12: Summary of decision making time spent by male and female
without consideration of session influence

Gender	N	Mean (s)	Std. Deviation (s)
Male	10	35.7	17.7
Female	4	65.5	55.2

The findings of the statistical analyses of each session about emotional decision making time were similar to those for navigation time in that the average time spent by females was higher but not statistically significant ($F_{train}(1) = 3.547$ (p > 0.05); $F_{outdoor}(1) = 2.132$ (p > 0.05); $F_{indoor}(1) = 3.844$ (p > 0.05))(See Table 13).

Table 13: Decision making time spent by male and female separated by sessions

	Gender	N	Mean (s)	Std. Deviation (s)	F	Sig.
Training	Male	10	48.9	19.1	3.547	0.084
	Female	4	72.6	26.5		
Outdoor	Male	10	32.9	13.9	2.132	0.170
	Female	4	74.1	92.3		
Indoor	Male	10	25.2	10.9	3.844	0.074
	Female	4	49.9	38.2		

In summary, females appear to take longer to make a choice among positive, negative or neutral emotions than males. More experimental trials with balanced number of male and female participants are needed to further investigate this finding.

5.4.3 VE Experiences

There is another important factor that may influence performance time － VE experiences. Based on the information from the questionnaires the 14 participants were classified into two groups, which were VF&FR and OC&NE (as shown in session 4.1 Participants)

From Table 14, participants with less VE experience spent significantly more overall time than those with more experience (t-test (40) = 2.542, p = 0.015 < 0.05)) (Appendix F).

Table 14: Summary of overall time spent by participants
classified by VE experiences

VR Exp.	N	Mean (s)	Std. Deviation (s)
OC&NE	9	222.4	92.4
VF&FR	5	157.6	44.9

Table 15 － overall time analysis by each session shows that participants in VF&FR spent less time than those in OC&NE group. However, the difference was not significant for the outdoor session, and not for the training or indoor VE session ($F_{train}(1)$ = 2.104 (p > 0.05); $F_{outdoor}(1)$ = 5.148 (p < 0.05); $F_{indoor}(1)$ = 2.319 (p > 0.05)) (Appendix F & Table 15).

Table 15: Overall time spent by participants classified by VE experiences

	VRExp.	N	Mean (s)	Std. Deviation (s)	F	Sig.
Training	OC&NE	9	210.2	76.4	2.104	0.173
	VF&FR	5	156.8	37.7		
Outdoor	OC&NE	9	295.8	98.4	5.148	0.043
	VF&FR	5	187.9	49.8		
Indoor	OC&NE	9	161.0	42.2	2.319	0.154
	VF&FR	5	128.2	30.3		

5.4.3.1 VE Experience by Navigation Time

Table 16 shows that participants with less VE experience spent 50 seconds more than those with more experience (t-test(40) = 2.487, p = 0.017 < 0.05) (Appendix F).

Table 16: Summary of navigation time spent by participants classified by VE experiences

VR Exp.	N	Mean (s)	Std. Deviation (s)
OC&NE	9	174.0	77.0
VF&FR	5	120.9	39.8

In Table 17, the participants in the VF&FR group seemed to spend less time than those in the OC&NE group in each session on navigation time, similar to the overall time comparison. The VE experiences effect was statistically significant only for the outdoor VE ($F_{train}(1)$ = 3.882 (p > 0.05); $F_{outdoor}(1)$ = 4.924 (p < 0.05); $F_{indoor}(1)$ = 1.498 (p > 0.05)) (Appendix F).

Table 17: Navigation time spent by participants classified by VE experiences

	VR Exp.	N	Mean (s)	Std. Deviation (s)	F	Sig.
Training	OC&NE	9	154.7	56.8	3.882	0.072
	VF&FR	5	100.9	26.8		
Outdoor	OC&NE	9	243.6	80.2	4.924	0.047
	VF&FR	5	156.9	43.1		
Indoor	OC&NE	9	123.9	30.1	1.498	0.245
	VF&FR	5	104.8	22.8		

5.4.3.2 VE Experience by Emotional Response Time

Given the decision making time analysis across sessions, the results are opposite to that of navigation time. From Table 18, the average time spent on deciding emotional valences by participant with less VE experiences was only 11 seconds more than those with more experiences (t-test (40) = 1.030, p = 0.309 > 0.05) (Appendix F).

Table 18: Summary of decision making time spent by participants
classified by VE experiences

	VR Exp.	N	Mean	Std. Deviation
Popup time	OC&NE	9	48.3	40.6
	VF&FR	5	36.7	20.8

The same results were obtained from ANOVA table in Appendix F, which are $F_{train}(1) = 0.001$ (p > 0.05), $F_{outdoor}(1) = 0.570$ (p > 0.05), $F_{indoor}(1) = 1.114$ (p > 0.05), and the decision making time analysis by each session shown in Table 19. Especially in the training session, the time spent by the two groups of participants is almost the same.

Table 19: Decision making time spent by participants
classified by VE experiences

	VR Exp.	N	Mean (s)	Std. Deviation (s)	F	Sig.
Training	OC&NE	9	25.3	55.6	0.001	0.982
	VF&FR	5	21.7	55.9		
Outdoor	OC&NE	9	61.2	52.3	0.570	0.465
	VF&FR	5	13.1	31.0		
Indoor	OC&NE	9	27.3	37.2	1.114	0.312
	VF&FR	5	12.3	23.4		

According to the statistical results of the analyses, both gender issue and VR experiences may have influenced performance time in the beta-testing of VEMap application but on different aspects. Gender may play a role in emotional response decision making. It seems that females may spend more time on choosing their emotion valence. Furthermore, based on observation made during trials, females liked to talk about their feelings during the session, which can add time too, while males were more likely to just type in the numbers. On the other hand, VE experience may influence navigation in the VEs during this test. For example, some participants with less experience had difficulty controlling their movement and orientation while navigating at beginning of training session. A typical strategy was to use the keyboard to reach certain point, and then moved the mouse to face a certain direction. In contrast, some in the VF&FR group, especially those proficient with video games, could control moving and orienting together continuously without stopping until they reached the marker where they had to stop to provide emotional valence as used by the popup screen.

6 Discussion

VEMap project is designed to capture a user's emotional responses in VEs by collecting GSR data as objective measures and self-reports as subjective measures, and to express the information about emotional arousal and emotional valence by a sequence of coloured dots on the VEMaps of the user. In order to validate the VEMap concept and assess the usability of this application, beta-testing with 14 participants was carried out with the results of the beta-testing presented in Chapter 5. Based on the results several points of discussion are carried.

6.1 Similarity between Real lab and Virtual Lab

According to Bishop (2003), the higher the degree of similarity between the real and virtual worlds, the more experience users can transfer between those virtual worlds and their associated physical worlds. For this reason, it makes sense to apply VE technology to architectural design and urban planning. Both designers and clients can design, evaluate and modify virtual models through the aid of a standard computer without the extra cost on making physical models or sample buildings. However, due to the limitations of technology, VE model fidelity can be compromised due to the size of VE computer file as models with more details must reside in large files that make it difficult to render images and viewpoints smoothly. Rendering problems can lead to obvious latencies in human-interface interaction. Image rendering latency can cause uncomfortable feelings in users, for example, motion sickness. If users feel uncomfortable within a VE then it is likely to affect not only navigation and decision-making time, but also how well they think the VE represents the real world. For this reason, designers of VEs must decide whether they want to include more objects to make the VE seem more detailed or less objects so that motion sickness is minimized. The users' view of VNCEP need to carry out the beta-testing for VEMap required relatively sample VE models in order to minimize the chances of participants experiencing motion sickness.

In the indoor (VE Use-IT Lab) session only large pieces of furniture, such as the sink, large screen display etc., were included in the VE to facilitate image rendering. To help improve

the simulation, similar colours to those of the real objects were used when modeling the VE objects used in the Use-IT Lab (Figure 25).

Figure 25: Virtual Use-IT Lab versus Real Lab

Based on the results of the similarity question, participants reported that the virtual Use-IT Lab was somewhat similar to the real Use-IT Lab ($p < 0.01$). None of the participants had ever visited the Use-IT Lab prior to participating in the beta-testing study. The setup of the lab was modified slightly for the purpose of conducting the beta-testing, and thus differed from the virtual representations which had been created earlier. This may have contributed to a similarity value that might be lower than if the VE representation was an exact match in terms of physical layout. Nevertheless, participants indicated experiencing a similar response to the virtual representation they did to the real world. In addition, upon completion of the indoor trial, participants explored the real lab with increased fascination compared with when they first arrived for the study. This highlights how VE technology—VEMap in this paper can encourage involvement and increase interest in real-world surroundings.

6.2 GSR Data Transformation and VEMap Creation

Although the best sites on the hand for measuring GSR are the medial phalanges of the index and middle fingers, the medial phalange of fifth finger and middle part of thumb might be the most optimal measurement sites for avoiding influence on GSR caused by finger movement of the other three fingers for keyboard control (Figure 14). However, based on a pre-test among Use-IT Lab group and recorded GSR data of the beta-testing, it might be reasonable to collect emotional arousal levels with either finger combinations.

Data transformation of individual participants is necessary in order to produce the VEMap. Taking the RMS values of original GSR data (See Formula 1) smoothes the raw readings and eliminates noise which might be extremely large. From the RMS GSR data summary in Table 3, it is obvious that minimum GSR values, the maximum GSR values and GSR data ranges vary by person, and even vary between sessions for the same participant depending on the VEs. The minimum GSR value of each participant, which was recorded while they were navigating through VEs instead of being under the calm situation, may not be the participant's true SCL — the real baseline of GSR for that participant. However, subtracting the minimum GSR value of each session, which might be seen as the relative baseline GSR of that session, appears to be a reasonable first approach for the purpose of the beta-testing (See Formula 2).

Separating adjusted GSR data into seven groups corresponding to seven colour intensities according to personal GSR range of one session seemed to express individual emotional responses with greater resolution, especially for those participants with small range of adjusted GSR data. For example, there might be no arousal level changes for that participant with a GSR range of 145.4nS compared to that with over 20000nS if applying the same grouping standard on these two samples. Further research should be carried out in order to find the most appropriate way of transforming GSR data in order to best capture an individual's true emotional responses.

Forty-two VEMaps were created, three (training session, outdoor session and indoor session) for each participants. Figure 22 shows two samples of the 14 VEMaps produced from training session for two participants. By examining the VEMaps for individuals the designer may learn more specific responses to designed areas. For example, the designer may learn that the user is comfortable while facing the sofa court, but generates very strong unpleasant feelings while moving towards the TV set and crossing the door to the other room. Consequently, the designer may modify the design of this area to reduce the strong negative emotions for this user. Once the VEMaps for the 14 particpants were created for each of the three test VEs, three overall VEMaps were created incorporating the emotional responses for all 14 participants. Figure 23 shows the sample of overall VEMap for training session. If

taking this meeting room of training session (right room) as a public place with 14 potential users, it may be easy for interior decoration designer to find out that the exit part of this room generates the most negative emotional responses, and needs to be changed immediately.

6.3 Match between VEMap and Verbal Feedback

According to comparison between VEMap and simultaneous verbal feedback, VEMap might accurately capture the participants' emotional responses, for example, whether they like this place or not (self-report); if they like it, which part they like the most (GSR testing). When participants reviewed their VEMaps, participants indicated that the representation was fairly accurate.

Although the VEMap is intended to capture an individual's experience, some common patterns were found. For the indoor session and training sessions, the most frequently commented environmental feature focused on colour of areas. For example, almost every participant provided positive feedback about the bright and warm colour in the meeting room of training session. On the other hand, almost 50% of participants disliked the blue doors in the virtual corridor of indoor session which interestingly is the actual colour of the real doors. Another common type of comment focused on the type of furnishings and features. In the outdoor session, the swimming pool and park were the most popular places. Participants were usually positive while moving through these two blocks based on verbal feedback and VEMaps.

Besides these two popular places, which most participants had in common, there were different responses to other locations. For instance, one participant provided a positive emotional response, and experienced a high arousal level in the construction area of the outdoor session. The participant's explanation was that the construction simulated the imagination. The participant indicated that this matched similar real world experiences when passing construction zones. However, most participants disliked this virtual area because of associated feelings of danger. This might suggest that VEMaps could be used to better understand individual differences in responses to design features.

During the VEMap evaluation procedure not all participants provided simultaneous verbal feedback which made it difficult to evaluate the true validity of the VEMap application. The situation may be better in a real-world application because clients who have thousands of dollars invested into a property development may be more involved in the design and may be more likely to use a verbal protocol so that the design team can better understand the client's likes and dislikes of the design.

6.4 Factors affecting the Usability of VEMap

Gender and VE experience are two factors that may affect the usability of VEMap. The results show that VE experience may influence time spent on navigating through a virtual world. More experienced users may be able to "walk" fluently in VEs with little or even without pre-training. Therefore, the experience of pre-training session might be necessary only for those with less VE experience. Moreover, this pre-training session could be varied depending on the user's VE experience (e.g. shorter for users with basic computer training; longer for those with no computer experiences). Gender may also have an effect on decision–making. Females seemed to take longer to respond when considering their emotional valence. For designers, it is a good time for getting comments and feedback on their designs when the pop-up screen is showing up. The more comments the designers can obtain from clients, the better they can understand clients' emotional responses (Females seem to do this more naturally than males). Thus, the popup screen may need to remind male participants to also talk about their emotional responses to the environments or designs.

Generally, from a usability perspective most participants could manipulate all the controls fluently even after training. There was one male participant among 14 who experienced slight motion sickness after finishing two sessions. He recovered and got back to the final session after a two-minute break. According to his comments, this sickness was caused by the conflicts of sitting still on the chair and virtually moving in VEs. Despite the low incidence of motion sickness experienced in VEMap evaluation, more attention should be paid to avoid this uncomfortable reaction among participants as it may confound data collection. Therefore, another benefit of simple pre-training might be assist on minimizing motion sickness that could arise during VEMap evaluation.

7 Recommendation and Future Work

As mentioned in Chapter 5 and Chapter 6, the data obtained from the VEMap evaluation procedure can be considered to be valid. Based on the results of the beta-testing, VEMap seems to be able to express users' emotional responses to VEs. Given this encouraging outcome, two main application areas for VEMap are put forward. One recommendation is for VEMap to be used in architectural design as an evaluation tool for improving communication between designers, clients and other stakeholders. For example, there is a new engineering building under construction across from the Davis Center on the campus of University of Waterloo. Designers could use VEMap to get feedback about the setup of each floor, the color of walls from potential users before the building is finished.

Secondly, this VEMap method could be used to promote real estate sales. The benefits would be that VEMap could aid visualization of virtual models; and help potential clients figure out their favourite condo or house with the help of VEMap. This innovative technology could be used to increase users' interest and involvement, especially for the younger generation, who would like to access advanced technology.

Given the cost associated with the ME6000 device used to collect GSR data, real estate companies might consider using a version of VEMap that just collects emotional response (via the pop-up screens) if they are only looking for general emotional response to a VE. In the case of getting user feedback for promotional or advertising purposes, accurate emotional arousal may be less important, given its cost and actual contribution. Emotional status showing on the VEMap may be enough for helping clients find their favourite designs. Future studies of VEMap might compare VEMaps generated using GSR and emotional valence versus using only emotional valence.

The other main recommended application is for urban planning research. Not only can individual feedback be indicated by VEMap, but composite VEMap for a number of participants can provide overall information (Figure 23). It might be useful for designers to

get the common emotional responses from a group of clients for spaces and buildings that will be used by the general public rather than a single user.

Getting feedback from the public is valuable for planners and municipal councils as residents may have completely different ideas on how public spaces should be used. As the construction example mentioned above, a designer might think that people dislike this site because of possible danger and design fencing to block views of construction. On the contrary, some people might enjoy construction sites and so designers may want to build observation windows into any fencing for those people. Consequently, getting feedback from the public may be significant to avoid misjudgment of urban planners and council for sites that planners and council may assume will be either positively or negatively received by residents.

As an investigation tool or evaluation tool, for instance, VEMap could have been helpful for making the final design of Waterloo Uptown Public Square. Residents could have announced their feedback on which idea should be kept among "Bell and whistle", "Water Wall", or "Skating Rink" etc. In the end, all the responses can be mapped on one base map. The result might be meaningful for making final decisions on such a public place.

Finally the VEMap application could be helpful in the design of VE application areas for which users' emotional responses may increase or decrease user interaction, such as 3D game design.

As for future work, there should be two main parts as well. The first one is further programming of the VEMap application. So far, VEMap is made by several separated software and platform including *MegaWin®*, VNCEP®, *Excel®* and *Surfer®*. In order to simplify the mapping procedure, customized and integrated software is required. Future research work would ensure that designed experiments have an equal number of males and females as well as equal numbers of VE experience levels. Ideally, the data would be more valuable if further research can be made using VEs in real design situations instead of being limited to the very basic VEs that were used in beta-testing this version of VEMap.

References

Al-Kodmany, K. (2002), Visualization Tools and Methods in Community Planning: From Freehand Sketches to Virtual Reality, *Journal of Planning Literature,* vol.17, pp. 189-211

Basdogan, C., Sedef, M., Harders, M. & Wesarg, S. (2007), VR-based Simulators for Training in Minimally Invasive Surgery, *IEEE Computer Graphics and Applications*, vol. 27, no. 2, pp. 54-66.

Bell, J. T., Fogler, H. C. (2004), The Application of Virtual Reality to (Chemical Engineering) Education, *IEEE Virtual Reality Conference 2004*, pp. 217-218.

Bishop, I. D. & Rohrmann, B. (2003), Subjective Responses to Simulated and Real Environments: A Comparison, *Landscape and Urban Planning*, vol. 65, no. 4, pp. 261-277.

Boucsein, W. (1992), Electrodermal Activity-The Plenum Series in Behavioral Psychophysiology and Medicine, New York: Plenum Press.

Brown, A. S. (2006), Virtual O.R, *Mechanical Engineering*, vol.128, no.11, pp.32-35.

Chen Y. (2007), Virtual Reality-New Promotion Method in Real Estate Sale (In Chinese), *Urban Development*, vol. 16, pp. 74-75.

Christie, M. J. (1981), Electrodermal Activity in the 1980s: A Review, *Journal of the Royal Society of Medicine*, vol. 74(8), pp. 616-622.

Dindo, L., Fowles, D.C. (2008), The skin conductance orienting response to semantic stimuli: Significance can be independent of arousal, *Psychophysiology*, vol. 45, pp. 111–118.

Dos Santos, C. T., Osorio, F. S. (2004), AdapTIVE: An Intelligent Virtual Environment and Its Application in E-commerce, *Proceedings of the 28th Annual International Computer Software and Applications Conference*, vol. 1, pp. 468-473.

Drettakis, G., Roussou, M., Reche, A. & Tsingos, N. (2007), Design and Evaluation of A Real-world Virtual Environment for Architecture and Urban Planning, *Presence*, vol.16, no. 3, pp. 318-332.

Harrold, D. (2000), Virtual Reality Saves Money, *Control Engineering*, vol. 47, no. 11, pp. 36-45.

Howard, T. L. J. & Gaborit, N. (2007), Using Virtual Environment Technology to Improve Public Participation in Urban Planning Process, *Journal of Urban Planning and Development*, vol. 133, no. 4, pp. 233-41.

Johnson, A., Moher, T., Cho, Y.J., Lin, Y.J., Haas, D., & Kim, J. (2002), Augmenting Elementary School Education with VR, *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 6-9.

Laparra-Herna′ndez, J., Belda-Lois, J.M., Medina, E., Campos, N., Poveda, R., (2009), EMG and GSR signals for evaluating user's perception of different types of ceramic flooring, *International Journal of Industrial Ergonomics*, vol. 39, pp. 326–332.

Leetaru, K. (2004), ShadowLight: A Flexible Architecture for Immersive Rapid Prototyping and Design, http://shadowlight.ncsa.uiuc.edu/index.html, retrieved on Aug 12th, 2007

Mandryk, R. L. (2005), Modeling User Emotion in Interactive Play Environments: A Fuzzy Physiological Approach, Doctoral Thesis, The University of Simon Fraser (Order Number: AAINR16967).

Mazuryk, T. & Gervautz, M. (1996), Virtual Reality: History, Applications, Technology and Future, Technical Report, TR-186-2-96-06, Institute of Computer Graphics, Technical University of Vienna, Austria.

Meehan, M. J. (2001), Physiological Reaction as an Objective Measure of Presence in Virtual Environments, Doctoral Thesis, The University of North Carolina at Chapel Hill (UMI Order Number: AAI3007847).

Montgomery, D.C. (2005), Design and Analysis of Experiments, 6th Edition, published by John Wiley &Sons, Inc.

Morie, J.F., Williams, J., Dozois, A. & Luigi, D.P. (2005), The Fidelity of Feel: Emotional Affordance in Virtual Environments, *The 11th International Conference on Human-Computer Interaction*, Las Vegas, NV.

Nold, C. (2004), BioMapping, retrieved on Aug 22nd, 2007 from http://www.biomapping.net/new.htm.

Norman, D. A. (2004), Emotional Design: Why We Love (or Hate) Everyday Things, New York: Basic Books.

Prokasy, W.F., & Raskin, D.C. (1973), Electrodermal Activity in Psychological Research, New York: Academic press.

Schroeder, R. (1993), Virtual reality in the real world: History, applications and projections, *Futures*, vol. 25, no. 9, pp. 963-973.

Sebok, A., Nystad, E. & Droivoldsmo, A. (2002), Improving Safety and Human Performance in Maintenance and Outage Planning through Virtual Reality-based Training Systems, *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants,* NY, pp.8.14-8. 22.

Simpson, D.M. (2001), Virtual Reality and Urban Simulation in Planning: A Literature Review and Topical Bibliography, *Journal Of Planning Literature*, vol. 15, no. 3, pp. 359-376.

Soo, S. M. K., Chan, S. C. F. & Ng, V. T. Y. (2006), Co-assembler - Supporting Collaborative Product Configuration for Customer-oriented E-commerce, *Proceedings 10th International Conference on Computer Supported Cooperative Work in Design*, Piscataway, NJ, pp.1-5.

Thacker, P.D. (2003), Fake Worlds Offer Real Medicine: Virtual Reality Finding a Role in Treatment and Training, *The Journal of the American Medical Association*, vol.290, no.16, pp.2107-2112.

Vervliet, B., Vansteenwegen, D. & Eelen, P. (2004), Generalization of Extinguished Skin Conductance Responding in Human Fear Conditioning, Learning and Memory, vol. 11, pp 555-558.

Wang, Y.J., Minor, M.S. (2008), Validity, Reliability, and Applicability of Psychophysiological Techniques in Marketing Research, Psychology & Marketing, vol. 25(2), pp. 197–232.

Wilson, J. M. & Jennings, W. C. (2000), Studio Courses: How Information Technology Is Changing the Way We Teach, on Campus and Off, *Proceedings of the IEEE*, vol. 88, no. 1, pp. 72 – 80.

Xu, N. (2004), http://www.86vr.com/case/realestate/200410/3852.html (In Chinese), retrieved on Nov. 2nd, 2007

Ye, J., Campbell, R. I., Page, T. & Badni, K. S. (2006), An Investigation Into the Implementation of Virtual Reality Technologies in Support of Conceptual Design, *Design Studies*, vol.27, no.1, pp.77-97.

Zhao, Y. M., Guo, L. J., Wang, X. L. & Pan, Z.G. (2004), A 3D virtual shopping mall that has the intelligent virtual purchasing guider and cooperative purchasing functionalities, *Proceedings 8th International Conference on Computer Supported Cooperative Work in Design*, vol. 2, pp. 381-385.

U.S. Congress Office of Technology Assessment (1994), *Virtual Reality and Technologies for Combat Simulation—Background Paper* (OTA-BP-ISS-136), Washington, DC: U.S. Government Printing Office.

Appendix A
Study Information Letter

Project information:
Title: VEMap: A Visualization Tool for Evaluating Emotional Responses in Virtual Environments
Instructor: Dr. Carolyn Macgregor, Assistant Professor, Systems Design Engineering
Investigator: Hong (Julie) Zhu, M.A.Sc. Candidate, Systems Design Engineering

Objective:
The objectives of this research are to measure and visually represent emotional responses while navigating through an immersive computer simulated 3D environment. Emotional responses will be measured by combining both subjective (participant responses captured by video recorder and questionnaires) and objective measures (galvanic skin response test). All the emotion data will be plotted on a 2D map of the corresponding environment generating the VEMap, and represented by coloured dots. Therefore, emotional responses, containing the information about emotional status and arousal level at certain location in that virtual environment, could be displayed directly through VEMap.

Potential Participant
Participants should have normal or corrected to normal vision to be able to see the paths in the virtual environments. If you have an active implantable medical device, such as a heart pacemaker you cannot participate in this study."

Time Commitment
The whole experimental procedure will take you approximately 30 minutes.

Procedure:
If you agree to take part in this study, you will be asked to explore three virtual environments by following paths. These environments include one simple training session at first, followed by an indoor and outdoor environment in a random order. Meanwhile, a Galvanic Skin Response (GSR) sensor will be wrapped around the fingers of one hand for collecting skin conductance response data. You will be asked to self-report your opinions on these environments. Your verbal feedback will be captured by a video recorder on a fixed position. At the end, you will be asked to complete a questionnaire including demographic questions and your comments on this study through comparing your virtual experiences obtained from the experimental trials and your physical experiences. You may decline to answer any of the questions if you so wish.

Benefits
Participants will have the opportunity to use high-tech visualization equipment and the chance to explore novel virtual environments.
Otherwise, researchers can use the information to improve methods for measuring emotions in users evoked by virtual environment (VE). If this approach is effective, designers of VE products (e.g. computer games) can improve the products based on this information.
.

Risks

During the experiment, participants will be asked to explore three large scale virtual environments, which might result in symptoms similar to motion sickness. Participants are encouraged to take a break after exploring each environment, for instance, play with Wii, lie down on the couch, or breathe fresh air next to windows. Participants can also withdraw from experiment at any time if they are uncomfortable during experimental procedure. Otherwise, all equipment used in this experiment is standard devices and pose no health risks to the participants.

Compensation

Participants will be able to select one prize from a selection of prizes, which are Chinese traditional artworks including Painted Hand Fan, Bronze Bracelet with Decoration, Glass Bottles with Inner Hand Drawing and Peking Opera Mask, each with a value of approximately $5. Participants can still select one prize if they withdraw from this experiment.

Confidentiality

All information you provide is considered completely confidential. Your name will not appear in any thesis or report resulting from this study, however, with your permission anonymous quotations may be used. All the data including paper records, video records and electronic data collected during this study will be retained for one year in a locked office, Use-It lab (E2-3367). Only members of Use-It Lab will have access.

Right to Withdraw

You may decide to withdraw from this study at any time without any negative consequences by advising the researcher. All data collected will be excluded from the results and destroyed.

If you have any questions regarding this study, or would like additional information to assist you in reaching a decision about participation, please contact me at 519-888-4567 Ext. 35607 or by email at h22zhu@engmail.uwaterloo.ca . You can also contact my supervisor, Professor Prof. Carolyn MacGregor at 519-888-4567 ext. 33742 or email cgmacgre@engmail.uwaterloo.ca .

Ethics Review

This study has been reviewed and received ethics clearance through the Office of Research Ethics at the University of Waterloo. If you have any comments or concerns resulting from your participation in this study, please contact Dr. Susan Sykes of this office at 519-888-4567 Ext. 36005.

Appendix B
Consent Letter

I have read the information presented in the information letter about a study entitled "VEMap-A Visualization Tool for Evaluating Emotional Responses in Virtual Environments" being conducted by Hong (Julie) Zhu of the Department of Systems Design Engineering at the University of Waterloo. I have had the opportunity to ask any questions related to this study, and received satisfactory answers to my questions.

I was informed that I will be asked to complete a demographic questionnaire for providing personal information including age, gender, and computer game experiences.

I was informed that I will be asked to follow the paths in order to finish this experiment. These paths are set up in advance, and will be shown clearly in those environments.

I was informed that a Galvanic Skin Response (GSR) sensor would be wrapped around my fingers of one hand for collecting skin conductance data.

I was informed that my voice will be recorded to ensure an accurate recording of my responses during the experimental trials.

I was informed that I will be asked about a question about comparing virtual experiences I got from experimental trials and my physical experiences.

I was informed that this experiment would last approximately 30 minutes.
I was informed that there may be a potential risk to experience motion sickness during experiment.
I was informed that excerpts from the video recording may be included in the thesis and/or publications to come from this research, with the understanding that the quotations will be anonymous.
I was informed that I may withdraw my consent at any time without penalty by advising the researcher.
I was informed that this project has been reviewed by, and received ethics clearance through, the Office of Research Ethics at the University of Waterloo, and that if I have any comments or concerns resulting from my participation in this study, I may contact the Director, Office of Research Ethics at 519-888-4567 ext. 36005.

With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.
[   ]YES   [   ]NO

I agree to have my verbal feedback video recorded.
[   ]YES   [   ]NO

I agree to the use of anonymous quotations in any thesis or publication that comes of this research.
[   ]YES   [   ]NO

Participant Name: _____ (Please print)

Participant Signature: _____

Witness Name: _____ (Please print)

Witness Signature: _____

 Date: _____

Appendix C
Questionnaire for VEMap Study

General Questions:

Gender: [ ] Male [ ] Female   Age: ____
Hand Dominance: [ ] Right [ ] Left

1. How would you rate your use of virtual environment systems?
[ ] Very Frequent (more than once per week)
[ ] Frequent (more than once a month)
[ ] Occasional (less than once per month)
[ ] Never
2. If you have used any virtual environment systems before, please list them out below:
-

_____

_____

_____

_____

3. Do you have same or similar feelings in the virtual Use-It lab compared your experience
    in real one?
  （You can explore the real lab before answering）Please rate your feelings
    1.   2.   3.   4.   5.
Completely   Completely
Different   Same

Comments:_____

_____

_____

_____

_____

Appendix D
Participant Feedback Letter

Date:

Dear _____,

I would like to thank you for your participation in this study. As a reminder, the purpose of this study is to test users' emotional responses while navigating through virtual environments. Emotion data will be plotted on a 2D map of the corresponding environment generating the VEMap.

This study will result in improved understanding of the feelings that people experience while exploring virtual environments (VE). Improved understanding of emotional responses in VEs will allow for improved real world designs based on virtual design feedback. Furthermore, improved techniques for measuring VE emotional responses can be used by researchers to measure level of presence experienced by virtual users.

Please remember that any data pertaining to you as an individual participant will be kept confidential. Once all the data are collected and analyzed for this project, I plan on sharing this information with the research community through seminars, conferences, presentations and journal articles. If you are interested in receiving more information regarding the results of this study, or if you have any questions or concerns, please contact me at either the phone number or email address listed at the bottom of the page. If you would like your personal VEMap and a summary of the results, please let me know now by providing me with your email address. I appreciate any suggestions or feedback after you receive your personal VEMap. If requested, I will send the result to you upon study completion.

As with all University of Waterloo projects involving human participants, this project was reviewed by, and received ethics clearance through, the Office of Research Ethics at the University of Waterloo. If you have any comments or concerns resulting from your participation in this study, please contact Dr. Susan Sykes in the Office of Research Ethics at 519-888-4567, Ext., 36005.

Student Researcher Name: Hong (Julie) Zhu

University of Waterloo
Department: Systems Design Engineering

Contact Telephone Number: 519-888-4567. Ext. 35607

Email Address: h22zhu@engmail.uwaterloo.ca

Appendix E

Statistical Analysis of Time Spent by Participants Classified by Gender

T-test of overall time spent by male and female participants

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
all time	Equal variances assumed	.516	.477	-1.486	40	.145	-42.12737	28.34707	-99.41893	15.16419

ANOVA of overall time spent by male and female separated by session

		Sum of Squares (s)	Df	Mean Square (s)	F	Sig.
Training	Between Groups	3847.51	1	3847.51	.801	.388
	Within Groups	57664.80	12	4805.40		
	Total	61512.31	13			
Outdoor	Between Groups	6649.39	1	6649.39	.675	.427
	Within Groups	118190.05	12	9849.17		
	Total	124839.44	13			
Indoor	Between Groups	4907.37	1	4907.37	3.576	.083
	Within Groups	16466.08	12	1372.17		
	Total	21373.45	13			

T-Test of navigation time spent by male and female participants

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
navigation time	Equal variances assumed	.008	.929	-.506	40	.616	-12.28008	24.27921	-61.35019	36.79002

ANOVA of navigation time spent by male and female separated by session

		Sum of Squares (s)	Df	Mean Square (s)	F	Sig.
Training	Between Groups	486.88	1	486.88	.156	.700
	Within Groups	37487.87	12	3123.99		
	Total	37974.76	13			
Outdoor	Between Groups	142.08	1	142.08	.021	.888
	Within Groups	82869.16	12	6905.76		
	Total	83011.24	13			
Indoor	Between Groups	800.12	1	800.12	.991	.339
	Within Groups	9690.55	12	807.55		
	Total	10490.67	13			

T-test of decision making time spent by male and female participants

		Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference		Lower	Upper
popup time	Equal variances assumed	14.041	.001	-2.680	40	.011	-29.84732	11.13858		-52.35924	-7.33540

ANOVA of decision making time spent by male and female separated by session

		Sum of Squares (s)	Df	Mean Square (s)	F	Sig.
Training	Between Groups	1597.03	1	1597.03	3.547	.084
	Within Groups	5403.23	12	450.27		
	Total	7000.26	13			
Outdoor	Between Groups	4847.53	1	4847.53	2.132	.170
	Within Groups	27282.30	12	2273.53		
	Total	32129.83	13			
Indoor	Between Groups	1744.43	1	1744.43	3.844	.074
	Within Groups	5445.88	12	453.82		
	Total	7190.31	13			

Appendix F

Statistical Analysis of Time Spent by Participants Classified by VR Experiences

T-test of overall time spent by participants classified by VR experiences

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
all time	Equal variances assumed	5.748	.021	2.542	40	.015	64.74314	25.47388	13.25851	116.22777

ANOVA of overall time spent by participants classified by VR experiences

		Sum of Squares (s)	Df	Mean Square (s)	F	Sig.
Training	Between Groups	9177.12	1	9177.12	2.104	.173
	Within Groups	52335.19	12	4361.27		
	Total	61512.31	13			
Outdoor	Between Groups	37478.52	1	37478.52	5.148	.043
	Within Groups	87360.93	12	7280.08		
	Total	124839.44	13			
Indoor	Between Groups	3461.19	1	3461.19	2.319	.154
	Within Groups	17912.26	12	1492.69		
	Total	21373.45	13			

T-test of navigation time spent by participants classified by VR experiences

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
navigation time	Equal variances assumed	4.810	.034	2.487	40	.017	53.15053	21.37074	9.95864	96.34241

ANOVA of navigation time spent by participants classified by VR experiences

		Sum of Squares (s)	Df	Mean Square (s)	F	Sig.
Training	Between Groups	9282.60	1	9282.60	3.882	.072
	Within Groups	28692.15	12	2391.01		
	Total	37974.76	13			
Outdoor	Between Groups	24151.95	1	24151.95	4.924	.047
	Within Groups	58859.29	12	4904.94		
	Total	83011.24	13			
Indoor	Between Groups	1163.93	1	1163.93	1.498	.245
	Within Groups	9326.74	12	777.23		

T-test of decision making time spent by participants classified by VR experiences

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
popup time	Equal variances assumed	1.412	.242	1.030	40	.309	11.59258	11.25700	-11.15866	34.34382

ANOVA of decision making time spent by participants classified by VR experiences

		Sum of Squares (s)	df	Mean Square (s)	F	Sig.
Training	Between Groups	.30	1	.30	.001	.982
	Within Groups	6999.96	12	583.33		
	Total	7000.26	13			
Outdoor	Between Groups	1458.07	1	1458.07	.570	.465
	Within Groups	30671.76	12	2555.98		
	Total	32129.83	13			
Indoor	Between Groups	610.84	1	610.84	1.114	.312
	Within Groups	6579.47	12	548.29		