Affective appraisal of virtual environments

Joske Houtkamp

4 April 2012

M.C. Escher, Prentententoonstelling, adapted by B. de Smit and H.W. Lenstra Jr., 2003
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4.2 Related Work</td>
<td>81</td>
</tr>
<tr>
<td>4.4.3 Hypotheses</td>
<td>82</td>
</tr>
<tr>
<td>4.4.4 Method</td>
<td>83</td>
</tr>
<tr>
<td>4.4.5 Results</td>
<td>85</td>
</tr>
<tr>
<td>4.4.6 Conclusions and discussion</td>
<td>88</td>
</tr>
<tr>
<td>4.5 EXPERIMENT: VISUAL DYNAMICS AND SOUNDS IN AN OUTDOOR ENVIRONMENT</td>
<td>91</td>
</tr>
<tr>
<td>4.5.1 Introduction</td>
<td>91</td>
</tr>
<tr>
<td>4.5.2 Related Work</td>
<td>92</td>
</tr>
<tr>
<td>4.5.3 Method</td>
<td>93</td>
</tr>
<tr>
<td>4.5.4 Results</td>
<td>96</td>
</tr>
<tr>
<td>4.5.5 Conclusions and discussion</td>
<td>101</td>
</tr>
<tr>
<td>4.6 EXPERIMENT: EXPERIENCING DANGER ON A VIRTUAL HIGHWAY</td>
<td>104</td>
</tr>
<tr>
<td>4.6.1 Introduction</td>
<td>104</td>
</tr>
<tr>
<td>4.6.2 Related Work</td>
<td>104</td>
</tr>
<tr>
<td>4.6.3 Method</td>
<td>106</td>
</tr>
<tr>
<td>4.6.4 Results</td>
<td>106</td>
</tr>
<tr>
<td>4.6.5 Conclusion &amp; Discussion</td>
<td>110</td>
</tr>
<tr>
<td>4.6.6 Appendix: Images of the video clips (original – enhanced)</td>
<td>114</td>
</tr>
<tr>
<td>4.7 EXPERIMENT: THE EFFECT OF SOUNDS ON ENGAGEMENT AND AROUSAL IN A VIRTUAL TRAINING</td>
<td>115</td>
</tr>
<tr>
<td>4.7.1 Introduction</td>
<td>116</td>
</tr>
<tr>
<td>4.7.2 Related Work: sound in virtual environments</td>
<td>116</td>
</tr>
<tr>
<td>4.7.3 Experiment 1: realistic sounds in a virtual training environment</td>
<td>119</td>
</tr>
<tr>
<td>4.7.4 Results of experiment 1</td>
<td>123</td>
</tr>
<tr>
<td>4.7.5 Conclusions on experiment 1</td>
<td>124</td>
</tr>
<tr>
<td>4.7.6 Experiment 2: enhanced sounds in a virtual training environment</td>
<td>125</td>
</tr>
<tr>
<td>4.7.7 Results of experiment 2</td>
<td>126</td>
</tr>
<tr>
<td>4.7.8 Conclusions on experiment 2</td>
<td>128</td>
</tr>
<tr>
<td>4.7.9 Conclusions and Discussion</td>
<td>129</td>
</tr>
<tr>
<td>4.8 EXPERIMENT: THE INFLUENCE OF CYBERSICKNESS OF THE AFFECTIVE APPRAISAL OF A VIRTUAL ENVIRONMENT</td>
<td>131</td>
</tr>
<tr>
<td>4.8.1 Introduction</td>
<td>131</td>
</tr>
<tr>
<td>4.8.2 Related work</td>
<td>131</td>
</tr>
<tr>
<td>4.8.3 Method</td>
<td>132</td>
</tr>
<tr>
<td>4.8.4 Results</td>
<td>136</td>
</tr>
<tr>
<td>4.8.5 Conclusions and Discussion</td>
<td>138</td>
</tr>
<tr>
<td>4.9 EXPERIMENT: EFFECTS OF STRESS AND DARKNESS ON THE APPRAISAL OF A VIRTUAL ENVIRONMENT</td>
<td>139</td>
</tr>
<tr>
<td>4.9.1 Introduction</td>
<td>139</td>
</tr>
<tr>
<td>4.9.2 Related work</td>
<td>139</td>
</tr>
<tr>
<td>4.9.3 Hypotheses</td>
<td>140</td>
</tr>
<tr>
<td>4.9.4 Materials and instruments</td>
<td>141</td>
</tr>
<tr>
<td>4.9.5 Procedure</td>
<td>144</td>
</tr>
<tr>
<td>4.9.6 Results</td>
<td>144</td>
</tr>
<tr>
<td>4.9.7 Conclusion and discussion</td>
<td>148</td>
</tr>
<tr>
<td>4.10 SUMMARY OF THE RESULTS</td>
<td>150</td>
</tr>
<tr>
<td>4.10.1 Experiments 4.2 – 4.5: Affective appraisal of environmental visualisations</td>
<td>150</td>
</tr>
<tr>
<td>4.10.2 Experiments 4.6 and 4.7: Affective/emotional reactions to virtual training environments</td>
<td>151</td>
</tr>
<tr>
<td>4.10.3 Experiments 4.8 and 4.9: Effects of response moderators in environmental visualisations</td>
<td>152</td>
</tr>
<tr>
<td>4.10.4 Outline of the experimental results</td>
<td>152</td>
</tr>
</tbody>
</table>

5 CONCLUSIONS & DISCUSSION..................................................................154

5.1 INTRODUCTION毛泽东主席曾言: “一个人穷一点不要紧,只要他有理想,就一定能够克服他的困难．”．154
5.2 ANSWERS TO THE FOUR SUB QUESTIONS毛泽东主席曾言: “一个人穷一点不要紧,只要他有理想,就一定能够克服他的困难．”．155
5.3 ANSWER TO THE MAIN RESEARCH QUESTION AND CONCLUSIONS毛泽东主席曾言: “一个人穷一点不要紧,只要他有理想,就一定能够克服他的困难．”．164
5.4 RECOMMENDATIONS FOR DEVELOPMENT AND USE OF DESKTOP VIRTUAL ENVIRONMENTS毛泽东主席曾言: “一个人穷一点不要紧,只要他有理想,就一定能够克服他的困难．”．168
5.5 LIMITATIONS OF THIS RESEARCH AND RECOMMENDATIONS FOR FUTURE RESEARCH毛泽东主席曾言: “一个人穷一点不要紧,只要他有理想,就一定能够克服他的困难．”．170

6 REFERENCES....................................................................................172

7 SUMMARY.....................................................................................185
1 Introduction

1.1 A retrospect: the inspiration for this research project

In 2003, the Rotterdam City Building Department started the project Visualisatie RandstadRail, to assess the benefits of 3D visualisations in large building projects. One of the goals was to explore the use of 3D computer models in the design and decision process to improve the public safety of future built environments, and to test designs for aspects of public safety.

At that time an extensive building project was being prepared around Rotterdam Central Station, involving new railway tracks, a tunnel and new subway stations. In front of the main entrance of the railway station, complex and extensive building activities were to take place. The familiar and open appearance of the area would be transformed by high wooden fences and building cranes, creating a chaotic impression and making orientation difficult. The disorderly appearance of the area might make the area unpleasant, especially at night, and be linked with the occurrence of vandalism and lack of surveillance. The department feared that travelers would consider the area unsafe and avoid using public transport in the area.

The 3D models for this project were developed and rendered in a style that was, and still is, common in architectural practice and in which geometrically accurate models and realistic textures are used to create a very believable image. The variable attributes, such as time of day, weather, season, number and appearance of people, traffic, and also common urban features such as billboards, litter, graffiti, and signs of vandalism, were left to the rather arbitrary decision of the designer, often based on aesthetic preferences. These attributes however would have made a rendered model an accurate representation of reality. In this case, for the models of the Rotterdam project, the conventions of the artist’s impression were applied. This means that the general appearance of the models was very clean, bright, uncluttered and well-ordered (Figure 1.1).

![Figure 1.1 A screenshot from the virtual environment used for the project “Visualisatie RandstadRail” (MCW Studio’s, I-GWR, 2006).](image-url)
During the project, experts in urban planning used the models to assess objective factors for public safety such as sightlines, illumination, navigational aids, and escape routes (Durmišević, 2002). These require geometrically accurate modeling, but a low level of detail. The models developed for this project presented the area in a manner optimal for spatial comprehension of the area, mainly because variable elements such as weather conditions, people etc., were optimised, and because viewers were able to take a bird’s eye view of the area.

The steering committee intended to present the same models to area residents, to assess how safe they would feel in the environment shown, so the subjective public safety factor (Durmišević, 2002). However, the subjective safety of an area is determined by for instance decay, lighting conditions, and other persons present; these features of the area and ambient (dynamic) factors require detailed modeling and textures. The perceived safety is also dependent on personal factors such as physical discomfort and appreciation of the area; these require use of emotion-inducing effects and techniques. Obviously, the models used were not designed for the assessment of subjective safety, and could not serve both goals, spatial as well as affective (subjective) assessment. When viewers want to make a mental representation of the area in less favourable conditions, they have to make an effort to replace many depicted elements by imagined elements. The function of 3D models for visualising complex information and preventing ambiguity among different viewers is reduced severely in view of this affective requirement.

The same applies for virtual environments used for training purposes. In some simulations participants are trained to familiarise themselves with the layout of a building, remember escape routes, and know the correct procedures for evacuation; for that purpose a geometrically accurate but somewhat abstract representation without embellishments of the main spaces, corridors, elevators and obstructions, is sufficient. However, when the simulation is intended to convey the danger of a situation and train the participant in acting in a stressful situation, features in the representation that are in reality important for the ambience of the area, and for instance diminish overview, should be included.

Since then, we have seen similar effects in many projects: the apparent realism of 3D models or virtual environments convinces clients that they are valid representatives of the real environment, to be used for any purpose. But in fact, they are designed for spatial tasks or other cognitive tasks, not for assessment of the affective qualities of an environment. They do not contain the required information for that purpose, which can be visual, but also of another modality. These observations led to the following questions. Which factors in the content and representation determine the validity of 3D models for assessment of affective qualities of environments? What information is required for the affective assessment of 3D environments? How does the modeling process and representation on a monitor influence affective qualities of environments? These questions were the starting point for this study.

1.2 Research domain

In this research, we focused on two categories of 3D computer graphics representing environments:

1. visualisations of natural environments and landscapes, built (man-made) environments, and architectural visualisations;
2. virtual environments used for training or serious gaming.

Both categories are generally designated as virtual environments, VE’s or 3D models, referring to the modeling techniques and processes, as well as to the software and hardware, that are common to both. They allow a viewer to navigate and assume viewpoints as in a real environment, in real time. Often the term virtual environment suggests that the virtual world also offers other interaction possibilities, which are related to the training goal or are meant to make the user feel more immersed in the environment.
1.2.1 Environmental and architectural visualisations

Visualisations of built and natural environments and of architectural designs have been extensively used for research in environmental psychology, and as tools for planning and decision making, before the advancements made in computer graphics. Until circa 1990 slides and photographs were widely accepted as substitutes for real environments (Aoki, 1999; Bell, Greene, Fisher, & Baum, 2001), although researchers have drawn attention to deficiencies in these simulations (Danford & Willems, 1975; Sheppard, 2005a; Wergles & Muhar, 2009). In architecture and in urban and landscape planning, drawings, plans and elevations were traditionally used to visualise future environments and buildings.

Since the last 20 years, 3D computer visualisations have become increasingly commonplace in design and planning of urban environments and landscapes, and are regarded as potent communication tools (Paar, 2006; Pietsch, 2000). They support planning decisions and allow greater involvement of the public by providing images of designs of an urban area or landscape that can be easily recognized and understood (Sheppard, 2005b). It is thought that presentations that better match daily visual references (Bishop & Rohrman, 2003) improve communication between stakeholders, and elicit more detailed responses (Van den Brink, van Lammeren, van der Velde, & Dâne, 2007). Interactive visualisations are sometimes even available online, for the purpose of public participation. Examples are the models used for building projects in several Dutch cities such as Venlo and Tilburg (Chapter 4.4).

The interpretation of more abstract visualisations requires education and experience which makes them less suitable for non-professionals. 3D visualisations convey information on the assumed future in a manner that is readily understandable for non-expert viewers, by offering viewpoints as in reality and details of for instance sky, terrain, buildings, and vegetation (Sheppard, 2005b; Paar, 2006). Highly accurate and detailed 3D visualisations are therefore considered especially important for laypersons, whereas experts are thought to have enough knowledge to create a mental model from drawings and maps. 3D visualisation is expected to offer better cognitive understanding of spatial relations (topology) and vertical dimensions (geometry) (Bos, Bregt, Bulens, & Lammeren, 1998) than 2D images, especially when users can navigate through the environment (Bleisch, Dykes, & Nebiker, 2008; Müller, Sack-da Silva, & Bruns, 2007). Photorealistic representations, including material textures, illumination and atmospheric conditions, are considered to improve recognition, support understanding, and elicit appraisal of aesthetic and affective qualities of the area (e.g. Sheppard, 2005a).

By revealing detailed and sometimes complex information of a future environment, 3D visualisations decrease ambiguities that may occur when 2D visualisations (such as maps and elevations) are combined into a 3D mental representation (Tress & Tress, 2003).

Validity, “realism”, and accuracy

Although the advantages of 3D visualisations are widely accepted, there is equally extensive concern on the validity of 3D models as surrogates of the real world in various conditions (Wergles & Muhar, 2009). Validity generally refers to whether an instrument or finding is sound, defensible and well-grounded or appropriate to the issue at hand (Sheppard, 2005a). A valid simulated environment is defined by Wergles and Muhar (2009) as one that produces a cognitive, affective and behavioural response in the observer equivalent to the response produced by the real environment. This criterion is called either response equivalence (Appleyard, 1977), representational validity (Craik, Appleyard & McKechnie, 1980; Daniel & Meitner, 2001) or predictive validity (Sheppard, 2005a).

Research on the validity of visualisations has focused on the relation of validity with realism and with level of detail, abstraction, and accuracy, and on connected issues such as how much realism is sufficient for decision making (Appleton & Lovett, 2003; Bodum, 2005; Paar, 2006). Scientific literature on these topics is confusing, in the first place because authors use different terminology. Secondly, the approach to the research question varies widely, from the point of view from
environmental psychology, concerned with the subjective experience of realism, to a developers’ angle, trying to simulate natural phenomena.

The term “realism” is often used to designate not only a combination of features of a representation, but also the appraisal of the viewer of the representation as a very accurate simulation. We propose to make a distinction between an evaluation of the representation and the viewer’s experience, and use “realism” in this context only as an indication of the viewer’s experience, so as “perceived realism”. To indicate that a representation is objectively very similar to a real environment, the term “photorealism” is more correct.

Accuracy and abstraction, which are measurable features of the 3D models, are often considered as important determinants of the perceived realism of a visualisation. Accuracy reflects whether the elements of a visualisation that are relevant for the task and user have been modelled correctly. This may refer to the geometry of the environment, but also for instance atmospheric circumstances and vegetation. Accuracy does not have a univocal relationship with perceived realism. Sometimes the omission of features or mistakes in the representation may decrease the accuracy of a visualisation, but not the perceived realism. In general however, it is assumed that highly accurate visualisations are perceived as more realistic by users.

The term “abstraction” refers to the information content of the visualisation and its elements, so the selection of information included in the visualisation (Pietsch, 2000). This may overlap with accuracy, for instance when details are added to the environment and objects, such as textures, or more precisely modeled shapes. However, abstraction is often used as a concept reflecting the opposite of “realism”, as in the so-called realism axis (Figure 1.2; Kibria, 2008). In this case, realism is used as “photorealism”, meaning the extent to which the representation resembles a high quality photograph. “Abstract” means “disassociated from any specific instance” (Merriam-Webster.com, retrieved 26/08/10). According to this definition, abstract visualisations use generic techniques to indicate trees, houses, etc. from which a viewer may build a mental model of a real environment. Highly detailed representations can of course still be inaccurate.

![Figure 20: Visual representation in realism axis adapted from Dykes, MacEachren & Kraak (2005)](image)

*Figure 1.2 (Kibria, 2008). Different representations of trees, as examples of abstraction levels of entities, used in visualisations.*

Sometimes the level-of-detail (LoD) of a visualisation is used as an indicator of the abstraction level of a visualisation, as for instance in Figure 1.3 (Kibria, Zlatanova, Laure, & Dorst, 2009).
Whether a viewer perceives a certain level of detail as realistic depends on the viewing distance in the model, and other use and user related aspects such as task, and expectations.

Other concerns on the use of visualisations and validity for assessment are related to characteristics and effects of the medium itself. Whereas visualisations are supposed to be beneficial by removing ambiguity in interpretation by helping viewers to create a mental image of the represented area (Tress & Tress, 2003), there is evidence that 3D visualisations are perceived as more trustworthy and convincing than other, traditional representations (Sheppard, 2001), especially when they are very detailed and have a high level of “photorealism” (Zanola, Fabrikant, & Çöltekin, 2009). Moreover, this phenomenon does not only apply to laypersons. Recent experiments comparing response to architectural visualisations, both traditional (sketches and artist’s impressions) and computer generated visualisations (rendered 3D models and computer generated photomontage), showed that professionals as well as laypersons judged the perceived credibility of computer generated photomontage as higher than the other visualisations (Bates-Brkljac, 2009).

Most users and designers intuitively prefer realistic 3D representations of environments. Less abstract or more “realistic” representations are generally preferred in visualisations, even when they are not required for the tasks at hand (Smallman & St. John, 2005, Hegarty, Smallman, Stull, & Canham, 2009). Viewers unconsciously expect that the added realism in a representation, which makes the environment look familiar, results in near-effortless comprehension and provides an accurate assessment of the environment that is represented (Al-Kodmany, 2002). This faith in realistic displays, sometimes called Naïve Realism, is often misplaced, because for many tasks low-fidelity visualisation tools offer superior functionality and task performance (Ssmallman & St.John, 2005; Hegarty et al., 2009). This has raised questions, for example, about the ethics of using 3D visualisations for the purpose of convincing the general public. Any biases in data interpretation engendered by the medium should therefore be considered carefully in the context of planning and decision making.

In our opinion major deficits in current research on validity are that firstly, effort is focused on improvements in software, modeling techniques, and display systems, but that the evaluation of the response of the users, always mentioned as the predominant measure of validity, is seldom assessed separately. Responses to visualisations of future developments in an environment cannot be assessed directly, but the visualisation of the current situation can be subjected to testing; and for improvement of the models for further projects tests can be performed after the constructions are built or the changes in the landscape are effectuated.

Secondly, the validity of a visualisation can only be assessed, when the tasks for which it will be used, and the context in which the application is used, are defined accurately: the meaning and understanding of visuals may vary greatly depending on the knowledge domain, the roles of the users and the tasks that they perform (MacEachren et al., 2005).

It should be recognized that the affective appraisal of, and affective/emotional response to, a visualised environment must be considered and assessed separately. Accurate and detailed visualisations do not necessarily create the same “ambience” of a real environment. Vice versa,

<table>
<thead>
<tr>
<th>LoD 1</th>
<th>LoD 2</th>
<th>LoD 3</th>
<th>LoD 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block Model and no roofing</td>
<td>Models with roofing</td>
<td>Detailed textured model with roof</td>
<td>Detail models with interiors</td>
</tr>
</tbody>
</table>

*Figure 1.3 (Kibria et al. 2009) Four levels of detail in 3D models.*
features that influence the perceived ambience of a visualised environment do not automatically improve the accuracy of a visualisation.

Three concepts are pivotal in this research: validity, realism and accuracy. We defined them as follows: a valid simulated environment produces a cognitive, affective and behavioural response in the observer equivalent to the response produced by the real environment. Realism is the extent to which an observer experiences a simulated environment as real, and is often closely related to the affective appraisal of viewers. Accuracy is the extent to which elements of a visualisation have been modelled correctly, and may contribute to the validity of a virtual environment, especially concerning the cognitive aspects; however accurate and detailed visualisations do not necessarily create the same “ambience” of a real environment and lead to similar affective appraisals.

1.2.2 Virtual environments used for training or serious gaming

Virtual environments developed for training applications and serious games are the second category of applications used for this study. 3D environments on desktop computers are increasingly used to represent situations for training that cannot be created in the real world for reasons of safety, cost, time, etc. They can represent generic environments (a hospital, a prison), or specific locations like one distinctive oil refinery in Rotterdam or a specified U.S. naval ship. They generally include simulations of the behaviour of relevant objects like equipment, or processes, such as a burning fire, or atmospheric conditions such as a storm. The simulations generally do not include a competitive element, because the trainee’s goal is to follow procedures or handle a complex incident. Instructors may play an important role in the training by (vocally) impersonating the avatars in a scenario, and by suggesting other options to solve a problem (e.g. Houtkamp & Bos, 2007). Simulations may be extended with game elements, to engage the user and (it is assumed) improve learning.

The term game refers to an activity pursued for entertainment and enjoyment (Smith, 2010). Many definitions exist that describe a game (cf. Garris, Ahlers, & Driskell, 2002; Vogel, Vogel, Cannon-Bowers, Muse, & Wright, 2006) but they generally include elements such as interest, challenge and fantasy, play and feedback; games are goal-directed, competitive (against the computer, another player, or oneself) and conducted within a framework of agreed rules (Graesser, Chipman, Leeming, & Biedenbach, 2009; Lindley, 2004; Prensky 2001).

A serious game is a “game in which education (in its various forms) is the primary goal, rather than entertainment” (Michael & Chen, 2006). The term was first used by Clark C. Abt in his book Serious Games (1970) who referred mainly to board and card games. It was introduced in the context of computer applications in 2002 by the Woodrow Wilson International Center for Scholars, Washington D.C. The center focused on the use of electronic games in policy education and exploring management and leadership challenges facing the public sector. In contrast with entertainment games, which also involve some type of incidental learning or skills, serious games “endorse intentional learning according to predefined learning goals, which can be implicit as well as explicit” (Bente & Breuer, 2009).

Simulations were first developed for defense establishments and military forces (Stone, 2002). Presently, desktop gaming and simulation environments are increasingly being used in military contexts for training, instruction and simulation; for example for operations in unfamiliar (foreign) urban areas and cultures in situations where emotions influence performance (Chatham, 2007; Zielke et al., 2009), or to simulate asymmetric warfare (Mayo, Singer, & Kusumoto, 2005). Other important domains are education, healthcare, and training of security personnel.

User experience in games, and the concept of presence.

The popularity of games, and the expected effectiveness in learning, is related to the “gaming experience” of the players, and their affective/emotional response to the application. The success of simulations and serious games is often attributed to the convincingness of the situation that is represented, in which the trainee can experience the events as real. From this point of view, it is important to develop environments that support the intended affective response of the user; if the
trainee is drilled to perform in a stressful situation and experience a high level of arousal, the virtual environment must support the arousing qualities of the training. Research on the affective response of the user generally does not consider the environment separately, but includes all components of the training, such as the events in the training, and interaction with the game. The environment in which the events and interaction take place however, functions like a stage on which the trainee performs. The affective qualities of this stage may, for instance, create a specific ambience; the environment may be perceived as realistic, or, on the contrary, may not be appropriate to the seriousness of the events at all and distract the user.

Frequently used in the context of games is the concept of engagement, which we explore in one experiment (Chapter 4.5). Studies have shown that learner engagement is of utmost importance to learning success (Lim, Nonis, & Hedberg, 2006). There are many definitions for the term engagement, but all entail mindfulness, intrinsic motivation, cognitive effort, and attention. Engagement thus refers to the (emotional) state in which continuation of the activity doesn’t require any effort, and allows a user to keep the attention focused on the events and required actions. Engagement is also necessary for optimal performance (Garris et al., 2002) and to induce a state of flow. Flow is experienced when the user’s concentration is so intense that (s)he does not notice fatigue or the passing of time (Graesser et al., 2009).

Virtual environments, especially when used in games, can elicit feelings of presence (Lee, 2004; Lombard & Ditton, 1997), defined as the experience of “being there” in a mediated environment (IJsselsteijn & Riva, 2003). There are at least three discernible dimensions of presence (Konijn & Bijvank, 2009), of which spatial presence, a sense of being physically located in a virtual environment (IJsselsteijn, de Ridder, Hamberg, Bouw huis, & Freeman, 1998) is the most important for our study. The other two dimensions are social presence and self-presence. Immersion is sometimes used to denote specifically the physical extent of the sensory information in a virtual environment leading to a sense of presence (Baños et al., 2004).

Although presence is not necessarily due to quality or other features of the graphics, or sound tracks (Reeves & Nass, 1996), factors in the representation can contribute to the feeling of presence. Reversely, if the environment distracts the users, it negatively influences their engagement with the environment and events, and in turn may diminish their feeling of presence. The affective qualities of the environment should be supportive of the graveness of the incident in a scenario. If a training scenario is designed to be arousing, the virtual environment should be designed to increase this emotional response of the trainee.

A training that aims to prepare trainees to perform effectively in a stressful environment, for instance for practicing decision-making skills in high affect or dangerous situations, is called stress exposure training (Tichon, 2007). This type of training has been used successfully across many work domains, e.g. the military, aviation, and health care (DeCarvalho, Freiere & Nardi, 2010).

Our assumption for now is that the affective qualities of a virtual training environment can contribute to the engagement and presence of a trainee. When the affective qualities of the environment do not match the expectations of the trainee and the danger or graveness of the incident they may have a negative influence on these aspects of the user experience. We will return to the relation between the training environment, presence and emotions in Chapter 5.

1.2.3 Selected platforms and applications

If an application is developed for the purpose of providing an accurate simulation of an environment, it would seem obvious to choose a type of display that occupies a large part of the field of view of a user, or can even immerse the user. In reality the environment surrounds us completely, and the experience of this space, whether large or small, is a very important factor in the appraisal of the environment.

Platforms such as the CAVE, head mounted displays (HMD's) and projection domes, have become commercially available in the last 10 years, and virtual environments can now relatively easy be implemented on these display devices. Nevertheless, most visualisations for planning and design,
as well as environments for training and simulation, are still developed for desktop computers. Companies choose desktop displays and projection screens because of budget, modelers’ skills, the wide availability and ease of use of these displays in different locations, and clients’ familiarity with these displays and interaction devices. For communication of design plans, relatively simple models are chosen because they can be made accessible in internet applications for large numbers of users.

In our experiments we studied the applications as they are presented to users in real life situations, on monitors or projection screens. The participants used interaction devices matching the display type, so keyboards and mouse, or, in some of the training applications, a gamepad. One of the studies was performed on location, to include the influence of factors in the context of, in this case, the training application. When sounds were provided in the original 3D model, they were played with the speakers or headphones normally available with desktop systems.

Consequently, this research approach focuses on applied settings rather than fully controlled lab settings. It requires that variables and effects are studied on an aggregate level and sometimes include (minor) variables that are difficult to control. These choices together enhance the usefulness for practitioners, and the external validity of the studies performed.

### 1.3 Research domain, goals, research questions and research methodology

The theory underlying this study belongs to the domains of environmental and cognitive psychology, human perception, human-media interaction, and media theory. This dissertation contributes to research on visualisation and virtual environments in a number of ways. In the first place we emphasize the importance of the viewers’ affective/emotional response. In environmental psychology the validity of media to display urban environments or landscapes for evaluation has been discussed extensively, and different dimensions in the appraisal of virtual environments are acknowledged. Mahdjoubi and Wiltshire (2001) for instance discern three core factors as being relevant to simulation evaluation and decision making: incentive – motivational, affective and cognitive functions. However, studies have either described prerequisites for the validity of visualisations on a general level, or examined specific virtual environments, but have not distinguished the affective qualities of a virtual environment as a separate requirement as we have done.

Other new contributions to visualisation research are the focus on the affective appraisal of desktop virtual environments, which are widely used, and the structured approach to studying the effects of visual and auditory features on the affective/emotional response to desktop virtual environments.

A basic principle repeatedly referred to in the next chapters is that characteristics of an environment may influence the viewer’s affect and emotions. (Core) affect refers to the free floating mood we consciously experience (Russell, 2003). Oatley, Keltner and Jenkins (2006, p. 29) have summarised the prevailing insights in emotion research and propose that one may treat emotions “as multi-component responses to challenges or opportunities that are important to the individual’s goals, particularly social ones.” The environmental characteristics that determine viewers’ affect and emotions are called affective qualities of the environment. When the viewer appraises the environment, the process is called affective appraisal; the assessments themselves are called affective appraisals. The viewer may experience a change in affect and emotion as a result of viewing the environment, an affective/emotional reaction. We will refer to affective appraisal and affective/emotional reactions together as affective/emotional response.

It is important to realize that 3D models are representations of an environment. There is a distinction between media content and media form, as explained by for instance Reeves and Nass (1996), Reeves, Kim and Tatar (1999), Västfjäll, Larsson, and Kleiner (2002), Lee and Peng (2006). Most virtual environments used for visualisation and training, and all environments used for experiments in this study, offer visual information, no or little auditory information, and no other sensory information. They are also very static; dynamic features are limited. The medium is not transparent but influences our perception of the represented environment, and our response.
Viewers respond to features of the representation, as well as the represented environment. As will be discussed in later chapters, media forms and media contents interact with each other, and often the responses cannot be ascribed to either form or content alone.

An example of the effect of the medium is that screen size influences attention and arousal of viewer (Reeves et al., 1999); an example of the effect of absence of other sensory information is that odours and temperatures influence the affective response of viewers (Bell et al., 2001). Personal characteristics, for instance experience using the medium, also modify the affective/emotional response of viewers to representations (Bates-Brkljac, 2007). As a consequence, the affective qualities conveyed by virtual environments are not (automatically) similar to those of a real environment. They are often coincidental, or unintentional, and not necessarily a result of requirements and deliberate modeling. Modelers are generally unaware of the importance of affective qualities for the validity of a virtual environment that is used for a specific goal. Their understanding of the mechanisms to convey or assess the ambience of a place is often implicit, and based on personal experience and preference (Macfarlane, Stagg, Turner, & Lievesley, 2005).

Moreover, there is a general tendency to use new techniques that provide visually attractive effects, but do not necessarily improve the validity of a representation (Lange, 2001). Viewers are unaware of the importance of affective qualities of environments and are generally not able to separate their assessments of the representation of the content, and of the represented environment.

Ignorance of these mechanisms of both developers and viewers may lead to visualisations that are not valid for assessment of ambience or aesthetic qualities, and lead to environments for training that do not support the learning goals. As mentioned earlier, 3D models are often developed for a specific project, but are implicitly expected to function as a substitute for a real environment, for other goals. If in these projects requirements and task analysis are performed, expectations of the models are high; tasks are many, varied, and not defined accurately by the client. Different tasks however require different content and representations. Consequently, the exact fit between a specific task or aimed use, and representation is crucial.

Our first goal was to assess the effects of visual and auditory features in the content and representation of 3D environments, on the affective qualities of the environment, and on the affective/emotional response of the viewer. We endeavoured to reveal the underlying psychological dimensions of affective qualities of 3D environments, and how these dimensions are moderated by various characteristics of the environment, by features of its representation, and by characteristics, mental models, and state of the user. In Chapter 2 we further discuss relevant research on affect, emotions and affective/emotional responses.

The second goal was to derive features of virtual environments that were expected to be determinants of affective appraisal from literature and observation, and organise them in a comprehensive framework. This framework can be used in research, and in developing or adapting virtual environments when the affective/emotional response of the user is important for the validity of the 3D model.

Furthermore, our goal is to offer guidelines and designs for validation studies for visualisations of environmental designs and planning, and virtual environments for training. We also discuss the usefulness of verbal scales and other measuring instruments, for this specific purpose.

Throughout the research for this dissertation we tried to answer the following questions.

1.3.1 Research questions

Main research question

*How do visual and auditory features in a virtual environment on a desktop computer influence the affective appraisal and affective/emotional reaction of the user?*

The main sources of information in desktop virtual environments are visual and auditory. The reduced quality and quantity of this information, and the absence of other modalities, will impact the appraisal of the environment. We strive to understand the extent of this impact, and to determine which features are predominant in a given desktop virtual environment.
**Research Question 1**

*How do users appraise virtual environments, and how is this different from the appraisal of real environments?*

This question is essential for the validity of virtual environments, particularly for environmental visualisation. To answer this question we examine theories of environmental psychology and studies in landscape and urban visualisation (Chapter 2). We describe experiments conducted in our studies, focusing on specific elements of the 3D model, or on a specific type of environment (Chapter 4).

**Research Question 2**

*How do personal characteristics, momentary user states, and context of use moderate the effect of visual and auditory features on the affective appraisal of a virtual environment used for environmental planning or for training?*

Viewers, tasks and context of use of desktop virtual environments differ widely and moderate the effects of features of the virtual environment on the affective appraisal of the user. We delineate the role of mental models in this process, and describe how affective appraisal and reactions may be influenced by personal and contextual factors, in Chapter 2. We included these factors in the design of the experiments, where possible (Chapter 4).

**Research Question 3**

*How can features of environments and the representation, and personal and contextual moderators, be organised into a framework of determinants of affective appraisal?*

The determinants of affective appraisal are numerous and may influence each other. To structure these determinants, their relations and their effects, in Chapter 3 we develop a framework including the most important determinants of the virtual environment and the viewer. This framework extends existing frameworks on the evaluation of environments; it also allows analysis and comparison of the results of experimental research, such as the experiments described in Chapter 4.

**Research Question 4**

*Which measuring instruments are sufficiently sensitive and accurate to assess affective/emotional responses to virtual desktop environments?*

In the course of this study we used and adapted different measuring instruments. Scales designed for assessing affective qualities of environments are not yet tailored for virtual environments. They for instance do not distinguish reactions to features of the representation, such as low quality graphics, or small dimensions of displays. In Chapter 3 we explain the available measures, and in Chapter 4 we use a number of instruments in the experiments. Final conclusions on the validity of the tools will be discussed in Chapter 5.

**1.3.2 Research methodology**

Research methods and techniques used, were derived from experimental research, environmental assessment and human-media interaction. In many ways, this research complies with studies in environmental psychology. Environmental psychology is “the study of the molar relationships between behaviour and experience and the built and natural environments” (Bell et al., 2001, p.6). It distinguishes itself from other fields of psychology in both perspective and methodology (Bell et al., 2001). First, it puts an emphasis on studying the environment-behaviour relationships as a unit, which explains the term “molar” in the definition. Environmental stimuli are not separate from the observer; by their presence humans influence the environment which creates a relationship. When a specific environment-behaviour relationship is studied in a field or laboratory setting, it is assumed from the start that “such dissection of an integral unit cannot tell the whole story” (Bell et al., 2001). Moreover, in environmental psychology stimuli are considered to be not distinct from each other, but to occur in clusters; in other words, we do not, typically, experience singular environmental characteristics – they tend to co-occur (Winkiel, Saegert, & Evans, 2009). Another characteristic relevant here is that in environmental psychology the distinction between
applied and basic research is less sharp. This means e.g. that researchers often chose settings for study outside the laboratory, and employ an eclectic methodology.

The first part of this research was exploratory. It can be considered as inductive research, starting bottom-up with observations and measures, detecting patterns and causalities and leading to tentative hypotheses and a framework of determinants of affective appraisals of virtual environments. Studying the use of virtual environments over a period of time and in many different contexts, we observed that virtual environments on desktop computers were used for assessment of affective qualities of the represented environment, whereas the applications were not designed or evaluated for this purpose. After analysis of the characteristics of hardware, software, and development, which contribute to the transformation and often degradation of the virtual environment, a pattern emerged. This was developed into a framework of determinants of the affective appraisal of, and reaction to, a virtual environment. Theories from environmental psychology on the affective qualities of real environments, and from media psychology, were used to adjust the framework and elaborate into an operable scheme for further research (Trochim, 2006).

The framework was adapted to structure the important modifications that occur as a result of the process of creating, displaying and interpreting 3D environments, and to handle the complexity of the research material (Chapter 3, 4, and Houtkamp, 2005).

The second part consisted of experiments testing elements of the framework, with specific hypotheses. These led to statements on the impact of the individual elements and on the correctness of the framework and generalizability of the research findings. In the experiments both quantitative and qualitative measures were used, to reveal not only effects of determinants on the viewer’s response, but also to understand the relationship between determinants and responses, and identify mediating or moderating variables. When possible, and relevant, the experiments were set up to resemble the context in which the 3D environments would be used in reality.

Finally, adjustments were made to the framework and have led to general conclusions and answers to the research goals (Chapter 5).

1.4 Outline of the thesis

Chapter 2 (Real and Virtual environments: Affective appraisal and affective/emotional reaction) provides an overview of the most important theories on affective appraisal and affective/emotional reactions to real and virtual environments. It also considers mental models of environments that people compose, and the role of the representation in the appraisal of a 3D environment.

Chapter 3 (Affective appraisals of virtual environments: A framework for empirical research) proposes a framework for assessing the affective qualities of a virtual environment. It classifies characteristics of 3D models and of the users of these models, which may influence the affective appraisal of, and reactions to the represented environment. It explains methods of measuring affective/emotional responses in experimental and field research.

Chapter 4 (Experiments on the affective/emotional response to virtual environments). In this chapter the experiments are described that were performed for this study. The experiments concern a number of variables of the represented environment, the representation, as well as user states and user characteristics, which are response moderators. They measured the effects of the manipulations of these variables on the affective appraisals of the viewers, of their affective/emotional reactions, and/or their performance.

Chapter 5 (Conclusions and discussion) discusses the results of the experiments of Chapter 4 in relation to each other and to earlier research. Some general conclusions can be drawn about the effect and strength of certain features in the representation and content of the 3D environment. The utility of measuring instruments for assessing affective/emotional responses to an environment is discussed. Finally, we provide recommendations for developers of 3D visualisations, and of environments for training and serious games.
2 Real and Virtual Environments: Affective Appraisal and Affective/emotional reactions

“They’re immense stretches of wheatfields under turbulent skies, and I made a point of trying to express sadness, extreme loneliness.”

(Vincent van Gogh, about his painting “Wheat field under threatening skies”. Br. 1990: 903 | CL: 649 in a letter to Theo van Gogh and Jo van Gogh-Bonger Date: Auvers-sur-Oise, on or about Thursday, 10 July 1890)
2.1 Introduction

Environments influence our mood, emotions and behaviour. Places can be experienced as uplifting, depressing, consoling, scary, exciting, or boring. The affective qualities of places can attract humans and make them linger, or distress them, make them leave and never come back. Virtual environments are often expected to operate in the same way, and elicit similar responses in viewers. To understand the complex matter of humans’ affective/emotional responses to real and virtual environments, we delve deeper to reveal the underlying mechanisms of this relationship.

This chapter explains the concepts and theories that are used throughout this thesis and are the basis for the experiments described in Chapter 4. In the next paragraphs we discuss:

- affect and emotions;
- affective appraisal and affective/emotional reactions (together referred to as affective/emotional responses);
- affective/emotional responses to real environments;
- mental representations of the environment, or schemas;
- and affective/emotional responses to virtual environments.

The study of the affective appraisal of real environments belongs to the domain of environmental psychology. Research in this interdisciplinary field was very productive in the period from ca. 1970 to 1990, after which research seemed to slow down. In recent years, renewed interest in the relationship between the environment and human health and wellbeing, and between retail environments and shopping behaviour, has given the field a new stimulus (e.g. Nasar, 2008). The use of computer visualisations for representing environments for the purpose of planning or landscape assessment has generated interest from researchers with a background in computer graphics, human computer interaction, and GIS (Geographic Information Systems). This has led to a new type of studies, often focused on the validity and utility of 3D models in environmental planning and assessment. However, this topic also requires knowledge of other branches of psychology, concerned with affect and cognition, and of media studies. This chapter brings together theory from these domains and, with Chapter 3, lays the foundation for the experiments described in Chapter 4.

2.2 Affect and emotions

2.2.1 Introduction to affect and emotions in psychological research

In psychological research, perception, cognition and emotions are often separated to facilitate theorizing and experimental design. They are however processes that depend on each other and exert bidirectional and unidirectional influences (Franz, 2005; Meier, Robinson, & Clore, 2004).

Research on environmental experience makes a distinction between perceptual and cognitive processes, investigating how individuals acquire, retain, and process information in environments, whereas studies of affective/emotional appraisal examine how individuals respond to, and evaluate, environments (Hanyu, 1997). Although we study the effects of virtual environments in real-life context, in which perceptual, cognitive and affective processes continuously influence each other, we also make this artificial distinction here. It not only stresses the focus of our research questions, but exposes a friction that may occur when the 3D models are used for tasks that are either primarily affective, or primarily cognitive, such as spatial tasks. For instance, to make a viewer feel in awe when looking up at a high building might require exaggerating the height of the building in the model to compensate for the limitations of projections on a monitor, such as the display size, the context of viewing, etc. This of course decreases the accuracy of the model and its utility for tasks such as assessing dimensions of the 3D environment.

Although “emotion” is a commonly used word in everyday language, scholars and scientists have struggled for decades to come to a common definition (for instance Izard, 2010; Kleinginna & Kleinginna, 1981; Scherer, 2005). As Young (1973) puts it, “almost everyone except the psychologist
knows what an emotion is ....”. Izard (2010) recently conducted a survey among scientists of their definitions of the term “emotion”. He concluded that although there is to some extent agreement on emotion-activation, functions, and regulation, there is no consensus on a definition. The collected data suggest that it cannot be defined as a unitary concept. Izard’s combination of the responses of the scientists resulted in the following description:

Emotion consists of neural circuits (that are at least partially dedicated), response systems, and a feeling state/process that motivates and organizes cognition and action. Emotion also provides information to the person experiencing it, and may include antecedent cognitive appraisals and ongoing cognition including an interpretation of its feeling state, expressions or social-communicative signals, and may motivate approach or avoidant behaviour, exercise control/regulation or responses, and be social or relational in nature. (Izard, 2010, p.367)

Emotions are modes of relating to the environment, and have a function in responding to events by organising adaptive responses to stimuli that are important for the survival and well-being of the organism (Brosch, Pourtois & Sander, 2010; Ekman, 1992b; Lazarus, 1991). Oatley, Keltner and Jenkins (2006) propose that one may treat emotions “as multi-component responses to challenges or opportunities that are important to the individual’s goals, particularly social ones.”

An important distinction is made between an emotion and mood. In psychology, the term “emotion” or “emotional episode” is used for short-term states that typically have an object, meaning that they are directed to something or someone. Physiological responses accompanying emotions may last only a few seconds, whereas people who are prompted on their emotions report experiences between a few minutes and a few hours. These episodes are therefore more transient than moods. Moods are states which may last for weeks, of which the start and end are often hard to assess, and are often objectless or free-floating (Frijda, 1993).

In everyday vocabulary as well as psychological literature, the definitions of affect and emotion are blurred and indistinct. The Merriam-Webster online dictionary explains affect as “the conscious subjective aspect of an emotion considered apart from bodily changes; also: a set of observable manifestations of a subjectively experienced emotion”. (Merriam-Webster.com, Retrieved on 22/10/10). In this thesis we use “affect” to refer to “core affect”, the free floating mood we experience. This concept is more fully defined by Russell (2003) and explained in the next paragraph. In Russell’s theory core affect may at some point be attributed to a cause and begin a more intense emotional episode. Since this distinction is not very marked, many authors (e.g. Schwarz & Clore, 2007; Villani, Lucchetta & Preziosa, 2009) use the terms affect and affective state broader and more inclusive, encompassing emotions, mood and some aspects of motivational processes. In contrast, some psychological studies use the term affect more specific, referring to all types of subjective experiences that are valenced, that is, experiences that involve a perceived goodness or badness, pleasantness or unpleasantness (Desmet & Hekkert, 2007; Schwarz & Clore, 2007).

We conclude that there is no consensus on the definition of emotions and affect. We will use affect and affective state to refer to core affect, the free floating mood we experience. Emotion or emotional episode is used for short-term states that typically have an object, meaning that they are directed to something or someone, for example fear of an attacker.

The environmental characteristics that determine viewers’ affect and emotions are called affective qualities of the environment. When the viewer appraises the environment, the process is called affective appraisal; the assessments themselves are called affective appraisals. The viewer may experience a change in affect and emotion as a result of viewing the environment, an affective/emotional reaction. We will refer to affective appraisals and affective/emotional reactions together as “affective/emotional responses”.

This paragraph presented a rather abstract overview of concepts relating to emotions and affect. The next paragraphs discuss emotion theories further, and will provide a more concrete and comprehensible description.
2.2.2 Emotion theories and mechanisms underlying emotions

Theories explaining the nature and structure of emotions and the underlying mechanisms can be sorted into a number of categories. The simplest classification is suggested by Lopatovska and Arapakis (2011). They first distinguish between somatic theories and appraisal theories that explain about the nature of emotions and their relationship to emotion stimuli; and further between discrete and continuous approaches that concern the structure of emotions. Other authors use different classifications, for instance Brosch, Pourtois and Sander (2010), who also distinguish constructivist theories (Barrett, Lindquist & Gendron, 2007). We limit the discussion to the theories that are relevant for our research.

Somatic or affect program theories (Moors, 2009) claim that each basic emotion has a unique neural circuit, and that these circuits are installed by evolution to serve specific adaptational functions. These theories are supported by for instance Ekman (1992a, 1992b), Izard (2009), Panksepp (1998) and Zajonc (1984). Ekman (1984) for example suggests that emotions’ primary function is to mobilise an organism to respond quickly to prototypical events, similar to those that were encountered in the past. Studies in this field often seek to describe emotional expressions and perceptions of emotional expressions, such as facial expressions (Izard, 2009).

The basic premise of appraisal theories of emotion is that emotional processes are a result of the continuous, subjective, evaluation or appraisal of objects, behaviours, events and situations by an individual with respect to their relevance for his/her needs, goals, values, and general well-being (Ellsworth & Scherer, 2003). The cognitive activity involved in this process can be conscious or unconscious, intentional or unintentional and take a form of a judgment or a thought (Lopatovska & Arapakis, 2011). Humans evaluate events along a number appraisal dimensions, such as goal relevance, goal congruence, certainty, and coping potential (Lazarus, 1991; Oatley & Johnson-Laird, 1987; Scherer, 2005). The values on these variables combine to form an appraisal pattern. Because a stimulus is assessed according to its importance for the individual specific situation and context, it gives room to differences between, e.g., species, age groups, personal dispositions, and cultural contexts. The basic prediction is that emotions are associated with particular combinations of appraisals (Tong, Ellsworth, & Bishop, 2009). Unconscious appraisal of stimuli takes place prior to the emotion, whereas conscious attribution of the emotion to a cause and/or labeling of the emotion (for instance fear or anger) may also take place after the emotion occurs (Moors, 2009). A recent view is that appraisals are not only antecedents of emotions but that emotions can also shape appraisals; therefore Ellsworth and Scherer consider emotions and appraisals as components of the same affective state (Ellsworth & Scherer, 2003; Tong et al., 2009). All appraisal theories however share the assumption that emotions are related to the appraisal of events, or objects.

The theories may be combined by suggesting that there are two stages in the appraisal process (Oatley et al., 2006). The primary appraisal then is an immediate and unconscious evaluation of whether the stimulus is good or bad (Zajonc 1980; Mischel & Shoda, 1995). Russell (2003) assumes that these first automatic evaluations give rise to our core feelings of positivity or negativity. Other, secondary systems provide in a second stage more deliberate, conscious, complex assessments in terms of such matters as what caused the event and what to do about it, and give rise to specific emotions (Oatley et al., 2006).

A fundamental controversy exists relating to the structure of emotion. Basic emotion theories (also called discrete approaches) assume that a number of separate, discrete emotions such as fear, anger and happiness, are innate in humans (Ekman, 1992b). Other emotions are considered to be combinations of the basic emotions or as socially learned variants such as grief, and guilt (Bower, 1992). The basic emotions cannot be reduced to lower level components, although they can be labeled superordinately as pleasant or unpleasant. Certain classes of stimuli trigger predefined affect programs, which then elicit emotion-specific response patterns such as prototypical facial expressions and physiological reactions driven by specific neural response systems (Panksepp, 1998).

Dimensional theories (or continuous approaches) are eventually derived from Wundt’s model, developed at the end of the 19th century (Scherer, Dan, & Flykt, 2006), and oppose the notion of a
limited number of basic emotions. Dimensional theories claim that stimulus processing leads to an evaluation that distinguishes between two (or three) basic dimensions: positive and negative stimuli also referred to as valence or pleasure, and between stimuli that lead to different degrees of activation or arousal. A third dimension, dominance - submissiveness (or potency/control) is recognized in many studies and supported by several researchers, for instance Bradley and Lang (1994), but is still a topic of discussion (see for instance Yani-de-Soriano et al., 2006). Fontaine, Scherer, Roesch and Ellsworth (2007) even suggest a fourth dimension, unpredictability.

In Russell’s model, the two major dimensions pleasure and arousal are combined in a circular two-dimensional space (Russell, 1980; Russell, 2003), shown in Figure 2.1. The horizontal axis represents pleasure or valence (from unpleasant to pleasant), and the vertical axis represents arousal (from calm to excitement). At any moment in time, an individuals’ affective state can be located at a single position within the map. Affective states that are similar in nature are positioned close to each other in the circumplex.

![Circumplex Model of Affect](image)

**Figure 2.1 (Russell, 2003). The circumplex model of affect, showing two dimensions pleasure and arousal. A conscious experience of one’s affective state can be described as an integral blend of the two dimensions and thus as a single point in this map.**

The exact values of the variables arousal and valence can be considered as properties of stimuli, of neurophysiological states, and of conscious experience (Moors, 2009). The combination of values on both variables is called ‘affective quality’; the affective quality of stimuli causes in the person a specific state of core affect. Russell (2003) describes the affective state of an individual, or core affect as follows: “A neurophysiological state that is consciously accessible as a simple, nonreflective feeling that is an integral blend of hedonic (pleasure–displeasure) and arousal (sleepy–activated) values.”

Core affect influences reflexes, perception, cognition, and behaviour and is influenced by many causes that are internal and external, but people have no direct access to these causal connections. Therefore, core affect can be experienced as free-floating (called mood) or can be attributed to some cause (and thereby begin an emotional episode). Mood is defined by Russell (2003) as a prolonged core affect with no intentional object although he recognizes it as fuzzy, since neither duration nor degree of stability is defined. An emotion is the attribution or categorisation of core affect into one of the so-called emotion categories (such as anger, fear, sadness and joy) which are not given in nature, but are socio-cultural constructions. Russell’s theory is therefore considered a constructivist theory (Moors, 2009).

Russell and Feldman Barrett (1999) use the term prototypical emotional episode to refer to what, according to their view, most people consider the clearest cases of emotion, for example fear of an accident, rage at an attacker, or sadness because of a loss. A prototypical emotional episode is a complex set of interrelated subevents that is always concerned with a specific object. The object is real or imagined, in the past, present, or future.
The appraisal and dimensional theories of emotion theories fit our research most readily, because they are congruent to theories prevalent in environmental psychology and are used in media studies (Lopotovska & Arapakis, 2011); they also provide the theoretical basis for measuring the affective response to environments. In the next paragraph we explain the distinction between appraisals and emotions within the framework of these theories.

2.3 Affective appraisal and affective/emotional reactions

Russell (2003) stresses the affective component in the appraisal process. According to this view, objects, events, and places (real, imagined, remembered, or anticipated) enter consciousness affectively interpreted. The perception of affective qualities is a ubiquitous and elemental process, and cannot be cleanly distinguished from perception of nonaffective qualities.

Affective appraisals are the attributed emotional or affective qualities, or cognitions about possible object- or place-elicted emotions (Franz, 2005; Nasar, 2008). They are also verbal labels of these qualities or characteristics of stimuli.

“Affective appraisals are (...) judgments concerning the capacity of the appraised object to alter mood. The English language provides hundreds of words such as exciting, boring, disgusting, stressful, and relaxing that people can use to describe the affective qualities of places. Our thesis is that any such word could be rendered excitement producing, boredom producing, disgust producing, and so on.” (Russell & Snodgrass, 1987, p.249)

The affective quality of an object can be described in the same two dimensions as core affect (Russell, 2003).

Affective appraisals are not necessarily accompanied by an affective change or reaction (Russell and Snodgrass, 1987). As explained earlier, according to Russell (2003), the affective state of an individual, or core affect, can be experienced as free-floating (then called mood) or can be attributed to some cause (and thereby begin an emotional episode). Core affect can change without reference to any external stimulus. Conversely, a stimulus and its affective quality can be perceived without a change in core affect; the affective perception may then be called cold and detached. Whether or not this happens, and to which extent, is a complex question: it depends on the viewer’s affective state and characteristics such as personality, the strength of the stimulus, and other factors in the environment. An event or environment must be relevant for a viewer to elicit any emotion, whereas an affective appraisal may be “cold cognition”, or just a label produced by a viewer. The prerequisite of relevance is included in basic definitions of emotions, summarized by Oatley et al. (2006, p. 29) “as multi-component responses to challenges or opportunities that are important to the individual’s goals, particularly social ones.”

So, to perceive affective quality is to represent (for instance verbally) core affect, rather than to experience its effect on core affect. The two processes are however linked empirically, as in mood-congruent priming and misattribution of mood, described later in this chapter. The notion that core affect may be influenced by properties of stimuli in the environment, but also itself influences perception and cognition, is an important presumption in our research.

During the first phase of our research, we concentrated on the affective qualities of represented environments, and in experiments we assessed the affective appraisals of these environments by participants (Chapter 4.2- 4.4). These are mainly measured using verbal scales describing the ambience of an environment.

In the second phase we also assessed the related affective/emotional reactions, i.e. the changes in affective state or emotion of the participants as a reaction to the environments (Chapter 4.5 -4.7). In these experiments, we applied measuring instruments relating to the affective/emotional state of the participants.

Independently, the affective state of the viewer may influence the appraisal of an environment. We therefore performed experiments to reveal the effects of two important states of a viewer, stress and cybersickness, on the affective appraisal of a virtual environment (Chapter 4.8 and 4.9).
2.3.1 Engagement

One of the main goals of many virtual environments, especially games and virtual training applications, is to engage or involve the user. Engagement is a widely used concept in studies of game experience and interaction design, because it is thought to improve the performance of a user and to enhance learning (Garris et al., 2002). Engagement is a subjective experience and has affective components, but is not considered to be an emotion (Agarwal & Karahanna, 2000). Despite, or as a result of, the recent burst of research into subjective experiences during game play, no definitional agreement can be found in the literature on the terminology (Bianchi-Berthouze, Brien, & Toms, 2008; Garris et al., 2002; Jennet et al., 2008; Mallon & Webb, 2000; Rozendaal, Keyson, & de Ridder, 2007). Rozendaal, Keyson, de Ridder and Craig (2009) summarize the term engagement used in literature on interaction design, as “an exciting and enjoyable state of mind in which attention is willingly given and held”. Engagement, flow, immersion, presence and motivation are often used as indicators of similar or related concepts, and often the term “engagement” is used as a generic indicator of game involvement (Brockmyer et al., 2009).

Obviously it is difficult to relate engagement in a simple way to other aspects of the affective experience explored in this research. For now we define engagement as a state of being involved in virtual environment, without effort, and without extrinsic motivation. We also limit the discussion of the relation between engagement and emotions to the assumption that there is a mutual relationship: pleasant surprising experiences in a game are for instance thought to increase engagement; on the other hand personal involvement, interest or motivation, all aspects of engagement, are required for experiencing emotions.

The determining factors for engagement are user characteristics and environmental factors in the application; these also determine other aspects of the user experience in virtual environments and gaming such as immersion and presence (Baños et al., 2004). The factors are further discussed in Chapter 4.5, Visual dynamics and sounds in an outdoor environment.

2.4 Environmental psychology and affective/emotional responses to real environments

To understand viewers’ responses to virtual environments, we must first understand the relation of individuals with, and their responses to, real environments. How does the environment, or its features, determine the affective response of an individual? How are environmental features classified, and which factors or dimensions can be distinguished in the viewers’ response? Which psychological effects and personal characteristics modify the response?

Research on the affective appraisal of environments belongs to the field of environmental psychology, which, as explained in Chapter 1.3, distinguishes itself from other fields of psychology both in perspective and methodology. One characteristic is that this branch of psychology studies the “molar” relationships between humans and their environments (Bell et al., 2001), indicating the unity of the human-environment relationship. Environments are generally heterogeneous, contain countless elements, are dynamic, and do not have boundaries. They influence a person’s behaviour, and emotional response, often subtly and unconsciously. Russell and Snodgrass (1987) stress the emotional aspect of this complex relationship, and the pervasiveness of the emotional links to places: “Affective quality is the bottom line of the accounting of the many features in a place, and is, we believe, a guide for much of your subsequent relationship to that place – what to do there, how well it is done, how soon to leave, whether or not to return. Afterward, you often remember little more about a place than its affective quality.” The importance and pervasiveness of affective qualities of places is obvious in our descriptions of the environments in which we live, work and recreate; they invariably contain many labels such as dull, depressing, open, inspiring, relaxing, impressive, peaceful, etc.

Affective qualities are characteristics attributed by individuals to objects, environments etc., and are remembered better than many non-affective qualities (Russell & Snodgrass, 1987). They constitute the atmosphere or ambience of a place, which is not an affective state, but a “subjective
experience through the perception of external elements and internal sensations” (Vogels, 2008a). Although personal characteristics such as age, culture, social class and previous experience influence the judgment of the atmosphere of an environment, there is a common structure across individuals (Hanyu, 1997), so a general effect of certain elements in the environment on the perceived atmosphere of this environment.

Environments contain numerous physical and atmospheric features that determine the affective qualities perceived by the viewer. Environmental psychologists have developed several typologies to structure, categorize, define and explain these features. However, these studies are mostly focused on a specific type of environment, natural or man-made, on a specific function such as retail environments (D’ Astous, 2000) and healthcare facilities (Dijkstra, Pieterse & Pruyn, 2006), and on the effects of specific factors in an environment. This makes it difficult to compare the results and build a generic model from the outcomes. General classifications are per force not very comprehensive, nor detailed.

The complex relationship between humans and their environment can be captured in the Stimulus-Organism-Response theory and model, as used by Mehrabian and Russell (1974) (Figure 2.2). The theory holds that the influence of a physical environment is mainly of an affective nature, and that an individual’s emotional responses to an environment can be described by three main variables: pleasure, arousal and dominance, the PAD dimensions. According to the PAD model, pleasure, arousal and dominance are the three basic emotional dimensions, which summarize the emotion-eliciting qualities of environments and mediate approach–avoidance behaviour in them.

![Figure 2.2](image)

This general framework has been elaborated by including modifying factors, and for specific environments. Bitner’s (1992) adaptation of this model for so-called servicescapes (service environments) has been used extensively and successfully in marketing research, to identify cues in the environment and contextual factors that influence clients’ shopping behaviour (Kearny, Kennedy & Coughlan, 2007). For our study on the affective qualities of virtual environments, we adapted Bitner’s model, and used it as a framework for our experiments. It will be further explained and discussed in Chapter 3.

Several theories have been developed over the last three decades for explaining affective responses to, and preferences for certain environments or environmental characteristics. They can broadly be divided into evolutionary theories and cultural preference theories. In evolutionary
theories, advocated for instance by Kaplan and Kaplan (1982), Kaplan and Kaplan (1989), and Appleton (1975), landscape perception and preference are shaped by human biological needs to survive. Kaplan and Kaplan’s (1989) information processing theory suggests that humans prefer environments that provide rapid, comprehensible information (Bell et al., 2001). On the other hand, humans like scenes that are engaging and provide interest. Kaplan and Kaplan distinguish coherence and legibility as preferred characteristics for understanding the environment, and complexity and mystery as factors that enhance involvement or interest in the environment. Appleton’s (1975) prospect refuge theory suggests that a landscape that provides an unobstructed open view that allows observation of approaching predators, and also provides refuge or protected settings for humans to hide, gives evolutionary advantages. Attraction to a particular landscape is therefore a biological condition. Because humans have a common evolutionary background there is a universal set of landscape features shaping preferences and affective responses for all humans (Tveit, Ode & Fry, 2006).

Cultural preference theories assume that social and cultural background, motives, previous landscape experience, and other personal attributes of the observer, determine the perception and experience of a landscape (for a discussion, see Bourassa, 1990; Home, Bauer & Hunziker, 2010). Recent approaches to landscape aesthetics now accept a combination of cultural and biological forces to explain landscape preference, arguing that landscape preference depends not merely on the physical characteristics of the landscape, but also on a person’s relationship and involvement with it (Arnberger & Eder, 2011; Tveit et al., 2006).

2.4.1 Dimensions and factors in affective appraisal

Several studies by Russell and collaborators using semantic differential techniques revealed two orthogonal dimensions of affective appraisals of environments, pleasure (from pleasant to unpleasant) and arousal (from sleepy to arousing). A 45° rotation of the axes produces two other independent bipolar dimensions: exciting-gloomy and distressing-relaxing (Russell, Ward, & Pratt, 1981; Nasar, 2008). The two basic dimensions are analogous to the dimensions in the concept of the circular space defined for core affect (e.g. Feldmann, Barret, & Russell, 1999) (Figure 2.3).

![Figure 2.3 A spatial representation of descriptors of the affective quality of environments (Russell 1980).](image)

In earlier studies, for instance Mehrabian and Russell (1974), a third dimension, dominance or the sense of control, was included besides arousal and pleasure. The dominance dimension ranges from extreme feelings of lack of control or influence upon one’s surroundings, to feelings of being influential or powerful (Yani-de-Soriano & Oxall, 2006). Osgood, Suci and Tannenbaum (1957) also established a factor called “potency” besides “evaluation”, and “activity”. Dominance was later rejected by Russell c.s. as a dimension because it did not emerge clearly in factor analysis in several
studies, and can be interpreted as not purely affective, but cognitive in nature (Russell & Pratt, 1980). Yani-de-Soriano and Oxall (2006) have recently reviewed the literature and conclude that dominance is as legitimate an environmental descriptor as pleasure and arousal. However, dominance has often been ignored since Russell and Pratt (1980), and semantic scales were developed that did not include adjectives in this dimension. Only a small number of studies have continued to include the dominance dimension, for instance Franz (2006) and Küller (2001), who assessed the effect of colours and spatial characteristics (such as spatial dimensions) of indoor environments on the affective response of viewers.

Vogels (2008a, 2008b) more recently developed a questionnaire to quantify the “atmosphere” of an environment as experienced by human observers, which has been tested in a number of environments, mainly shops (Custers, de Kort, IJsselsteijn, & de Kruijf, 2010). Factor analysis revealed two dimensions, “cosiness” and “liveliness”, which resemble the valence and arousal dimensions of Russell’s model. The questionnaire includes both terms that refer to the emotion or mood evoked by an environment, as well as terms that describe the atmosphere of the environment, supposedly “a more stable factor and less complicated variable than mood” (Vogels, 2008a).

These models of the appraisal of an environment or its atmosphere reflect the appraisal of the totality of the environment, and do not relate or refer directly to single objective, visual properties of this environment or elements therein. They do not clarify which elements in, or features of, the environment determine the appraisal, and why.

The measuring scale developed by Rikard Küller (1977, 1991) does aim to relate factors in the appraisal of environments directly to specific visual features and elements of the environment, as they are perceived by viewers. As a consequence, the factors are not purely affective but include descriptors of physical patterns, such as complexity, or openness. Küllers eight factors are Pleasantness, Complexity, Unity, Enclosedness, Potency, Social Status, Affection, and Originality. They are derived from an analysis of a large number of verbal responses to a variety of environments. The SMB-scale (Semantic Environmental Scale or, in Swedish, Semantisk MiljöBeskrivning) developed by Küller is based on these factors. Verbal scales used to assess features and affective qualities of environments will be further discussed in Chapter 3.

Other studies have provided relationships between sets of specific visual properties and affective appraisals of environments. A very direct method of relating specific characteristics of a certain physical environment statistically to judgments of preference or landscape quality is called the physical-perceptual approach (Bell et al., 2001). Examples of this approach are studies to assess the effects of characteristics such as differences in height, naturalness, and vegetation coverage, on the preference of viewers.

Although this approach provides knowledge of the effects of specific physical features in an environment on viewer’s appraisal, it does not lead to understanding of the affective/emotional response of humans to certain types of environments (parks, cities), or high level features of environments (dense vegetation, crowding). Other studies have therefore attempted to relate variables that are psychological-perceptual in nature, to appraisals. These variables are more abstract descriptions or features of environments, such as complexity and coherence. Bell et al. (2001) call this the psychological approach. Measures of the viewers’ response chosen are for instance preference, interest, and perceived safety (Hanyu 1997, 2000).

Results of studies on elements and features of environments that determine affective appraisal are discussed further in Chapter 3.
2.4.2 Theory of adaptation level, individual differences, mood congruence and misattribution of mood

The dimensions in environmental appraisals, and some effects of characteristics of environments on the affective appraisals of environments, have been established in many empirical studies. Individual and contextual variations may however influence the appraisal of an environment considerably. The factors that have most relevance for this research are outlined below.

Theory of adaptation level

One important modification is due to the change in scenes experienced by the viewer. Although during the assessment of an environment many important elements are stable, the viewer’s position is often not fixed. An individual may walk through an area or a building, perceive the environment from different viewpoints and thus experience a sequence of different spaces, shapes, and light conditions. A response to a scene is altered by prior exposure to a different scene in a way predicted by adaptation level theory (Russell & Lanius, 1984). This effect is used intentionally in architecture and landscape design. For instance, entering a large hall is a more impressive experience after passing through a narrow and dark corridor. The theory of adaptation level predicts that the more extreme the affective appraisal of an anchor environment on a dimension, the greater the magnitude and direction of shift found in the appraisal of the target environment (Russell & Lanius, 1984).

Individual differences

Individual personality traits, for example the arousal-seeking tendency, can influence an individual’s reaction to her physical surroundings (Mehrabian & Russell, 1987; Russell & Snodgrass, 1987; Lin, 2010). It has been noted often that architects and laymen have very different aesthetic judgments on buildings, which is probably related to their training. Gifford, Hine, Muller-Clemm, Reynolds and Shaw (2000) have found that these groups do not only have different appraisals of buildings, but also use different physical cues for their assessment, and that the laypersons produce more heterogeneous ratings (as a group) than architects. Gifford (1980) showed that building interiors appealed significantly different to lay judges depending on their age, sex, educational level, and mood. Variability in the experience of nature is large, and may occur not only between different people, but also for the same person over time (Schroeder, 2007). The response to an environment also depends on the visitor’s familiarity with the environment (Conniff, Craig, Laing, & Galán-Díaz, 2010) and the person’s plans (Ward, Snodgrass, Chew, & Russell, 1988; Snodgrass, Russell & Ward, 1988) and expectations (Bitner, 1992), often partially or wholly formed before encountering that environment.

Mood congruence and misattribution of mood

Viewers’ appraisals of an environment are also influenced by their current affective states. Niedenthal and Setterlund (1994) found that current moods and emotions can lead us to selectively perceive emotion-congruent objects and events. Participants of an experiment were either made happy or sad through music. Consistent with the theory of emotion-congruence, in a lexical decision task they were consequently quicker at identifying words that related to their emotional state. Mood-dependent effects on memory and other cognitive functions have been extensively described by for instance Bower (1981), Eich and Macauley (2000). Important for our research is the phenomenon that cues inducing a specific emotion may also elicit past experiences, images etc. that happened or were witnessed when the person was in a similar affective state.

Furthermore, moods and emotions may influence the appraisal of events and objects. Positive or negative emotional states, colour evaluative judgments as good or bad, even when the subject of evaluation has no relation to the cause of emotion (Oatley et al. 2006; Clore & Huntsinger, 2007). However, this effect only seems to occur if a person cannot attribute her affective state to a cause. “Without a salient cause, affect tends to be promiscuous, attaching itself to whatever is available, which is why moods can influence even irrelevant judgments” (Clore & Huntsinger, 2007). If a
person’s mood is improved, for instance when given a free gift, his or her liking for the environment also increases (Russell & Snodgrass, 1987). The person confuses the shift in mood he or she experiences, to environmental properties, which is called a misattribution.

In our first seven experiments (Chapter 4.2 – 4.7) we have manipulated elements which we assumed to have an effect on most viewers. In two other experiments (Chapter 4.8 and 4.9) personal response moderators have been taken as variables in the experiments.

2.5 Mental representations of real and virtual environments

When an individual visits a place, she has a goal or a plan, and has expectations about the place in relation to this plan, based on earlier experiences and memories (Russell & Snodgrass, 1987). An important hypothesis in this study is, that the appraisal of an environment is not only based on the environment the viewer perceives at a specific moment, but also on her mental representation; this is assumed to be a mix or combination of what the viewer perceives at that moment and a mental “image” or “schema”. Mental images are based on earlier visits and memories of the environment or are formed by visits to other places associated by the viewer, by films, literature etc. (Imamoğlu, 2009). They may include prototypic models of what a certain type of building such as a church should look like (Purcell, 1986). They are thought to contain elements of the world, their spatial relationships, and metric information, that can be accessed and mentally manipulated to some extent, for instance rotated, and viewed from different perspectives.

Most research on the content of mental representations has focused on the spatial knowledge contained therein, consequently referred to as “cognitive maps”, or “cognitive collages”, or “spatial mental models” (Tversky, 1993). However, Genereux, Ward and Russell (1983) already pointed out that the internal representation of a place includes more types of knowledge: apart from information about its objective attributes, also knowledge about its affective quality, and about behaviours that occur there. Also İmamoğlu (2009) refers to place schemas as carrying information about physical (sensory, spatial), social and organizational (rules of place, type, order and appropriateness of activities, etc.), purposive and affective qualities of places. Kim and Richardson (2003), researching the “image” of holiday destinations, define this image as the totality of impressions, beliefs, ideas, expectations, and feelings accumulated towards a place over time.

When visiting a location, the pre-existing representation or image is updated and the viewer gains more complex cognitive knowledge about the place (Kim & Richardson, 2003). There is evidence that as a result of an unconscious and continuous process, the mental representation is appraised and updated simultaneously (Kurby & Zacks, 2012). In visual perception, information flows
from the sense organs (eyes) to lower levels of the brain such as the visual cortex, and to higher brain areas such as the temporal cortex, where visual memories are stored (Bolls, 2002). This is called bottom-up processing. At the same time top-down processing takes place, in which information stored in long-term memory is transferred to lower brain areas, needed to make sense of initial visual information if it is unclear or incomplete. Bolls (2002) concludes that it is biologically plausible for lower levels of the brain, which contain information about spatial and temporal properties of objects, and higher levels of the brain, which contain memories, to simultaneously participate in the receiving and sending of information during visual perception.

Categorisation plays an important part role in this process. To move around in the world and respond to its challenges, humans must quickly make sense of their environment, which is generally very complex and in constant change. To cope with this, we create an internal mental representation of the stimuli that are immediately present in our surroundings. Brosch et al. 2010: “any given external object in the environment, the distal stimulus (e.g., a stone) is not processed as such, but is represented in the organism as a physical stimulation pattern on the senses, the proximal stimulus (e.g., the pattern of light on the retina reflected by the stone). Perception is the transformation of the proximal stimulus into a percept, the accessible, subjective, reportable experience that takes the form of an activation of a certain category in the mind (e.g., the accessible visual experience of the stone).”

By categorisation, the information complexity of the external world is reduced, because a number of objects or concepts are reduced to one group. In addition, from the association with the category, information about the stimulus is inferred. Therefore, the act of categorisation is critical to cognition and allows us to give meaning to the world (Brosch et al., 2010).

To make sense of the overabundant information embedded within environments, people use their existing place schemas (İmamoğlu, 2009). They are regarded as codified experiences that help humans deal with large amounts of perceptual information by creating generalized chunks that can continuously be modified when introduced to novel information. In this way schemas provide structure to perception and cognition of places, and help create place experience (İmamoğlu, 2009).

Categorisation and the use of place schemas help to explain why individuals tend to respond to an environment as a whole, instead of to individual features of the environment. They are able to label an area as pleasant, uncanny or somber without judging and weighing the individual cues separately. However, some stimuli in the environment are more important for our well-being or goals than others, for instance those indicating danger or threat (e.g. enemies and predators) or favourable conditions (e.g. for gathering food). They deserve priority and require that the process of perception, appraisal and response is accelerated. These have been called “emotional stimuli”. Emotionally salient material is remembered more easily than neutral material (Oatley, Keltner & Jenkins, 2006). According to Brosch et al. 2010, “the perception of an emotional stimulus is both stimulus driven and concept driven, i.e., the result is shaped by sensory information as well as by memory-based conceptual information and online evaluation capacities.” Perceptual and emotional processing are according to this view, highly intertwined.

Important for our study is the notion that affective associations always accompany the information in the place schema (Snodgrass et al., 1988; Ward et al., 1988). The power of these schemas is clearly illustrated by the experiment of Danford and Williams (1975). They compared the responses to 1. a guided walk through a law school complex; 2. slides and an audio tour of the guided walk with information of the function; 3. the same input but without information of the function; and 4. only the label “law school complex “(so no visual input). They found a high degree of agreement between the descriptive and affective ratings for all conditions, which indicates that participants relied on the common schema for this type of environment.

To sum up, we assume that for a specific task a selection of the complete knowledge of a place, and of similar places, is used to create a new mental representation. The mental representation is a transient amalgam of perceptual input, pre-existing imagery of the location or related places, and associated affective qualities. It will be used as a working concept in this thesis, although it has
aspects of a black box: we can neither describe its mechanisms or the content more precisely. Research in environmental psychology and virtual environments have not yet reached a consensual description of its structure, working, and content. Not only is it difficult to externally visualise such a complex structure using existing research methods; it is also highly individual by nature, relying on individual experiences, and may also show substantial individual differences.

3D models and virtual environments contain only a fraction of the information of real environments. Slater (Slater, Steed, & Chrysanthou, 2002) has remarked about virtual reality: “In interacting with the real world we are trying to make sense of too much information whereas in VR we are trying to make sense of too little.” Viewers complement the 3D model mentally. Babin and Burns (1998) label this as imagery elaboration, which they define as the activation of stored information in the production of mental images beyond what is provided by the stimulus. Although they refer primarily to non-visual stimuli, it is very likely that this occurs with visual stimuli as well. Even though in real life individual memories and preferences are also merged with the perceived environment, for 3D models the proportion of the added or superimposed information from memory is higher, and the resulting mental representation is thus more varied over a number of different viewers. Moreover, it may be modified by aspects of the representation such as screen size, low quality textures etc. On the other hand, this process of completing the mental representation allows the designer of a 3D model to focus on designing specific experiences. If (s)he is familiar with cues that are meaningful for viewers, these can be included and emphasized to trigger associations that are expected to elicit the required emotional reaction.

In Figure 2.4 a schematic view of this process is presented. Visual input and visual memories, coupled with associated feelings, are merged in a novel mental representation, and assessed.

![Figure 2.4 The role of the mental representation in the assessment of a virtual environment. The viewer perceives a 3D model of an environment, modeled after a real environment. The model triggers memories and associated appraisals of similar places and visits, and leads to an updated, novel, mental representation. The appraisal of the viewer is based on both the model and mental representations.](image-url)
2.6 Affective/emotional responses to virtual environments and the role of the representation

The validity of media to display environments for evaluation has been discussed extensively in environmental psychology. Many experiments using different response instruments have assessed the validity of different types of visualisations, for landscapes and urban areas, and for different aspects of evaluation, such as preference, aesthetic qualities and public safety of the represented area (for examples and overviews, see e.g. Bates-Brljac, 2007; Bishop & Rohrmann, 2003; Daniel, 2001; Daniel & Meitner, 2001; Palmer & Hoffman, 2001; Pietsch, 2000; Scott & Canter, 1997; Wergles & Muhar, 2009).

As discussed in Chapter 1, whether a representation of an environment is valid can only be assessed when the tasks for which it will be used, and the context in which the application is used, are defined accurately: the meaning and understanding of visuals may vary greatly depending on the knowledge domain, the roles of the users and the tasks that they perform (MacEachren et al., 2005).

In Chapter 2.4 the affective appraisal of real environments was discussed. In Chapter 2.5 and Figure 2.4 the concept of a mental representation was introduced. Viewers base their appraisals of an environment also on their personal repository of images and associations related to the environment of the viewer at a certain point in time. A virtual environment is an impoverished environment, lacking much of the multimodal information contained in a real environment. It is plausible that the existing mental representation plays a bigger role in the assessment of a 3D environment compared to the assessment of a real environment, although to what extent, cannot be determined in general; it is very much dependent on the individual and his relation with the represented location. In the experiments described in Chapter 4.3, 4.4 and 4.7 we find evidence for the importance of the mental representation.

In Figure 2.4, the representation or 3D model might seem a transparent medium through which the information of the real environment is presented. However, the representation is not transparent, and influences the appraisal of a viewer. Viewers are generally not aware of the representation as separate from the content, since we are used to interpreting, and working with, drawings, photographs, and television images from early age; also we tend to assess the presented environment as a whole, as explained above in Chapter 2.5. The influence of media characteristics or technology factors is not always immediately obvious. For instance, humans appear to have large tolerance for variance in image fidelity (Lee, 2004; Dillon, Keogh, & Freeman, 2002). This was shown in experiments by Reeves, Detenber and Steuer (1993) who found no difference between a low-fidelity version of a scene and a high-fidelity version of the scene in terms of arousal, memory, and attitudes. These results may be explained by the fact that humans are accustomed to observe the environment through peripheral vision, which is relatively unsharp. Nevertheless media characteristics such as display size influence the emotional response to the content, sometimes to a great extent. This topic and aspects of the medium such as lighting techniques, level of detail, display size, etc., are discussed in Chapter 3 and (Houtkamp, 2005).

In some cases the impact of the representation is conspicuous, and the medium elicits appraisals deviating from the content, i.e. the represented environment. Pleasant environments may be shown in photographs with ugly colours or unattractive lighting, or unpleasant scenes may be depicted in an aesthetically pleasing photograph (Figure 2.5).
Frome (2006) has proposed the “hybrid illusion theory” to explain emotional responses to events in media such as film, literature and games, which also may be used for representations of environments. Although we are aware the events in a film or a book are not real, we do experience emotions as a result of the events, sometimes very intensely. This theory is an adaptation of the classic illusionism which proposes that we emotionally respond to artworks because we are under the illusion that they are real, and which assumes that our minds form unified evaluations, either beliefs or judgments, about the situations portrayed in artworks. The hybrid illusion theory holds that only certain parts of our minds react to artworks as if they are real. On a global level we are aware that artworks are representations, but when a representation stimulates the mind in ways that are similar to how this happens in reality, those subsystems react as if they are engaged with reality (Frome, 2006). The theory offers an explanation why viewers may describe their experience of a representation as realistic, and confirms that viewers cannot easily separate appraisals of content (the represented world) and representation.

Viewers may at times be aware of a distinction in their appraisals of the environment and the representation, especially if the differences are extreme or if they are familiar with the displayed environment. If the environment is unfamiliar, the viewer would have to rely more on the images and associations in their mental representations. Without a reference (the actual situation), it is likely that viewers are not able to distinguish clearly between the affective qualities of the representation and the affective qualities of the environment. As a consequence, the assessment of the representation and content influence each other. In reality, viewers judge the molar environment, meaning the complete environment, with all its elements collectively. But even if the viewer is aware of the different appraisals of the environment and the representation, the resulting emotional responses will influence each other or be mixed into an amalgam and not be experienced separately. This effect of the influence of mood on evaluative judgment, and misattribution of mood, has been explained in Chapter 2.4.2.

Previous experience and expectations are important modifying factors in our response to a representation. From media history, it is known that whenever a medium provides a new sensory stimulation the audience at first is impressed with this new feature and experiences high perception of realism (Ijsselsteijn, 2003). After some time users become accustomed to the experience and their media schemata, or knowledge representations of what media are and what they are capable of, are updated.

However, even when viewers are used to certain stimuli, the perceptual system reacts to media as to reality, entailing local appraisal processes. Reeves and Nass (1996) suggest that this reaction has evolved during human evolution. In the social and physical environment of our ancestors, where danger was always present, “acceptance of what only seems to be real, even though at times inappropriate, is automatic”. In Frome’s terms, the subsystems respond to what is perceived, although at a higher level a viewer knows the situation is not real.

Humans are able to create rich mental models of environments from visual cues like drawings, or auditory or verbal information, and fill in gaps in information, often without being aware of it.
This is a valuable ability in many situations, but also has disadvantages in the context of communication and visualisation, for instance of urban planning. Drawings and 3D models are created to eliminate ambiguous elements in a plan, and to reveal information that is absent or incorrect. But representations are always abstractions of reality, and only show a selection of elements of the environment; therefore many gaps are still filled in by viewers individually. Moreover, by the selection and the style and technique of the representation, a biased picture may be presented that nevertheless looks convincing.

2.7 Summary

To understand the effects of virtual environments on viewers, we explained important concepts of theories on emotions, environmental psychology, and media theory.

In our experiments (Chapter 4) we study in the first place affective appraisals of virtual environments. These are the attributed emotional or affective qualities, or cognitions about possible object- or place-elicited emotions (Franz, 2005; Nasar, 2008). They are verbal labels of these qualities or characteristics of stimuli. The affective quality of an object (and environment) can be described in the same two dimensions as core affect, pleasure and arousal (Russell, 2003).

An affective/emotional reaction refers to an internal state (such as pleasure) that a person feels in relation to the environment (Nasar, 2008). Affective appraisals are not necessarily accompanied by an affective or emotional change (Russell & Snodgrass, 1987). Whether there is a reaction depends on the strength of the stimulus, the viewer’s affective state, and characteristics such as personality, and the relevance of the event or environment to the viewer. The affective response can be mapped on the circumplex model of affect, showing two dimensions pleasure and arousal (Russell, 2003). A conscious experience of one’s affective state can be described as an integral blend of the two dimensions and thus as a single point in this map (Figure 2.1).

When viewing a virtual environment, mental images are activated and the viewer appraises the new mental representation. In a real environment individual memories and preferences are also merged with the perceived environment, but in 3D models the proportion of the added or superimposed information from memory is necessarily greater.

The appraisal of the represented environment is also influenced by features of the representation: a medium is not transparent. Viewers are generally not aware of the representation as separate from the content, and do not assess them separately; this results in one judgment. The impact of the representation on the appraisal of the environment is a complex issue and underlies the main questions in this research.

Chapter 3 elaborates on a framework that identifies the important features of 3D environments, modifications caused by the modeling process and representation, and response moderators (of the viewer). We selected a number of these features, and designed experiments to further explore their effects on the appraisal or affective/emotional reaction of viewers. They are described in Chapter 4.
3 Affective appraisals of virtual environments: A framework for empirical research


3.1 Introduction

In this chapter we prepare materials for the experiments that constitute the main body of our research. In order to understand and examine the affective/emotional response to virtual environments, we need to

1. identify and classify environmental dimensions and features that determine affective/emotional responses in real environments,
2. identify modifiers and determinants of the representation, viewer, and context, that also influence affective/emotional responses when viewing virtual environments, and
3. select appropriate measuring instruments for assessing the affective/emotional responses to virtual environments.

First we present an operational framework for our experimental research, including factors of the environment, the representation, the viewer, and the context, which are important in the process of appraisal of a virtual environment (Chapter 3.2). Environmental psychologists have developed classifications to capture the most important elements or stimuli in the environment, often visual features, to assess their influence on the human observer. These are discussed in Chapter 3.3.

Modifiers and determinants belonging to the representation, viewer, and context, which influence affective/emotional responses when viewing virtual environments, are explained in Chapter 3.4.

The last paragraph (3.5) discusses measuring instruments for the two types of responses of the viewer that are measured in experimental research in this domain: affective appraisals and affective/emotional reactions. The framework and measures together provide the foundation for the experiments that are presented in Chapter 4.

3.2 Building a framework for empirical research

The original Stimulus-Organism-Response model used by Mehrabian and Russell (1974) (Figure 2.2), has been adapted and extended by several researchers (e.g. Kottasz, 2006). An often cited adaptation is Bitner’s framework for understanding environment-user relationships in service organisations. Bitner has added the “holistic environment” to the model to indicate that the environmental dimensions are perceived together, in what is sometimes called the “molar environment”. Moderators influencing the response are put between the environment and internal responses. Bitner also added cognitive and physiological responses to the emotional responses in the original model. The model was adapted specifically for servicescapes by distinguishing employees and customers.

We used the main components to develop a framework for studying the viewer’s response to virtual environments. Figure 3.1 shows the simplified framework, Figure 3.2 the version including the relevant variables. In the new framework, the environmental features (1) are modified by features of the representation (2), but perceived together as a holistic virtual environment (3). This means that viewers do not always distinguish between features of the representation and the represented environment. Response moderators (4) and internal responses (5) include factors that are relevant
to viewing 3D environments; some are already discussed in Chapter 2.4.2. In the rest of this chapter, the other elements in this framework are explained.

**Figure 3.1** A framework for understanding the affective appraisal of and response to virtual environments. Elements 1, 2 and 3 together constitute the stimulus as in the Stimulus-Organism-Response model, Figure 2.2.
**Figure 3.2** Framework of determinants of the affective/emotional response to a virtual environment.

The items in italics are addressed in the experiments in Chapter 4.

<table>
<thead>
<tr>
<th>1 Environmental dimensions and features (characteristic of real environments)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambient conditions</td>
</tr>
<tr>
<td>• temperature</td>
</tr>
<tr>
<td>• weather (2)(5)</td>
</tr>
<tr>
<td>• noise/sounds (5)(6)(7)</td>
</tr>
<tr>
<td>• music</td>
</tr>
<tr>
<td>• olour</td>
</tr>
<tr>
<td>• time of day (dark/light) (9)</td>
</tr>
<tr>
<td>• etc.</td>
</tr>
<tr>
<td>Spatial layout / functionality</td>
</tr>
<tr>
<td>• buildings</td>
</tr>
<tr>
<td>• other elements e.g. (street) furniture, canals, trees</td>
</tr>
<tr>
<td>• reads</td>
</tr>
<tr>
<td>• etc.</td>
</tr>
<tr>
<td>Signs, Symbols and Artifacts</td>
</tr>
<tr>
<td>• signage</td>
</tr>
<tr>
<td>• materials</td>
</tr>
<tr>
<td>• decor (9)</td>
</tr>
<tr>
<td>• maintenance (2)</td>
</tr>
<tr>
<td>• use (people)</td>
</tr>
<tr>
<td>• etc.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 Representational modifiers and determinants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection / modification of information</td>
</tr>
<tr>
<td>Restrictions of the medium</td>
</tr>
<tr>
<td>• loss of multimodal information (sounds: (5)(6)(7))</td>
</tr>
<tr>
<td>• absence of body (6)</td>
</tr>
<tr>
<td>• absence of other persons</td>
</tr>
<tr>
<td>Requirements related to goal</td>
</tr>
<tr>
<td>• visual realism</td>
</tr>
<tr>
<td>• functionality</td>
</tr>
<tr>
<td>Choices in modeling process</td>
</tr>
<tr>
<td>• style (3)</td>
</tr>
<tr>
<td>• level of detail (2)(4)</td>
</tr>
<tr>
<td>• illumination (3)</td>
</tr>
<tr>
<td>• time of day (2)(9)</td>
</tr>
<tr>
<td>• weather, season (2)(5)</td>
</tr>
<tr>
<td>• perspective (6)</td>
</tr>
<tr>
<td>• dynamic elements (5)(6)</td>
</tr>
<tr>
<td>• freedom of movement, control over velocity</td>
</tr>
<tr>
<td>• narrative</td>
</tr>
<tr>
<td>Modifications by platform</td>
</tr>
<tr>
<td>• field of view (6) (8)</td>
</tr>
<tr>
<td>• depth perception (6)</td>
</tr>
<tr>
<td>• resolution</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3 Perceived holistic environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived virtual environment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 Response modertors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual characteristics and user state</td>
</tr>
<tr>
<td>• Goals</td>
</tr>
<tr>
<td>• Pre-existing mental representations and associations (3)(4)(7)</td>
</tr>
<tr>
<td>• Expectations (2)</td>
</tr>
<tr>
<td>• Mood and emotional state (9)</td>
</tr>
<tr>
<td>• Cybersickness (8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5 Internal response to environment and representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
</tr>
<tr>
<td>• beliefs</td>
</tr>
<tr>
<td>• categorization</td>
</tr>
<tr>
<td>Physiological response (5)</td>
</tr>
<tr>
<td>• Affective appraisal (2)(3)(4)(5)(8)(9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6 Behavioural response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approach or avoidance behaviours</td>
</tr>
<tr>
<td>Social interactions</td>
</tr>
<tr>
<td>Performance (for instance in training) (7)</td>
</tr>
<tr>
<td>Affective/emotional reaction (5)(6)(7)(9)</td>
</tr>
<tr>
<td>Engagement (5)(7)</td>
</tr>
<tr>
<td>Perceived realism of presentation (convincingness) (6)(7)</td>
</tr>
</tbody>
</table>

**Experimental studies:**

1. decay in an urban environment - affective appraisal environment
2. illumination in an indoor environment - affective appraisal environment
3. visual dynamics and sounds in outdoor environment - affective appraisal environment
4. depth and speed cues in training environment - affective appraisal environment
5. sound in fire fighter’s training environment - affective appraisal environment
6. effect of cybersickness - affective appraisal environment
7. stress and ambient darkness - affective appraisal environment, arousal
8. effect of cybersickness - affective appraisal environment, arousal
9. illumination in an indoor environment - affective appraisal environment, arousal
3.3 Environmental dimensions and features as determinants of affective/emotional responses in real environments

**Compare framework: 1. Environmental dimensions and features**

An environment contains numerous physical and atmospheric features that together determine the affective/emotional response of a viewer. Several classifications have been made to capture the most important elements or stimuli in the environment, often visual features, to allow assessment of their influence on the human observer. Some are very specific to a research goal, for instance the studies on the effects of healthcare facilities on the health and well-being of patients (Dijkstra et al., 2006). Others are related to specific explanatory theories (for instance Ulrich, 1983; Korpela, Klemettilä, & Hietanen, 2002). We discuss only the most important approaches below, and attempt to identify elements that are relevant for most natural and man-made, outdoor and indoor environments and for a variety of research goals.

Several categories are suggested to discern and cluster environmental elements or stimuli on a physical level. Nasar (1998) distinguishes between two types of visual attributes of environments that are relevant for the aesthetic quality of environments. He identifies (1) formal attributes that refer to physical properties and relations, such as shape, proportion, colour, scale; and (2) content variables that relate to physical features in environments, but more specifically reflect the individual’s internal representation of the environment and meanings associated with that representation and environment. Physical features in environments individually or together, contain clues for the interpretation of the content of forms, of a style, of a type of building, or its function. These clues may trigger existing mental representations of a place. Mental representations of places play an important role in the affective appraisal of environments and are discussed in Chapter 2.5.

Bitner (1992) also recognizes formal and content attributes in physical environments, but adds ambient conditions as a third category of environmental characteristics for servicescapes. She distinguishes the following categories:

- **Spatial layout and functionality** refers to the ways in which elements in the environment (in case of servicescapes for instance equipment and furnishings) are arranged, the size and shape of those items, and the spatial relationships among them. Functionality refers to the ability of the same items to facilitate performance and the accomplishment of goals.

- **Signs, symbols and artefacts** consist of actual signs and labels, and of objects that give implicit cues to users about the meaning of the place and norms and expectations for behaviour in the place, for instance quality of materials used in construction, or artwork.

- **Ambient conditions** include background characteristics of the environment such as temperature, lighting, noise, music, and scent. Ambient conditions are recognized as a factor that affects perceptions of and human responses to the environment (Russell & Snodgrass, 1987; Bell et al., 2001).

Similar categories were used by other researchers, for instance Harris, McBride, Ross and Curtis (2002), for studying hospital environments and Kottasz (2006) for museums.

A different approach aggregates physical features of environments into more abstract attributes of the environment that are relevant to human affect and behavior. Whereas they refer to concrete physical attributes of the environment, they depend more on the observer’s integration of the environment’s attributes (for instance Nasar, 2008; Hanyu, 1997; Hanyu, 2000; Tveit, Ode, & Fry, 2006; Ode, Tveit & Fry, 2008). As an example, Nasar (2008) distinguishes as salient physical attributes that stand out in people’s perceptions and affective appraisals: naturalness, upkeep, openness, complexity, order, and historic significance. Tveit et al. (2006) also includes concepts such as disturbance and ephemera, indicating seasonal change and weather related changes. Table 3.1 provides an overview and explanation of the salient attributes distinguished by Nasar (2008).
<table>
<thead>
<tr>
<th>Salient physical attributes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naturalness</td>
<td>refers either to the individual’s perception of an area as natural or to the predominance of natural elements</td>
</tr>
<tr>
<td>Upkeep (civilities)</td>
<td>refers to the perceived maintenance (and lack of signs of decay) of areas. The negative pole of this attribute is sometimes referred to as physical incivilities, which function as cues to social disorder</td>
</tr>
<tr>
<td>Openness</td>
<td>refers to the perceived vista, visual scope, and related attributes (such as spaciousness, building density, and defined space). (...)</td>
</tr>
<tr>
<td>Complexity</td>
<td>refers to the amount of structural information in a scene, the number of different noticeable elements, and the distinctiveness among those elements</td>
</tr>
<tr>
<td>Order</td>
<td>refers to the degree to which people see an environment as unified, coherent, congruous, legible, or clear</td>
</tr>
<tr>
<td>Historic significance</td>
<td>rests on the observer’s perception. An environment could either have authentic historic significance or simply look historic to the observer</td>
</tr>
</tbody>
</table>

Table 3.1 Salient physical attributes of environments Nasar (2008).

The relationship between physical attributes, abstract attributes and affective appraisal has been subject of many studies; however they generally study the effect of singular or collative attributes (structural properties of visual patterns) on appraisal. In an overview Hanyu (1997, 2000) mentions examples, for instance the studies on the relationships of collative attributes such as complexity, coherence, novelty, with surprisingness and preference and interest; of man-made nuisance objects, physical incivilities with fear; and of lighting condition with preference (Flynn, Spencer, Martyniuk, & Hendrick, 1973; Hendrick, Martyniuk, Spencer, & Flynn, 1977; Flynn, 1988). Only a small number of published studies have examined the variables (physical attributes, abstract attributes and affective appraisal) together, such as Hanyu (1997) for scenes at night and Hanyu (2000) for scenes in daylight. They show relationships between environmental characteristics and appraisals, of which most important are the positive correlation between pleasure and the combination of naturalness/openness of environments, and between arousal and disorder in environments. Hanyu (2000) concludes that the findings have not yet contributed to a comprehensive theory of environmental assessment. The results of the individual studies are so much dependent on the specific environment and the chosen variables that no general conclusions on the effects of any of the variables on the affective appraisal of environments can be presented here. Known effects of specific features of environments on the appraisal are included in the paragraphs on the experiments in Chapter 4 when relevant to the experiment concerned.

In the preparation of our experiments, we decided on a practical categorisation of physical environmental features that directly accommodates the variables we chose to manipulate. We therefore chose to apply Bitner’s (1992) three categories, 1. spatial layout and functionality, 2. signs, symbols and artefacts, and 3. ambient conditions. They cover all physical elements in a real environment, as well as the elements that are used to build a 3D model of an environment. Since the models we studied do not include representations of humans, categories describing human behaviour or appearance such as mentioned by Kottas (2006) are not required.

The other categorisations mentioned here (Nasar, 2008, and Tveit et al., 2006), are aggregations of physical features of environments, and represent more abstract properties of environments. They include perceptions and evaluations of environments and are therefore more closely related to the appraisals of viewers of environments, which are the dependent variables in our experiments. The features recognized by Nasar and Tveit can be found again as factors in the appraisals, so in the measuring instruments used in the experiments (Chapter 3.5). For instance, complexity, openness and upkeep are similar to factors in the appraisal of environments in the SMB-scale by R. Külter (1977).
3.4 Determinants of appraisal in virtual environments

*Compare in framework: 2. representational modifiers and determinants*

Environments generally contain dynamic, very detailed and abundant multimodal information. They allow natural interaction with objects in the environment. The bulk of this information and interaction is simplified or omitted in the process of modeling virtual environments. Furthermore, the representation may also add cues, which influence the user’s response. Aesthetically pleasing styles, or unexpected elements such as sounds, can also elicit an affective/emotional response of the viewer. In the next paragraphs, we give an overview of the factors that may influence the affective/emotional response to a virtual environment. We distinguish three categories of factors: those related to the modeling, to the display of the environment, and to the user (Table 3.2).

3.4.1 Modeling: selection and modification of information.

In the process of modeling 3D environments developers face restrictions imposed by the medium; they select information determined by the goal of the application, and apply styles and techniques they are familiar with and find aesthetically pleasing.

*Restrictions of the medium*

*The loss of multimodal information*

Current technology cannot provide realistic information to all senses in real time, and advances have been made mainly in visual and auditory display technology. Consequently, attention of researchers has focused on the two sensory modalities of vision and audition. Very little research is performed to assess the effects of the loss of other modalities on the appraisal of an environment such as tactile information and smell. In the last decades, static visual representations (photographs, slides or sketches) were used widely for environmental assessment. As a result, the navigation and freedom of movement provided by virtual environments are seen as an improvement, while the absence of other modalities is accepted as standard.

Although the quality of graphics software and display technology has increased considerably over the last years, the visual information provided by a desktop virtual environment is only a fraction compared to that of a real environment; some of the consequences are discussed below as effects of the display medium.

*Sound*

In sound design for virtual environments and gaming, a distinction is generally made between ambient sounds and sound events. Other types of sounds mentioned are those that augment or substitute perceptions made by the other senses, communication sounds, and sounds that are outside the virtual world (non-diegetic sounds), such as voice-overs or interface feedback.

Ambient or background sounds are used to set the mood or ambience of an environment or experience. These sounds provide a sense of location and make the virtual world believable to the user. The absence of sounds may draw attention to the mediated nature of the environment which diminishes the convincingness of the environment (Fencott, 1999). Auditory background and sound events are essential to establish a feeling of contact with and presence in an environment (Dinh, Walker, Hodges, Song, & Kobayashi, 1999; Larsson & Västfjäll, 2001; Sanders & Scorgie, 2002; Whitelock, Romano, Jelfs, & Brna, 2000; Serafin & Serafin, 2004). Sound makes the environment come alive, and envelopes the user, thereby enhancing her involvement (Blesser & Salter, 2007).

Sound events provide information on events, actions of other people in the environment, and feedback to actions of the user in the virtual world which is also important for orientation in the environment (Shilling & Shinn-Cunningham, 2002). They may indicate events outside the field of view of the user. Together they enhance situational awareness (Shilling & Zyda, 2002).
Little is known of the general influence of the presence or absence of sound on the affective appraisal of environments, but by providing information about spatial dimensions and materials used, sounds evidently play a role in assessments of for instance the perception of enclosedness and complexity (Küller, 1977). Küller also mentions that people associate long echoes with types of buildings like churches and indoor swimming pools. These associations may influence the perception of the affective qualities of a place.

Sound, or more specifically noise, can influence mood and the interpretation of events. The addition of sound of good quality enhances the user’s experience and may also influence the user’s opinion of the environment in general (Rohrmann & Bishop, 2002).

Sound displays for desktop systems usually involve small speakers. This complicates the presentation of sound because of possible noise pollution and the acoustics of the room in which the system is located. Headphones are more suitable to convey 3D spatialized sounds, which provide the user with information on the direction and distance of the origin of the sound (Sherman & Craig, 2003).

Phenomena called cross-modal enhancement and cross-modal illusion may explain why viewers of a virtual environment often do not miss sound or other modalities, and why designers do not worry about them. Experiments by Storms and Zyda (2000) have shown the occurrence of auditory-visual cross-modal perceptions, indicating that auditory displays influenced the quality perception of visual displays, and vice versa, whereas the subjects were not aware of those intersensory effects. Biocca, Kim and Choi (2001) conclude from experiments with (immersive) virtual environments that the mind attempts to integrate incomplete sensory cues to form a complete spatial mental model of the virtual environment (Ernst & Bülthoff, 2004). These phenomena may explain why we are able to give an affective appraisal of an environment even if we only have very limited information: our mind fills in the gaps to create a more complete mental representation, adding illusions of other modalities, and we evaluate the affective qualities of the mental representation.

Other sensory channels

Virtual environments on desktop systems generally do not include other sensory displays such as haptic (touch), vestibular (balance), olfactory (smell) or taste displays. Again, very little is known how much these senses contribute to our affective appraisal of the environment. To mention just a few examples of the importance of these channels: in reality we are aware of the texture and weight of materials we touch, we perceive the temperature and humidity of the air, we notice wind or draught, we are aware of smells that may alarm us (fire) or reminds one of a pleasant place. Bell et al. (2001) discuss comfort and discomfort as a result of high or low temperatures and wind, and the effect on for instance affective feelings and social behaviour.

Olfactory stimuli and their effects have received attention in marketing research, often to assess the influence on consumer behaviour (Turley & Milliman, 2000). Smell is very powerful in evoking memories that are associated with past experiences and the emotions that arise from them (Ward, Davies, & Kooijman, 2007). Recent efforts to develop olfactory-enhanced applications in multimedia are discussed by Ghinea and Ademoye (2010). Olfaction is used towards a variety of goals, including notification alerts, enhancing the sense of reality in immersive applications, and branding, but also multimedia application to inform and/or entertain users.

Further research is required to link haptic and olfactory information to dimensions of appraisal and emotional response. Familiar beliefs can be mentioned here, such as that the feeling of enclosedness will probably be greater if the temperature is high, and the pleasantness of an environment will be less if one feels a strong cold draught, or smells a bad odour. The focus of this study is on visual and auditory information, so this topic is not pursued here.

Absence of body

In many virtual environments the user has a so called first-person perspective (or point of view, POV) on the environment, as if from eye level of a viewer present in the environment. The advantages of the first-person perspective are supposedly a greater involvement with the
environment, and in games, in the gameplay (Clarke-Willson, 1997; Rouse, 1999). A disadvantage is that in a first-person perspective on a desktop VE the body cannot be used as a reference for scale and position. This makes it more difficult for a user to estimate dimensions and locate his or her position in a complex building or environment. Moreover, although the relative motion between the user and the virtual world is correct, the percept of self-motion can be lost. To compensate for this loss of information a third-person perspective (or POV) is often used in computer games and simulations. In this mode, the position of the camera is related to the avatar (digital representation of a person) so that the person is visible from some distance behind (Salamin, Thalmann, & Vexo, 2006). The implementation of a third-person perspective is more complex than of a first-person perspective, because there may be positions of the avatar in which the software has difficulties producing a full view: imagine the avatar leaning with her back to a wall (Rouse, 1999). To reduce the problem of occlusion the viewing angle of the camera is often increased (Bateman, Doucette, Xiao, Gutwin, Mandryk, & Cockburn, 2011). The advantages are that the extent of the area visible for the viewer is larger, that the avatar’s body provides a reference for estimating dimensions of the environment, makes it easier to locate the position of the avatar in the environment, and to see how the avatar moves in relation to other objects. The disadvantage, however, is that the POV of the user no longer corresponds to the POV of the camera. That is, the external field of view (the angular size of the observer’s physical visual field that is filled by the displayed virtual scene) is larger than the internal field of view (the field of view the graphics generator is using to render its images of the environment). A mismatch between internal and external field of views will cause image distortions as well as distortions in optic flow which may yield incorrect estimates of spatial properties and cybersickness (Emmerik, Vries & Bos, 2011). Although it is often assumed that this perspective is preferable for steering and driving tasks, for instance in driving games, Bateman et al. did not find positive effects of third person perspective on performance or preference (Bateman et al., 2011). How the perspective affects the appraisal of environments has as far as we know not been subject of study.

Absence of other persons

The absence of other people in many 3D models has without doubt an important effect on the affective appraisal of the environment. The representation of characters (or avatars) becomes increasingly convincing in computer games, but in serious games and visualisations they are still simple and often static elements. Merely by their presence in an environment, avatars influence environmental variables like complexity, unity and enclosedness. From marketing research, it is known that perceived crowding has a negative effect on evaluations of shopping environments (Turley & Milliman, 2000). Avatars also serve as visual references for the perception of spatial dimensions. The sounds humans produce provide information about dimensions of spaces, and materials used. On a behavioural level, humans provide clues for the function of an environment, the way to behave (for instance move reverendly and whisper), and may influence dimensions of appraisal such as social status, potency and pleasantness. Possibly the presence of other persons influences the affective appraisal of an environment through the feeling of co-presence.

Requirements related to the goal of the 3D model

Visual realism

The goal of developers is often to attain a high level of visual realism. Is photorealism required if the goal of a 3D model is primarily to evoke similar affective appraisals as the real environment? Or does photorealistic rendering, including many details and complex illuminations create distractions and cause an annoying delay in viewing the model? Several authors (Mahdjoubi & Wiltshire, 2001; Slater et al., 2002; Davide & Walker, 2003) point out that the illusion of realism for the observer of a representation of a virtual environment is not necessarily the result of a very close visual (for instance geometrical) similarity to the real world and of an increased level of detail. The illusion is a result of the representation of relevant properties of reality, essential elements that have been
selected dependent on the purpose of the representation. This “simplistic” approach may help to explain the realistic effect of 3D models that consist of only very basic shapes and few details: a few simple lines, planes and shadows may trigger a pre-existing mental image of a (three dimensional) environment, building, a room or corridor. As a consequence it advocates that the designer selects cues that capture the “essence” of buildings and spaces that are able to trigger required internal representations.

The illusion of realism does not necessarily mean that our appraisal of the 3D model is identical to the appraisal of the environment in reality; it might evolve when an observer feels that the representation contains a sufficient quantity of information to meet his or her demands or expectations. The information in 3D models is generally more abundant than in drawings, photographs or other static representations. By moving around viewers can obtain more effective depth cues and in general more information on spatial dimensions, illumination, position of elements etc. Hence, a high degree of photorealism in 3D models on desktop computers is not always required, but depends on the specific task (Scaife & Rogers, 1996). An affective appraisal of a building requires a different set of information than cognitive tasks such as estimating the size of a building. An accurate and detailed visualisation does not necessarily convey the same “ambience” as the real environment. Vice versa, features that influence the perceived ambience of a visualised environment do not automatically improve the validity of a visualisation (see also Chapter 1.1.1).

Functionality

The functionality offered in a 3D environment is often limited to moving around and exploring a building or an area. The experience is different from reality due to the loss of haptic feedback: manipulating objects produces information on the materials used. In computer games the player can usually manipulate all kinds of objects that are part of the environment. This adds to the involvement of a player and creates a sense of reality, but in which way it has an impact on the affective appraisal of the environment, is, again, difficult to say.

Choices made in the modeling process

Even within the restrictions of the software, hardware and the goal of the model, developers have numerous choices in designing a 3D model of an environment. As Sherman and Craig (2003) put it, representation is the choice of what to render. The process by which visualisations are designed, or built is not at all straightforward (MacFarlane et al., 2005). Visualisation developers need to make decisions about for instance level of detail, lighting, atmospheric circumstances, point of view, and dynamic elements, which can influence the affective qualities of the scenes, and give them different meanings to different individuals and groups (MacFarlane et al., 2005). Decisions in the development process on abstraction, accuracy, and realism are made on vague consensus of “sufficient realism” (Appleton & Lovett, 2003). The style consciously or unconsciously adopted by developers influences the ambience of an environment. For instance, the conventions of the artist’s impression have become standard in 3D visualisations in urban design, and obviously portray the environment too favourably (Rekittke and Paar, 2009).

Illumination, time-of-day, weather, season

The simulation of lighting is “the major and central conceptual and practical problem of computer graphics” (Slater et al., 2002, p.73). The complexity of illumination and shadows is such, that real time rendering demands a great deal of computing power. First of all, the location, type and intensity of the light sources, or the absence of light, influence the colours of a space and the perception of spatial dimensions. The impression of enclosure increases as lightness in a room decreases, but more diffused lighting makes the degree of enclosure lower than if the light is concentrated to one point in the room; by the use of light or dark colours on walls the impression of the shape of the room can be altered as well, and the use of many strong lights increases the complexity of a room (Küller, 1977). The time of day and season and the weather, as noticeable in a clear blue or threatening sky, not only influence perception of the colours of an outdoor environment, but may also affect the atmosphere of the environment, comparable to effects used in
films. Cinematographic techniques of lighting are especially effective for establishing a desired atmosphere, and for attracting attention to relevant areas and creating depth in an image (Giannetti, 2002).

**Dynamic elements**

The static, motionless impression of many 3D environments, where the only movement is the consequence of navigation by the user, is another obvious characteristic imposed by restrictions of hardware, software, but mainly by the required effort in modeling.

Visual dynamic attributes of a setting, such as moving traffic, clouds and water movement, are environmental elements which should be included in the representation (Hetherington, Daniel & Brown, 1993; Huang, 2004; Huang, 2009). The increase in the use of game engines to build virtual environments, offering easy implementation of so-called atmospheric phenomena, dynamic elements and sounds, raises questions about the effect of these visual and auditory elements on the perception and appraisal of the user. Representing subtle dynamic elements in an environment on a desktop monitor may be distracting on a monitor, and may give unwanted visual effects when the resolution of the display is not high enough.

**Freedom of movement and control over velocity**

Freedom of movement and control over velocity are often mentioned as the main advantages of 3D models compared to static representations, since navigating a model helps the user to gain an overview and understanding of the spatial layout of an area, provides an impression of distances, dimensions, and allows inspection of sightlines. Navigating requires viewers to focus their attention on the layout of a virtual environment. Conniff, Craig, Laing and Galán-Díaz (2010) found that self-navigators of virtual models noticed structural and layout alterations between two versions of a virtual environment, whereas (passive) observers of walkthroughs noticed more architectural details. Self-navigators also reported a greater sense of presence in the environment.

A factor that is difficult to assess without further research is the effect of speed of moving in the virtual environment. Judgment of complexity is different when we move at walking speed and can detect small-scale details, or when we “fly” or “jump” at high speed to another location in a 3D model and only notice large-scale elements (Nasar, 1994). When the user leaves the “walking mode” of moving through the model to fly or jump, she also loses the point of reference of viewing the environment at ca. 1.70m height and walking speed, and consequently experiences dimensions differently. Velocity and viewpoint can be used by developers to manipulate the affective appraisal of the environment. If the model contains few details, the speed of movement might be increased to keep the information rate at a level that still holds the attention of the viewer and in this way maintain the appraisal of the dimension complexity. The simulation of moving very fast over or through an environment may also add to the aesthetic pleasures of the model (Fencott, 2001), and may produce a sensation that makes the viewer of the model forget about elements in the model that are of lower quality. An example can be found in games, where the speed of navigation is slowed down when the player explores underground chambers to create suspense and tension.

**Narrative**

When we visit an environment in reality we create our own narrative, a sequence of views, actions, experiences and affective appraisals of the locations visited. The similarities and contrasts between locations affect our perception of the different locations, and of the place as a whole, and can create for instance surprise (Heft & Nasar, 2000). If a user of a 3D model is not restricted to certain locations but can place him- or herself in any position from which to observe the environment, and can fly or jump quickly to other positions, this may influence the affective appraisal of an environment.

** Modifications caused by the platform**

The display of a 3D model on a monitor and the means of moving through the VE using a keyboard and mouse, or joystick, have an important effect on our perception of the represented
environment. As far as visual information is concerned, determinants are the generally small size of a desktop monitor, the distance between viewer and the monitor, the restricted (internal or geometric) field of view, the issue of depth perception on a flat screen, perception of spatial dimensions, the resolution of the images, and the colours that are influenced by the characteristics of the display. Visual perception of images on a desktop monitor and means of moving using a keyboard and mouse or joystick obviously have an effect on orientation and way finding. These topics have received attention in research over the last years because of reported problems of the feeling of getting lost in virtual environments (Darken & Peterson, 2002), which negatively influences the affective appraisal of an environment (Evans & Gärling, 1991).

**Field of view**

When navigating a 3D model, the viewing distance to the display is important for the experience of the environment. The design eyepoint is the point – or volume in space – where an observer has an undistorted view on the visual stimuli. When an observer moves away from this design eyepoint, for instance by leaning forward or backward, this will cause image distortion (Van Emmerik et al., 2011). Van Emmerik et al. distinguish two kinds of field-of-view (FOV): the external and internal FOV (Figure 3.3). The external FOV is defined as the angular size of the observer’s physical visual field that is filled by the displayed virtual scene. It is directly related to the size of the display used and the viewing distance of the observer. They consider the internal FOV the design eyepoint. This is independent of the viewer’s position and screen dimensions because it refers to the FOV the graphics generator uses to render its images of the environment.

![Figure 3.3 Schematic representation of the external and internal field of view. The external field of view is determined by the size of the screen and the distance of the observer to it. The internal field of view is specified in the graphics generator. The different angles formed by the solid and dashed lines show examples (Emmerik et al., 2011).](image)

A position that deviates from the design eyepoint (for instance because the viewer is farther away from the display) will distort perceived dimensions and shapes in the virtual environment.

The extent of our visual field without movement of eyes or head is approximately 200 degrees horizontal and 135 degrees vertical (Werner & Rossi, 1991). Movement of the eyes and the head add substantially to the horizontal and vertical extent of the area humans can observe. The view a desktop display offers is much more restricted. The guidelines for the use of a desktop VDS state that the viewer should be about 50cm from the display (American National Standards Institute 1988). A 17" display viewed at this distance has a FOV of 37 x28° (Pfautz, 2002). Although the quality of our perception diminishes near the edges of our visual field, at any moment we view a large area of the real environment outside the area of the virtual environment on the monitor, and the periphery may interfere with our perception of the virtual environment. A position farther away from the display diminishes the experience of being surrounded by buildings, a forest, etc.; the viewer’s peripheral view is not covered by the virtual environment but by the real location.

A decrease in the external field of view may have negative effects on navigating virtual environments; Jansen, Leverts, Toet and Werkhoven (2011) have demonstrated that a decrease in the vertical angle impairs maneuvering through a virtual environment.
To explore an environment on a desktop monitor requires actions with a mouse or keyboard and an effort to complete a mental model of the environment. This characteristic of a desktop system and the chosen geometric field of view generated in the display thus seem to have most effect on our perception of spatial dimensions and on the awareness of the viewer of the position in the room or environment. Neale (1996) concluded from experiments that 60 degrees for the horizontal field is the optimum geometric (or internal) field of view for perspective displays to perceive the basic characteristics of space accurately. The designer of a model can select a wider geometric field of view to allow more of the environment to be displayed and to improve spatial orientation, but this may lead to distortions in perceived distances, angles and shapes (Waller, 1999; Neale, 1997). Another possibility to improve spatial orientation is the use of symbolic enhancements in the environment such as visual momentum techniques (for instance landmarks) used in film to provide continuity between scenes. Or, a solution may be found in using a third-person perspective that produces the effect of a wider field of view on the virtual environment, but again leads to distortions in the perceived environment (Pazuchanics, 2006). Distortions of the model or variations in the field of view can be used on purpose, to compensate for the effect of a chosen geometric field of view, or to strengthen or diminish a certain feeling such as enclosedness.

**Depth perception**

Although we generally use the term 3D models we actually view computer models on a monitor as perspective representations on a 2D viewing plane (Whyte, 2002). The perception of depth is one of the most important factors for estimating spatial dimensions, which in turn determines much of our assessment and appraisal of a place. The degree of enclosedness, or impressiveness of an environment is determined by spatial dimensions.

Display on a monitor can have unwanted effects on complexity, for instance, when a large room contains many different elements such as furniture and decorations, but they all seem to clutter on a single plane as a consequence of a diminished perception of depth.

Depth cues in reality generally provide redundant and sometimes conflicting information, and it is important to be aware of the relative weight when designing a 3D model.

Sherman and Craig (2003) describe four varieties of depth cues people use in a real environment, two of which are applicable to the most common desktop monitors, namely monoscopic depth cues and motion depth cues. Monoscopic depth cues include interposition (the partial blocking of a more distant object by a nearer object), which is the most important; shading, size, linear perspective, surface texture gradient, height in the visual field, atmospheric effects and brightness. Motion depth cues come from the parallax created by the changing relative position between the eyes and the object being observed and have a very strong effect.

Stereoscopic depth cues are based on the fact that our eyes have a slightly different view on the world due to their different position in the head. The brain correlates the difference in the position of objects on the retinas as depth information. Stereo depth cues can be simulated by modern displays using several techniques. The basic requirement is to display offset images that are filtered separately to the left and right eye. For a more complete description of depth cues we refer to the literature on this subject (e.g. Sherman and Craig 2003; May and Badcock, 2003).

Evidence from experiments by Waller (1999) and Witmer and Kline (1998) is not conclusive on the question whether distances in virtual environments can be accurately estimated. Important factors that emerge from their studies are feedback, which help participants to correct their estimations in a following session, a sufficiently wide geometric field of view, varying between 50 and 80 degrees for the horizontal field, and the ability to move in the environment. By moving around observers can view objects from different angles and relative positions, and estimate distances traversed by the time that passes. A final factor we mention is the availability of some sort of reference for scale, necessary since an observer of a 3D model in first-person point of view cannot use his or her body for comparison. Familiar references for scale in daily life are other people, doors, windows, furniture, trees etc. If in a virtual environment the dimensions of these elements, for
instance doors, are unusual, they may confuse the observer (Slangen-de Kort, IJsselsteijn, Kooijman, & Schuurmans, 2001).

Resolution, other characteristics of the display and their effects on detail and colour

Technical characteristics of the hardware used as well as environmental factors influence the display of a 3D model on a monitor. Resolution of the display used, viewing distance, colour settings and ambient light influence the perception of detail and the colours of the virtual environment. Resolution (pixel density) affects the image quality, and thus the level of detail of the represented environment, and of the perceived quality of the representation (Bracken, 2005). Colour and colour contrast are important factors in the affective appraisal of the environment and are relevant for many dimensions of appraisal. There is some recent empirical research directed specifically at the effectiveness of manipulations of environmental colour, and at the effects of colour in virtual environments (Franz, 2005; Franz, 2006). Generally, warm and saturated colors are seen as more arousing, while the dominance dimension corresponds to light-dark, both within and between color hues. Effects of colours on the dimension valence are less clear. Positive relations between brightness and saturation, and valence were found by Mehrabian and Russell (1974). For colour hues, research shows more mixed results, complicated by theories of colour harmonies. The effects of colours are furthermore complicated because of an assumed dependency on the colour-bearing object category and on the culturally dependent symbolic meaning (Franz, 2005). Bell et al. (2001) mention results from studies in environmental psychology indicating that perception of spaciousness, or conversely, crowding, may be influenced by colour. Lighter rooms were seen as more open and spacious.

The perceived holistic environment

In the framework, the third column is the perceived holistic environment. It is assigned a separate column to indicate that it represents a virtual environment with a unique combination of environmental and representational features, viewed by a specific viewer, with specific goals and tasks, at a certain point in time.

3.4.2 Response moderators

Compare in framework: 4. Response moderators

The effects of the factors mentioned above are influenced by the emotional state and personal characteristics of the observer. Mood and affective states may modify the appraisal of a place (Russell & Snodgrass, 1987). Expectations of a place and pre-existing mental representations influence the creation of a mental image, which according to our theory plays an important role in the affective appraisal. Experience with the medium, including the techniques for navigation and expectations of graphic conventions and styles may also be an important factor (Sherman & Craig, 2003). In Chapter 2, we already referred to the “media schemata”, people’s previous experiences with and expectations towards media that determine whether a person accepts a mediated stimulus at face value and responds accordingly (Ijsselsteijn, 2003).

Several factors of the context or environment in which the 3D model is displayed, may influence our perception and appraisal of the virtual environment itself, such as the ambient illumination, the visual characteristics of the monitor itself and its direct surroundings, the ambient sounds, and the presence of other persons in the room. Research by Miwa and Hanyu (2006) has shown that interior design of a room, especially lighting conditions, influence persons’ mood, which spreads to their appraisal of objects and persons in that room. This effect, or misattribution of mood, has been discussed in Chapter 2.4.2.

Finally, the task of the viewer or user is a response moderator. When an individual visits a place, she has a goal or a plan, and has expectations about the place in relation to this plan (Russell and Snodgrass, 1987). Humans evaluate events, and most probably also environments, against factors
such as goal relevance and goal congruence (Lazarus, 1991; Oatley & Johnson-Laird, 1987; Scherer, 2005). Failure to attain the goals may therefore influence the appraisal negatively.

3.5 Affective appraisals, affective/emotional reactions and their measuring instruments

In the proposed framework people respond to the environment with internal responses and (sometimes) behavioural reactions (column 6 of the framework). In the internal response, a distinction is made between reactions to the environment (or the content of the representation) and the representation itself, as explained in Chapter 2.6. In the Stimulus-Organism-Response model the affective/emotional response is the “organism”, which mediates between stimulus and behavioural response, manifested in approach or avoidance behaviour.

Internal responses that are relevant for this research, are affective appraisals of the environment, affective/emotional changes (for instance a change in arousal or valence), and evaluation of the representation. Behavioural responses, such as approach and avoidance actions, were largely out of the scope of this study. Because they are indications of, respectively, positive and negative evaluations they can be used as measures in empirical research (Chapter 3.5.2).

This paragraph reviews the affective/emotional responses as well as the instruments to assess them.

3.5.1 Assessment of affective appraisals

Verbal rating scales and other verbal response measures

Studies on the affective appraisals and other affective/emotional responses to environments, have mostly used verbal self-reports methods like questionnaires and verbal rating scales.

Subjective feelings are the awareness of the emotional state one is in, so can be measured through self-report. An important advantage of verbal instruments is that rating scales can be assembled to represent any set of emotions, and can be used to measure combinations of emotions (Desmet, 2003).

The measuring instruments themselves have been subject of research, showing a number of problems, of which the most important are:

- verbal self-reports are either retrospective and thus give average of assessments over time, or represent a salient moment or location to which the viewer reacts (Zimmerman, Guttormsen, Danuser, & Gomez, 2003);
- or they are concurrent, interrupting the user during the task (Zimmerman et al., 2003);
- questionnaires can only assess the conscious experience of an affective/emotional response, but much of the affective experience is non-conscious (Zimmerman et al., 2003);
- emotional experiences are not primarily language-based. Cognitive effort is required to put emotional experience into words, and this effort can contaminate the measures (Hazlett & Benedek, 2007);
- words and expressions as used in verbal rating scales may be interpreted differently by individuals, while denying them the opportunity to choose an appropriate phrase (Hazlett & Benedek, 2007; Wergles & Muhar, 2009);
- participants may feel pressed to give answers e.g. for desirability reasons (Zimmerman et al., 2003);
- a verbal rating system makes it difficult to use in cultures with other languages, or be translated and again validated (Bradley & Lang, 1994);
- verbal rating scales may be long and time consuming, which may compromise the validity of the results (Poels & DeWitte, 2006).
An important category of measuring instruments is the semantic differential scale, first designed by Osgood, Suci and Tannenbaum (1957). It offers a set of bipolar adjectives derived from a factor analytic study of a large number of verbal responses describing the meaning that individuals attach to objects. Several researchers have raised doubts on the usefulness of the semantic differential and other verbal scaling techniques for appraising environments (Trent, Neumann, & Kvashny, 1987). Danford and Willems (1975) for instance have found that similar ratings for scenes can be obtained by presenting sites in reality, slides, or even merely a verbal indication of the function and location of the setting. In our view similarity in ratings can be explained by the place schemas, that, as discussed in Chapter 2.5, contain affective qualities associated with them as well. The results of the experiment however indicate that the measuring instruments do not reveal the effects of representational types, even when they are so widely different as slides and verbal descriptions.

Despite these objections, semantic differentials and other verbal rating systems are the most widely used measuring instruments for the appraisal of environments, and much effort has been spent on refining and validating scales. Some of the aforementioned disadvantages can be compensated by proper design of questionnaires. The objection that affect cannot be assessed by verbal measures may be partly rebutted; the words people use to describe current emotions show consistent patterns (Russell, 1980) and situational manipulations have been found to reliably elicit different patterns of self-reported emotional states (Larsen, Berntson, Poehlmann, Ito & Cacioppo, 2008). For unconscious appraisals, other methods can be used to complement the conscious response, as we have done in some of our experiments. In between the tasks graphic scales (such as the SAM, discussed later in this paragraph) can be introduced to assess a change in participants’ response over time. The tasks and instructions in the experiment must be verbalized carefully so participants do not get the impression that they may fail or be judged on their response.

When all these precautions are taken, it is important to assess the internal reliability of the scales before data analysis, and consider the data from the different sources together. In Chapter 5 we will reflect on the measures applied in the experiments, and possible consequences of the disadvantages mentioned on the conclusions of the experiments.

**Verbal rating scales developed for the appraisal of environments**

The scales used to assess the appraisal of environments can be divided into first, scales assessing the affective qualities and affective appraisals themselves, and second, scales judging types of physical attributes of environments that stand out in people’s perceptions and affective appraisals of environments (Nasar, 2008). Most scales however, contain adjectives that may fall into both categories, demonstrating that the distinction is very difficult to uphold when applied in real environments.

In Chapter 2.4, the semantic differential designed by Russell and Pratt (1980), has been introduced. It consists of 40, or in the shortened version, 20, adjectives that are rated on a (unipolar) scale from 1-8, that describe the affective qualities of an environment. The resulting 8 unipolar scales are joined into 4 bipolar scales by treating items from one unipolar scale as positively keyed, and items from its bipolar opposite scale as negatively keyed items of a single scale. The 8 scales represent the following qualities of environments, as perceived by viewers (adjectives between brackets):

- Arousing quality (intense, arousing, active, alive, forceful),
- Exciting quality (exhilarating, sensational, stimulating, exciting, interesting),
- Pleasant quality (pleasant, nice, pleasing, pretty, beautiful),
- Distressing quality (frenzied, tense, hectic, panicky, rushed),
- Sleepy quality (inactive, drowsy, idle, lazy, slow),
- Gloomy quality (dreary, dull, unstimulating, monotonous, boring),
- Unpleasant quality (dissatisfying, displeasing, repulsive, unpleasant, uncomfortable),
- Relaxing quality (tranquil, serene, peaceful, restful, calm)
The resulting values can be mapped on the circumplex model presented in Figure 2.3. The scale has been used extensively, also by other researchers (e.g. Kim & Richardson, 2003, Baloglu & Mc Cleary, 1999).

By means of factor analysis of a large number of responses to a variety of environments, Küller (1972, 1991) developed a verbal rating scale to measure and describe the experiences of (visual qualities) of the built environments systematically (Küller, 1972). The SMB-scale (Semantic Environmental Scale or, in Swedish, Semantisk Miljö Beskrivning) consists of 36 seven-grade rating scales giving estimates of the following eight qualities: pleasantness, complexity, unity (coherence), enclosedness, potency, social status, affection, and originality.

Table 3.2 The eight factors of the SMB scale (Küller, 1991)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
<th>Adjective positively correlated to the factor</th>
<th>Adjective negatively correlated to the factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasantness</td>
<td>The environmental quality of being pleasant, beautiful and secure</td>
<td>stimulating, secure, idyllic, good, pleasant</td>
<td>ugly, boring, brutal</td>
</tr>
<tr>
<td>Complexity</td>
<td>The degree of variation or, more specifically, intensity, contrast and abundance</td>
<td>motley, lively, composite</td>
<td>subdued</td>
</tr>
<tr>
<td>Unity</td>
<td>How well all the various parts of the environment fit together into a coherent and functional whole</td>
<td>functional, of pure style, consistent, whole</td>
<td></td>
</tr>
<tr>
<td>Enclosedness</td>
<td>A sense of spatial enclosedness and demarcation</td>
<td>closed, demarcated</td>
<td>open, airy</td>
</tr>
<tr>
<td>Potency</td>
<td>An expression of power in the environment and its various parts</td>
<td>masculine, potent</td>
<td>fragile, feminine</td>
</tr>
<tr>
<td>Social status</td>
<td>An evaluation of the built environment in socioeconomic terms, but also in terms of maintenance</td>
<td>expensive, well-kept, lavish</td>
<td>simple</td>
</tr>
<tr>
<td>Affection</td>
<td>The quality of recognition, giving rise to a sense of familiarity, often related to the age of the environment</td>
<td>modern, new</td>
<td>timeless, aged</td>
</tr>
<tr>
<td>Originality</td>
<td>The unusual and surprising in the environment</td>
<td>curious, surprising, special</td>
<td>ordinary</td>
</tr>
</tbody>
</table>

The scale contains many adjectives that relate directly to features of the environment and that can be considered as not purely affective, for instance “new”, or “open”. Five of the factors (complexity, unity, enclosedness, and social status) are very similar to the salient physical attributes recognized by Nasar and the concepts of Tveit et al. (2006), discussed in Chapter 3.3. Therefore the scale may reveal more of the motives for appraisal, so the salient physical attributes of the environment, than for instance the scale of Russell and Pratt; on the other hand it does not assess the affective response of the viewer on the level of “core affect”. The SMB-scale has been used in a number of experiments assessing different types of environments, as well as virtual representations of environments, for instance (Karlsson, Aronsson, & Svensson, 2003; Koramaz & Gulersoy, 2009; Küller & Wetterberg, 1996; Küller, Mikellides, & Janssens, 2009; Nordvik & Broman, 2009; Roupe, Wernemyr, Westerdahl, Suneson, & Allwood, 2001; Westerdahl, Suneson, Wernemyr, Roupe, Johansson, & Martinallwood, 2006).

The scales of Russell and Pratt and of Küller are the measures used most often in research on the appraisal of environments. Other researchers have adapted these scales or developed new
scales, of which we describe the most important here. Rohrmann and Bishop (2002) and Bishop and Rohrmann (2003) have assessed impressions and affective qualities of virtual and real environments using a scale based on the scale described by Russell and Pratt, which they modified after pre-tests. Two factors were added, that are not only affective, but also characterize the environment: order and spaciousness. A third factor added is familiarity, which describes the relation between the viewer and the environment, and therefore does not describe a purely affective quality of the environment either.

The factors (attributes between brackets) are Pleasure (beautiful, pleasant, pretty), Naturalness (artificial, natural, unnatural), Familiarity (familiar, unfamiliar, well-known), Order (orderly, spacious, understandable), Inertia (drowsy, lazy, slow), Arousal (active, arousing, intense), Threat (forceful, threatening, unpleasant), and Disliking (displeasing, dissatisfying, uncomfortable). The negative contributors are in italics. They are rated on a scale from 1–7.

Vogels (2008a, 2008b) developed a measuring instrument of 38 (Dutch) adjectives to quantify experienced atmosphere, which includes four factors,

- coziness, which is related to items like “cozy”, and “pleasant”;
- liveliness, which is related to e.g. “exciting” and “inspiring”;
- tenseness, related to e.g. “tense” and “terrifying” and
- detachment, related to e.g. “businesslike” and “chilly”.

The instrument of Vogels was developed to measure ambience in indoor environments with different lighting conditions. It contains mainly affective qualifiers, like the scale of Russell and Pratt, however some of the adjectives can be considered as typifying features of the environment that are not purely affective, such as “businesslike”. Although the four factors do not overlap completely, they can be recognized in dimensions arousal and pleasure in the circumplex model of affect (Russell, 2003). The discrepancy may be the accumulation of effects of several differences: the focus of Vogels’ measure (indoor environments versus outdoor environments), differences between languages (Dutch versus English) and the development in language over time (Vogels developed her measure 10 years later). Also the selection criteria for including the adjectives may have played a role: Russell and Pratt were strict in selecting only what they consider as purely affective descriptors.

De Kort, IJsselsteijn, Kooijman and Schuermans (2003) presented viewers of virtual and real environments 29 bipolar adjective 6-point items. The results indicate five factors: evaluation, ambience, arousal, privacy and security. Again, the items include adjectives that can be considered objective characteristics of environments, such as “bare”-“decorated” and “grey”-“colourful”.

The measures used by Rohrmann and Bishop (2002) and by Kort et al. (2003) were developed to assess different types of environments, also outdoor environments. Like the SMB-scale they include some qualifiers to describe features of the environment that viewers consider a source for their appraisal. So with these scales, viewers for instance also assess the degree of naturalness or openness of an environment, and not only whether the naturalness or openness makes them feel pleasant or aroused. They are therefore more descriptive of physical attributes of the stimuli, whereas the scales of Russell and Pratt and Vogels reflect the affective qualities of the environments.

When we started this research, the scales of Russell and Pratt and Küller were used most often as measures in research on the appraisal of environments. We decided to use them as examples of the two classes of scales distinguished in this paragraph, the first revealing the purely affective appraisal of an environment, the second including the assessment of salient physical attributes. We selected them for our experiments, described in Chapter 4.2-4.10. The SMB-scale of Küller was used for the experiments using representations of built environments, when understanding of the judgment of participants of the salient physical attributes of the environment was relevant for the goal of the studies. The scale of Russel and Pratt was used for virtual outdoor environments, in
which we also investigated aspects of the affective state of the participants, for instance their level of arousal or engagement.

**Appraisal of environments: open-ended questions**

In addition to verbal rating scales, other methods to elicit verbal response are used to assess the appraisal of environments. Wergles and Muhr (2009), following Bishop and Rohrmann (2003) and Scott and Canter (1997), advocate the use of open questions so participants are incited to recall and conceptualize what they have seen, and verbalize the information in their own words. This provides richness in descriptions and a wide range of reactions, providing extra information, and possibly clarifying results of scales as well. Open-ended questions are possibly more successful in discriminating between representational variants, as suggested by Trent et al. (1987). In their study viewers of architectural sites in reality offered significantly more descriptions of ephemeral aspects of the scenes, and slightly more evocative (subjective) responses, than viewers of slides of the same sites; the answers to closed questions however did not differ significantly between the two conditions.

A second argument in favour of open-ended questions is that similar ratings for an environment or its surrogate do not necessarily mean that the motivation for the assessment is the same. Different determinants may result in similar assessments, especially if the measurement scale used is little differentiated, for instance when preference is measured. In their study, Wergles and Muhr (2009) therefore focused on revealing which elements determine the observers’ evaluation; “whether it is indeed the same that they like or dislike in the depicted and the real environment”. An obvious disadvantage is that the answers to open-ended questions are more difficult to analyze, and that the results cannot quickly be compared for different environments.

In the experiments described in Chapter 4.2 to 4.5, we have asked viewers to describe the environment in their own words, which led to new verbal assessments not included in the scales presented. Participants for instance often used words like “inanimate” and “lifeless” to typify the virtual environment, which is very to the point for 3D visualisations.

**3.5.2 Assessment of affective/emotional reactions and engagement**

**Scales and questionnaires**

We have not only assessed the affective appraisal of environment, but also emotional responses or reactions to the environments presented, and emotional states such as engagement.

For effects on the arousal and pleasure of participants, we used visual self-report and physiological measures. For engagement, questions from standard questionnaires can be used, for instance from the ITC-Sense of Presence Inventory (ITC-SOPI). The ITC-SOPI is developed to measure users’ experiences of media, without reference to objective system parameters (Lessiter, Freeman, Keogh, & Davidoff, 2001). It consists of four factors; engagement, sense of physical space, ecological validity, and negative effects. To assess the level of engagement we also used questions on involvement (the attention devoted to the virtual environment and the involvement experienced) of the Igroup Presence Questionnaire (IPQ, see Schubert, Friedman, & Regenbrecht, 2001).

**Visual self-report**

To address problems that may occur with verbal response measures noted earlier, there have been attempts, for instance in HCl, marketing research and product design, to develop better self-report measures, such as visual scales. The Self-Assessment Manikin (SAM) is a picture-oriented scale and is often used because it is easily and quickly administered (Bradley & Lang, 1994). The SAM consists of three dimensions, pleasantness, arousal and dominance. For each dimension 5 cartoonlike figures are used. Respondents indicate which figures best represent their emotional states. Due to lack of effects found on the dominance dimension, the dominance dimension is often not used, or the results are not reported (Poels & Dewitte, 2006). We used the SAM in three experiments, described in Chapter 4.5, 4.6 and 4.7.
Visual self-report reduces the cognitive effort associated with verbal measures, but does not completely eliminate it. Verbal instructions are still required to administer the scales, and may influence the perception of an emotional reaction by the participant (Poels & Dewitte, 2006).

**Behavioural measures**

To record environmental perceptions and evaluations, self-report measures can be supplemented with behavioural and physiological measures. For evaluative appraisals, verbal measures alone identify affective qualities of environments, which are, in the words of Lazarus (1984), cold cognitions that lack emotional involvement. Behavioural and physiological measures may help establish the level of involvement (Nasar, 2008).

Behavioural measures can involve observing how long individuals stay in one area, where they look at, which way they choose to go when confronted with an intersection, etc. They are based on the approach-avoidance distinction that is well known in psychology since the earliest scientific studies (Elliot, 2006), and adopted by (among others) Mehrabian and Russell (1987). These kinds of measures are extensively used in the domain of marketing research, to assess the effects of atmospherics (the physical elements contributing to the store’s ambience) on the consumer’s behaviour (e.g. Bitner, 1992; Turley & Milliman, 2000; Yani-de-Soriano & Foxall, 2006). They examine the effects of environmental cues on affective states, which then result in approach or avoidance behaviours, according to the Stimulus-Organism-Response Model (e.g. Eroglu, Machleit, & Davis, 2001; Eroglu, Machleit, & Davis, 2003).

These measures were not used in the experiments described in our study, except in one experiment described in Chapter 4.6; this concerns the examination environment for Road Inspectors, where participants were asked to cross a road in the virtual environment as soon as they thought it was safe. For the experiments assessing the affective qualities of environments, the use of verbal rating scales is more appropriate; in the other experiments assessing the emotional response to virtual environments behavioural measures could not be included into the experimental tasks adequately.

**Psychophysiological measures**

Emotions are accompanied by bodily reactions, including facial expressions and physiological reactions caused by changes in the autonomic nervous system. For instance, facial muscle movements, heart rate, and skin conductance can be determined to measure arousal, valence, or discrete emotions such as fear and anger (Tullis & Albert, 2009; Larsen et al., 2008; Gunes & Pantic, 2010). Facial expressions can be observed and classified using a coding system, but it has been argued that changes in muscular activity evoked by some stimuli are often too subtle to be measured with this method, and individuals can mask the emotions they are experiencing (Bolls, Lan, & Potter, 2001; Larsen et al., 2008). Facial electromyography (EMG) is a more precise measure of facial muscle activity. Two facial muscles are generally involved in this method, the corrugator muscle on the brow, involved in frowning, and the zygomatic muscle around the cheeks, controlling smiling. In the experiment “Visual dynamics and sounds in an outdoor environment” (Chapter 4.5) we monitored these two muscles as indicators of the affective/emotional response to a virtual environment. Facial activity in zygomatic and corrugator regions increases for pleasant and unpleasant stimuli, respectively (Hazlett & Benedek, 2007; Poels & de Witte, 2006) and is related to arousal as well (Ribeiro, Teixeira-Silva, Pompéia, & Bueno, 2007).

Other physiological measures often used in psychological research are skin conductance and cardiovascular responses. Skin conductance reflects the level of sweat at the surface of this skin. Sweat secretion increases as a result of more activation of the autonomic nervous system, which is an indicator of arousal (Ravaja, 2004). Cardiovascular measures include for instance heart rate and blood pressure, and are also indicators of arousal. Heart rate was measured in one experiment (Effects of stress and darkness on the appraisal of a virtual environment, Chapter 4.9), to confirm that the task which was meant to make participants feel stressed, had been effective.

Physiological measures are valued as objective measurements of emotional reactions, which may occur unconsciously and fleetingly. They can be performed continuously during an experiment,
are not dependent on language, do not require memory, and do not interfere with message processing (Ravaja, 2004). However, several limitations of these physiological measures are recognized.

First, the autonomic nervous system varies not only with emotional responses, but with other mental states such as effort, and attention (Berntson & Cacioppo, 2000). Ecological validity may be compromised because specialist equipment is required, usually in a lab, and some instruments used are considered unpleasant by participants. Also, the interpretation of the data is often complex, because the variations in physiological reactions may be caused by other factors than the experimental stimuli (Poels & de Witte, 2006). The recent overview of Gunes and Pantic (2010) on methods for automatic emotion recognition also shows the complexity of this topic, and the many different problems that occur in linking bodily and facial expressions and available physiological measures with specific affective states. In our experiments, physiological measurements were always combined with self-report measures and observation, as often advised (e.g. Herbelin, Benzaki, Riquier, Renault & Thalmann, 2004).

3.6 Summary

In this chapter we presented a framework for understanding and studying the affective appraisal of and affective/emotional response to virtual environments (Figure 3.2). We first discussed environmental dimensions and features as determinants of affective appraisal in real environments (column 1 in the framework), and chose three categories to accommodate the variables of our experiments: spatial layout, sign and symbols, and ambient conditions. Next we described modifiers and determinants related to the representation and specific to 3D models of environments (column 2), and their effects on the viewers’ response. User states and characteristics are considered as response moderators (column 4), influencing viewers’ internal responses to the virtual environment (column 5) and their behavioural responses (column 6). We discussed advantages and disadvantages of measuring instruments for the assessment of the affective appraisals of environments, particularly verbal and psychophysiological measures.

Using the framework explained in this chapter, we designed and conducted a series of experiments using different 3D visualisations and virtual training environments, which constitute Chapter 4. The comparative discussion of the results is presented in Chapter 5.
4 Experiments on the affective/emotional response to virtual environments
4.1 Introduction

The first part of this research project was exploratory, combining studies of the use of 3D models for visualisation and training, and of literature on environmental and cognitive psychology, human-computer interaction, and media theory. It led to a framework of determinants of affective appraisals and responses of virtual environments (Chapter 3).

In this second part we describe the experiments conducted to examine distinct elements of the framework, each with a specific hypothesis.

Goals in the second part of this research are to
1. examine effects of a selected number of determinants on the affective appraisal and response of the viewer;
2. determine the validity and reliability of some experimental designs and the measuring instruments used; and
3. assess the correctness and appropriateness of the framework of determinants. Does the framework provide a useful and comprehensive schema for studying the viewer’s response to 3D environments and for understanding the effects of the determinants?

These objectives require a broad and exploratory approach in the design of the experiments. This approach provides evidence of both quantitative and qualitative kind, also on contextual variables that are not purposefully included in the design and hypotheses, but that occur as a result from the chosen material or setting in the real world. A narrow focus and in-depth research of a small number of proposed determinants may lead to answers to only very restricted research questions. With a larger diversity of evidence, gained from the experiments, we expected to attain a higher level of comprehension and to be able to draw general conclusions. The resources available for the experiments were thus divided over eight experiments.

The experiments were performed between 2004 and 2010. They took place either in a lab environment (taking into account the context of use of the virtual environment in reality), or in the field. As a consequence, not all external conditions and variables were under control of the experimenter, but this was a deliberate choice. Robson (2002), quoting Andersen al. (1991): “... you might be better able to vary anticipatory stress experimentally and control other factors in a laboratory study, but there is much to be gained by transferring the enquiry to the dentist’s chair”. Real world research, applying both quantitative and qualitative designs, has received increasing attention and recognition in social research in the last decades (Robson, 2002). In this research project, quantitative and qualitative designs were sometimes used together in so called mixed designs. This approach also has disadvantages, such as the occurrence of confounding variables and the influence of experimental “noise”, which may result in small experimental power and difficulties in drawing conclusions. We deal with these issues in the individual experiments as far as possible.

The experiments assessed effects of modifying factors in the representation and of response moderators, on the affective appraisals and response of the viewers. The design of the experiments is based on the framework of determinants presented in Chapter 3. The framework contains a large number of potential dependent and independent variables. We could not proceed from results of earlier empirical work in this domain, because this is unfortunately scarce. Therefore, we made a selection of evidently important variables, based on our observations in several projects and on acknowledged issues in the development of 3D environments. A third factor that determined the selection was the limited availability of appropriate 3D environments. As a result of this approach, the independent variables show a variation over the categories “Environmental dimensions and features”, and “Representational modifiers and determinants”. The eight major experiments are discussed separately in Chapter 4.2 to 4.9. Table 4.1 lists the categories of determinants addressed in the experiments.

The dependent variables are related to the domains chosen in this research: first, the affective appraisal of environment, which is relevant for visualisation of the environment, and secondly, affective/emotional reactions important in serious gaming, such as arousal and engagement.
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Environmental dimensions and features</th>
<th>Representational modifiers and determinants</th>
<th>Response moderators</th>
<th>Affective/emotional response</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2. Decay in an urban environment</td>
<td>Ambient conditions: weather</td>
<td>Choices in modeling process: style, level of detail, time of day, weather</td>
<td>Expectations</td>
<td>Affective appraisal of environment</td>
</tr>
<tr>
<td></td>
<td>Symbols: maintenance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.3. The influence of lighting on the affective qualities of a virtual theater</td>
<td>Symbols: décor</td>
<td>Choices in modeling process: illumination</td>
<td>Pre-existing mental representations</td>
<td>Affective appraisal of environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.4. Comparing affective appraisals of real and virtual environments</td>
<td></td>
<td>All modifying factors</td>
<td>Pre-existing mental representations</td>
<td>Affective appraisal of environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.5. Visual dynamics and sounds in an outdoor environment</td>
<td>Ambient conditions: sounds, weather</td>
<td>Restrictions of the medium: loss of auditory information</td>
<td>Affective appraisal of environment, engagement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Choices in modeling process: dynamic elements</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.6. Experiencing danger on a virtual highway</td>
<td>Ambient conditions: sounds</td>
<td>Restrictions of the medium: loss of auditory information, absence of body</td>
<td>Emotional reaction, perceived realism</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Choices in modeling process: dynamic elements, perspective</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Modified by platform: field of view, depth perception</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.7. The effect of sounds on engagement and arousal in a virtual training</td>
<td>Ambient conditions: sounds</td>
<td>Restrictions of the medium: loss of auditory information</td>
<td>Emotional reaction (arousal, engagement) perceived realism</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.8. The influence of cybersickness on the affective appraisal of a virtual environment</td>
<td>Modified by platform: field of view</td>
<td>Degree of cybersickness</td>
<td>Affective appraisal of environment</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.9. Effects of stress and darkness on the appraisal of a virtual environment</td>
<td>Ambient conditions: time of day</td>
<td>Choices in modeling process: lighting</td>
<td>Mood and emotional state</td>
<td>Affective appraisal of environment, emotional reaction (anxiety) Physiological response</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The design of the experiments varied, depending on the specific questions to be answered, on the materials used, and the type of participants and location:

**Materials**
- The 3D environments used for six of the experiments (4.2 – 4.7) were developed by companies specialized in visualisation and serious gaming, ensuring the state-of-the-art quality of the models at the time of the experiments. The models were produced for a specific purpose and for users who are experts in a specific domain, which provided us with research questions and hypotheses that are also externally valid. The other two models (4.8 and 4.9) were produced with professional 3D software and game engines especially for the purpose of this research.
- Most environments were displayed on desktop monitors. In experiments 4.6, 4.7 and 4.8, the environments were displayed on projection screens, as usual in training situations. The size of the projection screens was small, and not intended to create (physical) immersion.

**Participants**
- Experiments 4.6 and 4.7 were conducted on location, and with domain users. It was expected that the effects of the determinants were also dependent on the domain knowledge of the users, in this case leading firefighters and traffic officers. In experiment 4.9 only young male participants were recruited, because cortisol measurements were made. Cortisol levels are known to differ for age and gender. Participants in the other experiments were mostly recruited from participant databases, including university students, with an age range of ca. 18 – 80 years old. An effort was made to select participants with a relevant background and interest and a roughly equal balance of women and men.

**Location**
- One of the conditions of experiments 4.3 and 4.4 concerned a real environment, to allow comparison of the model with the environment. The other experiments were conducted in a laboratory setting or in a separate room in a more informal office environment. Experiments 4.6 and 4.7 were conducted on location.

In paragraph 4.10, a summary of conclusions is presented. The patterns that emerge from the aggregated results lead to generalized conclusions in Chapter 5 on the three research goals identified at the beginning of this paragraph:

1. examine the effects of a selected number of determinants on the affective appraisal and response of the viewer;
2. determine the validity and reliability of some experimental designs and the measuring instruments used; and
3. assess the correctness and appropriateness of the framework of determinants.

The experiments are divided into three clusters: the first four experiments concern the effects of environmental and representational determinants on the affective appraisal of environmental visualisations; two experiments concern the effects on the affective/emotional reactions to virtual environments used in a training context, and the final two experiments study the effects of response moderators on the appraisal of virtual environments, especially environmental visualisations. The bar indicates to which cluster the experiments in paragraphs 4.2-4.9 belong.

![Cluster Bar Chart](image-url)
4.2 Experiment: Decay in an urban environment


4.2.1 Introduction

The experiment focuses on the bright, clean, static and lifeless appearance of many 3D environments used for visualisation of urban areas, often commented on by users who are not familiar with computer graphics (Rekittke & Paar, 2009). It is a consequence of restrictions of hardware and software, as well as a continuation of the style of the artist’s impression common in architectural visualisations. This evident discrepancy with real environments, which are disorderly, dynamic and lively, raises questions on the validity of the models when used for assessment of subjective experiences in urban areas, such as aesthetic quality or perceived safety.

To study the effect of these elements in a 3D model we changed the general impression of cleanliness and good maintenance in a 3D model of an urban area in Rotterdam, which included an old industrial harbour area and a new business district. We added street furniture, traffic equipment, signs of vandalism; we also adjusted textures and adapted the ambient conditions from bright to dull weather. The manipulations in the old harbour area were designed to create a disorderly effect, and were expected to have an overall negative effect on the appraisal. The manipulations in the second area, the business district, did not include evidently negative elements. They were expected to make the environment more complex and original.

The manipulated variables belong to two categories of the framework in Table 3.2: environmental dimensions and features, and representational modifiers.

4.2.2 Related Work

Developers of 3D architectural visualisations often by convention adopt a style that is based on the artist’s impression familiar from earlier media. The style supports a general goal of 3D visualisations to inspect sightlines and the position of buildings in relation to other buildings, understand the layout of an area, and assess the main visual characteristics of individual buildings. In this style, bright, diffuse lighting and advantageous colours are applied; objects that can block the users view, and details in façades or street elements are omitted; indications of use and of the presence of people are absent or simplified. As a result, the general appearance of the represented area is orderly, uncluttered, bright, clean, and impersonal. In a real environment these visual characteristics would obviously influence the viewer’s appraisal.

In reality, manmade nuisance objects such as wires, poles, signs and cars, and signs of dilapidation (litter, weeds, rubbish) evoke negative affective response (Nasar, 1990). Physical incivilities, either deliberate (graffiti, vandalism) or inadvertent (litter, unkempt housing), are symbols of social disorder, are disliked, and are linked to perceptions of unsafety (Hanyu, 1997; Austin, Furr, & Spine, 2002; Park, Spicer, Guterres, Brantingham, & Jenion, 2010). Humans in general prefer naturalness, spaciousness and upkeep in environments (Nasar, 1987). The nuisance objects
influence the visual complexity and coherence of the area, which are salient physical attributes of environments (Chapter 3.3). Physical characteristics alone or together, contain clues for the interpretation of the content of forms, a style, a type of building or its function. These elements affect the viewers’ internal representation of the environment, which consists of the perceived environment and associations (Chapter 2.5).

In this experiment we try to establish in which way the style commonly used for 3D models of the built environment influences the viewers’ affective appraisal of the represented area. To assess the appraisals, we have used the SMB-scale, discussed in Chapter 3.5. The SMB-scale (Semantic Environmental Scale) comprises 36 adjectives in eight factors that describe the experience of persons in a man-made environment (Küller, 1977; Küller 1991). The factors are: Pleasantness, Complexity, Unity, Enclosedness, Potency, Social Status, Affection, and Originality. Another goal of this experiment is to assess the reliability and suitability of the SMB-scale, and the tasks designed, for future experiments in this line of research.

4.2.3 Method

Participants

The participants were 12 women and 12 men, aged 19 – 55. Mean age of the participants who viewed version A was 30.6 years (SD= 13.2); of the participants who viewed version D, 34.3 years (SD=14.9). They were university and college students or had an academic degree. All were familiar with the use of personal computers, but no one had experience in creating or manipulating 3D environments. They were paid a small amount of money for participating. The participants were randomly assigned to the conditions.

Manipulations

A 3D model of the “Kop van Zuid” district in Rotterdam, covering ca 200 by 800 m, was provided to us by VSTEP BV. The Kop van Zuid district is situated in the Nieuwe Maas river, and was part of an extensive harbour area. It still features warehouses and other industrial buildings of ca. 1900 (Figure 4.1)

![Figure 4.1 Photograph of the Kop van Zuid area, a) old harbour 2004, b) business district 2004](image)

The model represents two adjacent areas: 1. the old harbour area, including some of the modern buildings that are being constructed since the beginning of this century (Figure 4.1a), as well as the prominent Erasmus bridge (1996) and 2. a small part of a modern business district adjoining it (Figure 4.1b). The water surrounding the area has a rippling surface, which reflects the sky and gives the area views over the river and the surrounding city. The model is a good example of the common style for this kind of model, showing clean and even surfaces, pleasant colours, few details and favourable weather (Figure 4.2).
After visiting and photographing the actual area in Rotterdam we performed more than 300 manipulations in the original model, by changing textures, and adding objects. The manipulations represent two clusters of variables that oppose the two most obvious characteristics of unmanipulated model: the orderly cleanliness, and the general brightness. This resulted in three new models, as shown in Table 4.2 and Table 4.3.

Table 4.2 The four versions of the model.

<table>
<thead>
<tr>
<th>Version</th>
<th>Environmental decay</th>
<th>Sky</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (original)</td>
<td>no</td>
<td>blue</td>
</tr>
<tr>
<td>B</td>
<td>no</td>
<td>grey</td>
</tr>
<tr>
<td>C</td>
<td>yes</td>
<td>blue</td>
</tr>
<tr>
<td>D</td>
<td>yes</td>
<td>grey</td>
</tr>
</tbody>
</table>
Table 4.3 Examples from the different versions.

<table>
<thead>
<tr>
<th>Version</th>
<th>old harbour area</th>
<th>new business district</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (original)</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
</tr>
<tr>
<td>B</td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
</tr>
<tr>
<td>C</td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td>D</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
</tr>
</tbody>
</table>

Note: In the old harbour area vandalized street furniture and building cranes are added (versions C and D). The sky is made grey (versions B and D). In the view towards new area, fences and graffiti can be seen and the colour of the windows is varied (versions C and D). The sky is made grey (versions B and D).

The uniform textures that indicated clean surfaces of roads, pavements and walls, were replaced to give an impression of worn down and stained materials. Secondly, many objects were added such as:
- cars, parked in parking areas as well as occasionally on the pavement
- fences, large building cranes, piles of building material, as indications of building activity, at the back of Hotel New York,
- common street furniture, like bins, poles, wires, parking meters, a stained glass recycling bin and traffic signs.

Moreover, we introduced clearly negative elements (Figure 4.3, a and b):
- litter like discarded building materials, cardboard boxes, fire extinguishers
- graffiti on walls, in conspicuous places and of large format
- vandalized benches
- lantern posts that were partly pulled down.

The manipulations in the part of the model that depicts the old harbour area were intended to give a strong impression of dilapidation and neglect. In the second area, the new business district, the objects added were more neutral, such as large billboards and traffic signs. In both areas they created a more varied and visually irregular, less orderly effect.

To decrease the impression of pleasant, sunny weather, the blue sky was replaced by grey cloud (Figure 4.4). To match this lighting condition, the colours of several buildings were darkened, which also diminished the impression of a clean and well-kept area.

Together, these two clusters of manipulations create a more verisimilar and convincing representation of the area. In the first experiment we compared the two extreme conditions, due to limited availability of participants, so the version of the model presenting both clusters of manipulations (D) with the original model (A). In a later, smaller experiment, we also compared versions B and C. In version B only the sky was changed to grey, and in version C only the man-made nuisance objects and adjusted textures were included.

During the development of the models and preparation of the experiment, we consulted several individuals, experts as well as laypersons; when the versions were presented to them directly after each other, they considered the visual difference between the models extensive, obvious, and convincing.
**Equipment**

The model of the Kop van Zuid Area was created in Quest 3D and manipulated using Maya and 3DStudio Max. The experiment was conducted on a desktop computer with a 17” screen. The computer was a Pentium 4, running at 2.4 GHz, 512 MB RAM and supporting a GeForce 6600GT. The participants used keyboard and mouse to navigate the model. The model did not include any sounds.

**Materials and procedure**

The original version (A) and the manipulated version (D) were presented to the participants in a between subjects design, with 12 participants individually tested in each condition. After a short exercise to help them get acquainted with navigating a 3D model, they were given two simple tasks that included “walking” from one point in the model to two targets, using an isometric map, with suggested viewpoints (Figure 4.5). The first task led the participants through the old harbour area, the second, shorter walk, through the new business district. On reaching their targets the participants were asked, at both occasions, to rate the adjectives belonging to the SMB-scale to determine the affective appraisal of the area just seen. The adjectives were rated on a 7-point scale, ranging from “not at all”, through “neutral”, to “very much”. Also, they were asked to estimate the time they spent navigating, because a longer stay in the model might relate to higher interest of the participant in the environment. Assessments of distance traversed, distances between buildings, and building heights were asked. Other questions concerned the participants’ impression of the area in their own words, and notions about the probable function of the area, that might explain the ratings on the scale. Some questions addressed the represented time of day and weather, and the perceived quality of the model.

![Figure 4.5 Map of the area presented to the participants. a) the first part of the route, through the old harbour area. b) the second part of the route, through the new business district.](image)

**Hypotheses**

The manipulations performed in the first part of the area, the old harbour area, were designed to create a disorderly effect, and expected to negatively influence the SMB-factors pleasantness, unity, and social status, and to increase the factors complexity, and originality, as they would in a real environment.

Pleasantness is measured by 8 adjectives, for instance “good”, “pleasant”, “secure” and “idyllic”. The indications of vandalism and neglect were therefore expected to have a negative effect on this factor.

Visual unity and consistent style, measured in the factor Unity, was expected to be high in the original model and lower in the manipulated version D because the added elements disrupt the shape, colour, and style of the buildings and streets.
Social status, measured with for instance “well-kept”, and “expensive” was assumed to be negatively influenced by the indications of vandalism and neglect.

Complexity, meaning the environment’s liveliness and complexity, was supposed to increase as a direct result of the increase in visual variation.

The factor Originality, with adjectives such as surprising, and (negatively correlated) ordinary, was expected to result in higher scores, since the original model contained very few details and was rather monotonous and dull.

The manipulations in the second area, the new business district, were supposed to decrease the factor Unity, and increase Complexity and Originality, for the same reasons explained above. In this area we did not expect a negative effect on perceived pleasantness, because the manipulations were of a more neutral character, such as billboards, and traffic signs.

4.2.4 Results

Affective appraisal measured by the SMB-scale

The old harbour area. Since the main purpose of this experiment was to assess the effect of negative elements in the visualisation, we concentrate our discussion on the results of the old area, from Hotel New York to the “Nieuwe Luxor” Theatre, where most manmade nuisance objects and physical incivilities were inserted.

The SMB-scale is a validated instrument for the assessment of environments in reality, not yet for virtual environments. We performed a reliability analysis to check the internal consistency of the items (measuring a factor) for the environments used in this experiment.

Cronbach’s alpha for Pleasantness (.75), Enclosedness (.60), are not high, but sufficient. The results for the other factors were however lower, and in some cases so low, that even removing items did not result in a value higher than .6. Deleting items provided the following results: Unity: deleting item “functional”, .77; Potency: deleting item “potent”, .71; Affection: deleting item “timeless”, .64. The factors Complexity and Social Status, and Originality could not be improved to a satisfactory level by removing one or even two items. In 7 factors, the reliability was lower for the manipulated version than the original version.

Comparing the results on the SMB-scale for versions A and D, the T-test only shows a significant difference for the factor Social Status, although it also seems to confirm our hypothesis for Pleasantness (Social Status p=.01, Pleasantness p=.06, one-tailed). The results of the data analysis for the old part of the area are presented in Table 4.4. The last column shows the p-value. The critical p-level or significance level is the amount of evidence required to accept that an event is unlikely to have arisen by chance; if there is only a 5% chance (a probability of 0.05) of something occurring by chance we can accept it is a genuine effect (Field, 2009). Figure 4.6 shows a graph of the scores on the SMB scale, with adjustments, for the old part of the area.
Table 4.4 Mean scores on the factors of the SMB-scale.

<table>
<thead>
<tr>
<th>SMB-Factor</th>
<th>Version</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasantness</td>
<td>orig.</td>
<td>3.64</td>
<td>.91</td>
<td>1.57</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>3.11</td>
<td>.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>orig.</td>
<td>3.46</td>
<td>.95</td>
<td>.06</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>3.44</td>
<td>.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unity</td>
<td>orig.</td>
<td>5.15</td>
<td>.89</td>
<td>.52</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>4.96</td>
<td>.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unity - adjusted</td>
<td>orig.</td>
<td>5.06</td>
<td>1.14</td>
<td>.24</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>4.94</td>
<td>1.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enclosedness</td>
<td>orig.</td>
<td>3.40</td>
<td>1.09</td>
<td>-5.8</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>3.63</td>
<td>.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potency</td>
<td>orig.</td>
<td>5.02</td>
<td>.75</td>
<td>-68</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>5.23</td>
<td>.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potency - adjusted</td>
<td>orig.</td>
<td>5.11</td>
<td>.90</td>
<td>-1.07</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>5.50</td>
<td>.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Status</td>
<td>orig.</td>
<td>3.96</td>
<td>.70</td>
<td>2.43</td>
<td>.01*</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>3.35</td>
<td>.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affection</td>
<td>orig.</td>
<td>5.46</td>
<td>.83</td>
<td>-18</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>5.52</td>
<td>.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affection - adjusted</td>
<td>orig.</td>
<td>5.72</td>
<td>.94</td>
<td>-2.4</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>5.81</td>
<td>.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Originality</td>
<td>orig.</td>
<td>3.88</td>
<td>.79</td>
<td>-1.13</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>manip.</td>
<td>4.21</td>
<td>.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.6 Mean scores on the adjusted SMB-scale, old area. The lines between the measuring points are only added to facilitate comparison of the results.
The new business area. The elements introduced in the second part of the area, which has the appearance of a modern business district, are not so evidently negative, and consist of some graffiti as well as more neutral elements such as billboards and traffic signs. Figure 4.7 shows the means of the SMB-factors for this area. Although the scores do confirm our hypotheses (lower for Unity, higher for Complexity and Originality), no significant differences were found on the factors.

![Figure 4.7 Mean scores on the adjusted SMB-scale, new area. The lines between the measuring points are only added to facilitate comparison of the results.](image)

**Verbal comments on the areas (response to open questions)**

In the original version the most commonly used phrase to describe the environment was “ongezellig”, a Dutch word indicating “not cozy”, or “not intimate” (8 times). For the manipulated model “dull” was used 7 times, for the original version 5 times. The number of elements and details added was obviously not enough to significantly change the appearance of the rather desolate area. Words that described the areas as inanimate or desolate were used 5 times for the manipulated version, and twice for the original model. Maybe the traces of human presence make the absence of people more noticeable, especially since they are negative, such as graffiti and vandalism. Thus, in both versions the areas were considered as inanimate and rather boring. The additions apparently did not offer enough visual interest to the viewer, and the absence of people and movement, apart from the surface of the river and a car passing by now and then, is in both versions a dominant feature.

**Time of day, weather and season.** The models do not include evident indications of the time of day, weather and season. This did not bother the participants during the experiments. Most readily answered the questions on this subject based on their memories and interpretation of light, traffic condition, etc. For the original model (blue sky) most often the middle of the day was mentioned (7), whereas for the manipulated model, that shows a grey sky, most participants suggested daytime in general, or (late) in the afternoon (9). Some participants interpreted the quietness in the streets (lack of people and very little traffic) as an indication of either early morning, early evening, or a Sunday. Only one participant (in the manipulated version) thought an evening situation was depicted.

The blue sky of the original model is most often used to infer spring or summer (5). The grey sky of the manipulated model is more confusing: 4 participants thought it represented autumn or winter, and 5 participants answered that the models did not include information on the season or weather. The blue sky and supposed pleasant weather might have a positive effect on the viewer,
that influences the affective appraisal of the area. The effect will be studied separately in a next experiment.

Quality of the models. The models were considered convincing: the original model was called realistic or convincing by 7 participants, the manipulated model by 9. One would expect that the manipulated model would score higher on realism, but the number of participants is too small to be definitive on this point. The results do show that the manipulations were well performed and were not conspicuous as elements that were added separately. In verbal comments, participants mentioned their appreciation of details in the manipulated versions such as billboards, specific car types, and graffiti. A participant even remarked that she felt curious to inspect these details, such as, for instance, check the contents of the bins.

Perception of time and distances

The participants were asked to estimate the time they spent in the model, walking from Hotel New York to the Luxor Theater, and this was compared with the actual time they used. Participants navigating the original model thought they spent more time in the model than participants navigating the manipulated model, whereas in fact a shorter period elapsed. A paired samples t-test showed that the difference was significant ($t(11)=3.02$, $p<.05$) (Table 4.5).

<table>
<thead>
<tr>
<th></th>
<th>Mean time in model (secs)</th>
<th>Mean time estimated by participant (secs)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original model</td>
<td>334</td>
<td>528</td>
<td>194*</td>
</tr>
<tr>
<td>Manipulated model</td>
<td>421</td>
<td>450</td>
<td>29</td>
</tr>
</tbody>
</table>

A difference was also found in the estimates of distances in the models. Table 4.6 shows the results for both models. Dist1 represents the estimates by the participants for the total length of their route from Hotel New York to the first target, the Nieuwe Luxor Theater. Dist2 and Dist 3 are egocentric distances, between the viewer and another building. Dist 4 is the estimate of the distance between two buildings, from another position, thus an exocentric distance. In the manipulated version all estimates were higher. For Dist2 and Dist4, this difference was significant (resp. $t(21)=-2.57$, $p<.05$ and $t(21)=-3.07$, $p<.01$). Moreover, in the manipulated version the estimates are closer to the distances in reality. Adding details and objects obviously provided the participants with more cues for this task. Whereas dimensions of roads, buildings, and even elements such as doors and windows may vary considerably, added elements such as cars, bins, and traffic signs, were easier to interpret and were used as an aid for estimating sizes and distances. However, the difference in distance perception does not seem to have influenced the participants’ experience of space, openness, etc., in the SMB-scale.
Table 4.6 Mean distance estimations between the original and manipulated versions.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dist1</td>
<td>M 454.5 (SD 180.9)</td>
<td>M 641.7 (SD 406.1)</td>
<td>187.2</td>
<td>890</td>
<td>435.5* p&lt;.05</td>
<td>248.3</td>
</tr>
<tr>
<td>Dist2</td>
<td>M 32.9 (SD 13.9)</td>
<td>M 59.8 (SD 32.1)</td>
<td>26.9* p&lt;.05</td>
<td>94</td>
<td>61.1* p&lt;.01</td>
<td>34.2* p&lt;.05</td>
</tr>
<tr>
<td>Dist3</td>
<td>M 61.4 (SD 38.2)</td>
<td>M 114.6 (SD 95.7)</td>
<td>53.2</td>
<td>90</td>
<td>28.6* p&lt;.05</td>
<td>24.6</td>
</tr>
<tr>
<td>Dist4</td>
<td>M 38.4 (SD 12.8)</td>
<td>M 62.3 (SD 22.7)</td>
<td>23.9* p&lt;.01</td>
<td>94</td>
<td>55.6* p&lt;.01</td>
<td>31.7* p&lt;.01</td>
</tr>
</tbody>
</table>

Note:  
Dist1: distance traversed from starting point to first target (Hotel New York to Luxor).  
Dist2: Luxor Theatre to Toren op Zuid (estimate from position 1). The difference between the versions is significant, p<.01.  
Dist3: Luxor Theatre to location of participant at target 2.  
Dist4: Luxor Theatre to Toren op Zuid (estimate from position 2). The difference between the versions is significant, p<.05.  
In the last two columns the difference with the real distance is shown.

4.2.5 Small scale extra study

In version D of the model of the “Kop van Zuid”, used in the main experiment described above, two clusters of information were manipulated, the “cleanliness” of the area, and the weather conditions. After completing this experiment we were able to attract some extra participants and conducted a limited follow-up experiment to assess the effect of these clusters separately. Combined with the data of the previous experiments, we examined the effect of the variables “environmental decay” and “colour of the sky”, as shown in Table 4.2.

In version B only the sky was changed to grey, and in version C only the man-made nuisance objects and adjusted textures were included. The numbers of participants were 5 and 4 respectively; they were university students between 20 and 26, 6 male and 3 female.

Figure 4.8 a) Version B: with grey sky. b) Version C: with man-made nuisance objects
The old harbour area. The results indicate that version C, the combination of blue sky and added details was appreciated most, with significantly highest scores for the SMB-factors Pleasantness and Social Status. Table 4.7 shows the results for the versions B and C.

Table 4.7 Results for versions B (grey skies, no details) and C (blue skies, with details), old area.

<table>
<thead>
<tr>
<th>SMB-Factor</th>
<th>Version</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasantness</td>
<td>B</td>
<td>3.38</td>
<td>.91</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>4.44</td>
<td>.79</td>
</tr>
<tr>
<td>Complexity</td>
<td>B</td>
<td>3.55</td>
<td>1.63</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>3.63</td>
<td>1.16</td>
</tr>
<tr>
<td>Unity</td>
<td>B</td>
<td>5.50</td>
<td>.73</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>5.00</td>
<td>.86</td>
</tr>
<tr>
<td>Enclosedness</td>
<td>B</td>
<td>3.25</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>3.44</td>
<td>.97</td>
</tr>
<tr>
<td>Potency</td>
<td>B</td>
<td>4.95</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>5.06</td>
<td>.38</td>
</tr>
<tr>
<td>Social Status</td>
<td>B</td>
<td>3.20</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>4.31</td>
<td>.55</td>
</tr>
<tr>
<td>Affection</td>
<td>B</td>
<td>5.15</td>
<td>.89</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>4.63</td>
<td>1.05</td>
</tr>
<tr>
<td>Originality</td>
<td>B</td>
<td>3.3</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>4.25</td>
<td>1.65</td>
</tr>
</tbody>
</table>

A two way ANOVA (using as variables environmental decay and the colour of the sky) was performed for the SMB-factors for all four versions, thus including the data of the previous study. There is a main effect for the colour of the sky on Pleasantness \( F(1, 33) = 6.74, p=.015 \), and Social Status \( F(1, 33) = 13.85, p=.001 \). This indicates that the blue sky resulted in a higher perceived Pleasantness and Social Status of the area compared to the grey sky. No effects are found for the added nuisance objects and decay, or the interaction of these variables.

The new business district

The elements added in the new business district were, apart from the adjusted textures to give an impression of use and stains, neutral compared to the old area: large billboards, traffic signs, poles etc.

The results indicate that version C, the combination of blue sky and added details was appreciated most, with again significantly higher scores for the SMB-factors Pleasantness and Social Status.

A two way ANOVA was performed for the SMB-factors and single adjectives for all four versions. There is a main effect for the added elements on Potency \( F(1,33)=8.81, p=.006 \), Social Status \( F(1,33)=10.6, p=.003 \) and Originality \( F(1,33)=6.38, p=.017 \): added elements increase the scores on these factors of the SMB-scale. There is also a main effect for the colour of the sky on Pleasantness \( F(1, 33)=5.11, p=.031 \) and Social Status \( F(1, 33)=7.40, p=.011 \): a blue sky increases the scores on these SMB-factors. No interaction effects were found.

4.2.6 Conclusions and discussion

The manipulations performed were expected to influence the SMB-factors as they would in a real environment: the manipulations in the old harbour area were designed to create a disorderly effect, and were expected to have an overall negative effect on the appraisal. We expected lower scores on Pleasantness and Social Status, and higher scores on Unity, Complexity and Originality. The manipulations in the second area, the business district, did not include evidently negative elements and were expected to make the environment more complex and original, resulting in a lower score for Unity and higher scores for Complexity and Originality.
In direct comparison in a pilot study the visual difference between the versions was considered extensive, obvious, and convincing. However, the differences in response between the first two groups of participants of the experiment, who only navigated one of the versions (A or D), are small. For the factors Pleasantness and Originality they show a tendency as expected, but they are only significant on the factor Social Status in the old harbour area. For the new business district the results also show a tendency confirming our hypotheses (lower for Unity, higher for Complexity and Originality), but no significant differences were found on these factors. This may be due to the small number of participants, but it is probable that other factors play a role.

Firstly, the low reliability coefficient for several factors of the SMB-scale indicates problems with the scale. For the old area, the reliability of 7 factors of the scale was much lower for the manipulated version than the original version. Participants mentioned that they found some of the adjectives of the scale difficult to interpret. When answering the open questions in the experiment, the participants used terms which were very different, but suitable for characterizing the environment. It may be that the scale is less reliable, as a result of changes in language, which obviously have occurred in 30 years. Also, preferences may have changed. Participants were asked to judge an urban, industrial area. Such environments are now considered attractive and interesting, especially by young people, but were less so 30 years ago. This may have an effect on some of the adjectives of the scale, and on the internal consistency of the factors.

Secondly, the number and strength of the manipulations may have been too small to override the distinct style of the artist impression. The environment in the manipulated version was still considered dull, and desolate. The objects and details that were added were, however, sufficient in number to influence the perception of time and distances in the models.

Thirdly, we surmise that the effect of the representation on the appraisal of the represented environment is more complex than we hypothesized. Although the tasks were designed to make the participants feel involved with the environment, the participants seemed to confuse their judgment of the representation with that of the environment represented.

On one hand, they used words that describe not only the actual visual characteristics of the physical structures, but also the affordance of the area, and the affective qualities, such as empty, and lifeless. These are judgments on the urban environment that is represented in the model. On the other hand, they for instance showed interest in, and appreciation of, elements in the manipulated version such as bins, litter, and traces of vandalism, because for them these were new and attractive elements in 3D model, the representation. Cars deliberately parked on the pavement, which in reality would form a nuisance to the viewer, aroused attention, and formed no obstacle because navigating around the car took no effort. Participants expressed their interest in these represented elements by comments during and after the experiment. Their perception of time spent in the model supports this observation: the time possibly seemed to pass slower in the original version, because there was less to see and to draw the interest. The appreciation of these, inherently negative, elements may well have influenced the appraisal of the environment. The low reliability of some of the factors in the SMB-scale, especially in the manipulated version of the old area, may indicate this confusion. A similar effect was noticed by Park et al. (2010), who have tested perception of crime in a virtual environment. In a virtual environment deliberately manipulated to make viewers feel unsafe, some of the participants nevertheless chose to enter a hidden space with graffiti and a dumpster, which in reality is considered dangerous. In the virtual environment, the space intended to reflect danger also created curiosity, as in our model.

In our experiment, the appreciation of the attractive and unusual graphic characteristics may thus have influenced the appraisal of the environment itself. In this experiment most participants considered both versions (A and D) of the model convincing and realistic. Their verbal comments indicate that, although viewers were very much aware that they are not seeing a truthful depiction of an environment, they did not distinguish the representation from the environment represented, or assessed in what way they were influenced by the representation. This confirms remarks by...

From the second smaller study it may be concluded that the blue sky and bright colours were important determinants of the perceived pleasantness of the environment. Of all four versions, the environment with blue sky and added physical incivilities scored highest on Pleasantness and Social Status. The added details, intended to make the environment more dilapidated, did not diminish the pleasantness or social status of the area, but added some interest to the otherwise rather dull environment.

In reality a clear sky is known to influence emotions positively, and to increase the perceived pleasantness of a place (Cunningham, 1979; Knez, Thorsson, Eliasson, & Lindberg, 2009). Cunningham (1979) mentions several explanations for this effect, based on symbolic associations, aesthetics, and biological processes. Sunshine level can influence mood through its symbolic connection with pleasant or disappointing events, trips, or cancelling of plans and physical discomfort respectively. Sunshine can produce a more positive mood through aesthetic responses to the environment. The strong light illuminates the environment in a more stimulating and pleasing manner, enhancing colours and sharpening detail. The scenery on a cloudy day appears more dull and monochromatic.

The selection of the sets of independent variables in this experiment was motivated by observations of the use of visualisations of urban areas for purposes in which affective appraisal are important, such as the assessment of aesthetic quality and perceived safety. In a project of the Rotterdam municipality, described in Chapter 1, 3D models of the Rotterdam Central Station area were examined as a means to assess the public safety of an environment during the design process of large scale projects in public transport and civil engineering (Houtkamp, 2004). Although many factors can contribute to the subjective public safety, generally the appreciation of the buildings and their environment is considered to be an important determinant (Durmišević, 2002). The presence of litter and signs of dilapidation and vandalism, or physical incivilities, are indications of a lack of surveillance; the visual complexity of an urban environment with manmade nuisance objects can increase disorientation, and thus feelings of unsafety. Evidently, these elements influence the affective qualities of an area, which in turn influences other subjective experiences. Our experiment suggests that physical incivilities do not have the impact in a desktop 3D environment as expected, and that they do not per se improve validity of models for, for example, assessing perceived safety. At first sight, this result seems to deviate from recently conducted experiments by Park et al. (2010), who explored the use of virtual environments as research tools for the study of fear of crime. In their experiments, large displays (five by four meters) were used to project virtual environments of urban areas. They found that participants’ route choices in the virtual environment, were largely similar to reality: people avoided narrow alleys, physical incivilities, and threatening persons. However, the social incivilities (threatening persons) were more important cues than the physical incivilities. In our experiment no persons were displayed. Secondly, Park et al. (2010) feel that immersion is required to elicit these responses. Desktop displays are much less immersive than the large displays used in their experiment.

Experiments with larger numbers of participants are required to assess the effect of weather conditions and the level of detail in a virtual environment, on the perceived pleasantness of an area. Understanding of the effect is of importance when visualisations are used for assessment of aesthetic qualities of an area, or assessment of perceived public safety. In Chapter 4.4 the effect of weather conditions on visualisations will be addressed again.

Our experiment also shows that the characteristics of the medium, and style and conventions of modeling, influence the affective response to the environment represented. When 3D models are used to inform clients or citizens of architectural or urban designs, and of the affective quality of the future buildings, developers should be aware of this effect. If realistic models of high graphic quality are used, instead of, for instance, traditional artist’s impressions, user testing of the models to assess the impact of the visual effects on the appraisal of the environment represented, is essential.
The results of this experiment support the importance of discriminating between the environment in a 3D model and features of its representation, as proposed in the framework in Figure 3.3. Elements in a model may influence viewers’ response to a 3D model of an environment differently than expected. Viewers do not distinguish their judgments of the environment represented, and features of the representation, in a consistent manner.

**Acknowledgements**

We are grateful to VSTEP BV for providing the original models and to Matei Lunca for creating the adjusted models.
4.3 Experiment: The Influence of Lighting on the Affective Qualities of a Virtual Theater


4.3.1 Introduction

Lighting is recognized as a fundamental issue in 3D graphics (Slater et al., 2002). Challenges are to maintain a sufficient level of perceptual fidelity, while reducing the computational burden of models used for real time navigation. Illumination in a model therefore depends on choices made in the modeling process, which are generally restricted by budget, and not by assessment of the required quality or detail of the illumination. We wanted to assess to what extent quality of lighting conditions and shadows influence the affective appraisal of an indoor environment. The effects of lighting and shadows on the affective appraisal of desktop virtual environments are relevant for architects and visualisation developers, since they determine both the validity of the models, and the development costs of 3D models. They may also affect the validity of virtual environments for training purposes, where the environment provides the stage setting for actions and events, and must be convincing and convey the desired ambience.

We determined the affective qualities of a 3D model of the Royal Carré Theater for two different lighting conditions: a shadowless one with only ambient lighting, and a shaded representation with enhanced lighting. We also obtained appraisals from participants who visited the real theater. We expected the lighting effects to enhance the depth perception, the perceived quality of textures, and the information richness of the virtual theater. We also expected the enhanced lighting to increase the pleasantness, perceived social status and variety of the environment, and to make the environment appear more enclosing. Finally we hypothesized that the affective appraisal of the enhanced model would be more similar to that of the real theater, than the appraisal of the degraded model.

This experiment relates to the representation of the environment, and thus to column 2 of the framework presented in Chapter 3: Modifying factors and determinants.

4.3.2 Related Work

The interaction of light and shade provides a wide range of informational cues to the human visual system, and can also elicit a range of human responses such as the affective appraisal of an environment (Zimmons, 2004; Slater et al., 2002). In environmental psychology and architectural design, the effects of light and shadows on a varied range of human responses have been investigated. It was found that people use brightness and variability of luminance distributions for their affective appraisal of the lighting of environments (Flynn et al., 1973; Flynn, 1977; Flynn, 1988; Hendrick et al., 1977; Veitch, 2001). A uniformly lit environment appears more spacious (Flynn et al., 1973) and decreases feelings of fear in deserted places (Hanyu, 2000), whereas a non-uniformly lit environment appears more complex, thereby making it more interesting or pleasant (Hanyu, 2000; Newsham, Richardson, Blanchet, & Veitch, 2005; Protzman and Houser, 2005; Veitch, 2001). These principles are for instance deployed in the design of ambient lighting schemes to create a mood,
impression or visual interest in a given real-world environment (Flynn, 1977; Miwa & Hanyu, 2006). In computer games light and shadows are used to enhance mood, atmosphere and drama, and to evoke strong emotions such as suspense, dread, and comfort (Niedenthal, 2005). Lighting also affects the colour distribution in a scene, which in turn determines the perception and appraisal of architectural space. Although studies on the effect of colour are incomplete, warm and saturated colours are generally considered as more arousing (Franz, 2006). Brightness and saturation generally contribute to higher valence (Mehrabian & Russell, 1974), while luminance variations determine the dominance factor. Generally, cool, desaturated, and light colours are considered to enhance the experienced spaciousness, whereas dark, saturated, and warm colors have opposite effects (Franz, 2006).

![Figure 4.9 a) Corridor, Royal Carré Theater (reality). b) Hall, Royal Carré Theater (reality)](image)

### 4.3.3 Method

#### Participants

A group of participants, N= 33, 20 to 62 years old (mean 30.3), was randomly split into three. 12 participants traversed the degraded version of the model, and 12 participants the version with lighting. 9 participants visited the real theater.

#### 3D Models

We used two manipulated versions of a 3D interior model of the lobby, corridor and main theater hall of the late nineteenth century Royal Carré Theater in Amsterdam, the Netherlands. The interior decoration of Carré shows many characteristics of the baroque style, such as the extensive use of the colour red, typical decorations, and chandeliers (Figure 4.9). The model was created by VSTEP B.V. for a media presentation by Kopla Media B.V., to showcase the renovation plans for fundraising purposes, before the actual renovation work began in 2004. It was made in Quest3D and 3D Studio Max, on 2D CAD-drawings of the building. The model did not include sounds. We created two versions of the theater model (Figure 4.10, Figure 4.11). In one version all textures were given the same emissive value and all the shading was removed, resulting in a degraded quality of lighting, and a bland environment. The second version featured pre-computed lightmaps for the shadows and additional real-time light sources for extra saturation and diffuse glow. The lighting was kept subtle since the goal was to mimic the real environment, and not to surpass it in experienced intensity. However, the shadows were purposely made noticeable. The lightmaps were created with 3D Studio Max v7’s native engine, analogous to what VSTEP used at that time. Although this is not the cutting-edge of contemporary lighting technology, for our research a technique which is in use in the architecture visualisation industry is more relevant. The architect who supervised the renovations of the theater, from Greiner, van Goor, Huijten architects, confirmed that the enhanced model more closely simulated the festive effect brought about by theater’s illumination, and created an impression of excitement and expectancy.
Figure 4.10 Corridor, a) degraded version, b) enhanced version

Figure 4.11 View intro theatre hal: a) from side entrance, degraded version; b) from side entrance, enhanced version; c) from stage, degraded version; d) from stage, enhanced version
Two of our manipulations may have influenced the outcome. First, since the seats comprise a large number of polygons they could not be lightmapped. Therefore we added real-time omni lights at the location of the chandeliers, resulting in sufficiently realistic lighting on the seats. Second, the model uses a semi-transparent bitmap to create a hazy halo effect on and around the chandeliers. While this is arguably lighting information, removing the bitmap altered the appearance of the chandeliers, i.e. they appeared less rounded and showed more detail. Therefore we chose to retain the haloes in both versions, but kept them subdued in the degraded version.

**Equipment**

The experiment was conducted on a desktop Pentium 4 computer, running at 2.4 GHz, 512 MB RAM, supporting a GeForce 6600GT, and equipped with a 17” CRT screen. The computer was located in a room with artificial lighting.

**Procedure**

**Assessment of the models**

After a short exercise to practice navigating the 3D model, participants were positioned in the lobby of the theater, with instructions to enter the corridor, and pass through the door into the theater hall. From there they were transported to a seat on one of the higher rows of chairs. Participants assessed the affective qualities of the environment twice: in the corridor and in the theater hall. For this, we used the SMB-scale (Semantic Environmental Scale). It comprises 36 adjectives in eight factors that describe the experience of persons in a man-made environment (Küller, 1977; Küller, 1991). The scale is more fully described in Chapter 3.5. The adjectives were rated on a 7-point unipolar scale. In addition, questions were presented to elicit verbal reactions to the subjective impression of the spaces, the quality of the representation, and the perceived dimensions of the spaces traversed. Besides determining the effect of the lighting condition on the affective qualities of the theater and on distance perception, we wanted to assess participants’ awareness of the differences in lighting between the two versions of the model. Therefore, after completing their assessment of one of the models, we asked the participants to briefly navigate the same path in the other version of the model, and to answer questions on the perceived difference between the two conditions.

**Assessment of the real theater**

A similar procedure was used to assess the real theater. Participants were led to the lobby of the theater and received instructions on how to proceed. Questions about the representation had been removed from the questionnaire. In this condition several additional factors may have influenced the assessment of the theater, such as the presence of the experimenter and maintenance personnel, and the lighting conditions in the theater that were not controlled by the experimenter. We can therefore only draw preliminary conclusions on the correspondence between the experience of the real theater and the modeled theater.

**Hypotheses**

We expected the lighting effects to enhance the depth perception, the perceived quality of textures, and the information richness of the virtual theater. We expected to find differences in several factors of the SMB-scale: the variation and higher quality of the lighting effects was expected to result in higher scores on Pleasantness and Social Status of the enhanced model; the shadows and lighting effects make the environment more varied and would therefore result in higher scores on Complexity; and because uniform lighting increases spaciousness, the variation in lighting was expected to result in higher scores on Enclosedness.

We hypothesized that the affective appraisal of the enhanced model would be more similar to that of the real theater, than the appraisal of the degraded model.
4.3.4 Results

**Affective appraisal of the models measured by the SMB-scale**

The dramatic lighting effects inserted into the model show a significant effect only on the SMB-factors of Enclosedness and Originality in the corridor, and on none of the factors in the theater hall. In the theater hall, the enhanced model scores higher on Complexity, but not on a conventional level (p< 0.06). In the corridor, the enhanced model does not score higher on Complexity, and even lower on Social Status, which is contrary to our hypothesis. A problem that we also noted in earlier experiments (Houtkamp & Van Oostendorp, 2007), is that the reliability of some of the SMB-factors is low (Cronbach’s alpha <.6). In several factors we removed one item to improve the reliability. The results, including the adjusted factors, are shown in Table 4.8, Table 4.9, Figure 4.12 and Figure 4.13.

<table>
<thead>
<tr>
<th>SMB-Factor</th>
<th>Version</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>t</th>
<th>p-value (one-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasantness</td>
<td>degraded</td>
<td>4.16</td>
<td>.67</td>
<td>-.32</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.26</td>
<td>.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>degraded</td>
<td>4.21</td>
<td>.89</td>
<td>.24</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.13</td>
<td>.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity - adjusted</td>
<td>degraded</td>
<td>4.42</td>
<td>1.15</td>
<td>.65</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.17</td>
<td>.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unity</td>
<td>degraded</td>
<td>4.33</td>
<td>.85</td>
<td>-.85</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.60</td>
<td>.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unity - adjusted</td>
<td>degraded</td>
<td>4.36</td>
<td>.77</td>
<td>-1.4</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.83</td>
<td>.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enclosedness</td>
<td>degraded</td>
<td>3.67</td>
<td>.83</td>
<td>-1.75</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.33</td>
<td>1.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potency</td>
<td>degraded</td>
<td>4.54</td>
<td>.81</td>
<td>.13</td>
<td>.45</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.50</td>
<td>.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potency - adjusted</td>
<td>degraded</td>
<td>4.00</td>
<td>1.05</td>
<td>.21</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>3.92</td>
<td>.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Status</td>
<td>degraded</td>
<td>4.48</td>
<td>.51</td>
<td>1.52</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.02</td>
<td>.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affection</td>
<td>degraded</td>
<td>3.21</td>
<td>.64</td>
<td>-.69</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>3.46</td>
<td>1.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Originality</td>
<td>degraded</td>
<td>4.81</td>
<td>.43</td>
<td>3.30</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>3.81</td>
<td>.96</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.9 SMB scores for the hall of Carré (N = 17).

<table>
<thead>
<tr>
<th>SMB-Factor</th>
<th>Version</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>t</th>
<th>p-value (one-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasantness</td>
<td>degraded</td>
<td>4.80</td>
<td>.52</td>
<td>.74</td>
<td>.23</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.61</td>
<td>.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>degraded</td>
<td>3.56</td>
<td>.91</td>
<td>-1.67</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.21</td>
<td>.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unity</td>
<td>degraded</td>
<td>4.85</td>
<td>.73</td>
<td>- .15</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.90</td>
<td>.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enclosedness</td>
<td>degraded</td>
<td>3.54</td>
<td>1.18</td>
<td>- .21</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>3.63</td>
<td>.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enclosedness -</td>
<td>degraded</td>
<td>3.11</td>
<td>1.25</td>
<td>- .19</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>3.19</td>
<td>.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potency</td>
<td>degraded</td>
<td>4.65</td>
<td>.71</td>
<td>- .33</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.73</td>
<td>.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potency - adjusted</td>
<td>degraded</td>
<td>4.14</td>
<td>.93</td>
<td>.25</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.06</td>
<td>.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Status</td>
<td>degraded</td>
<td>4.56</td>
<td>.54</td>
<td>- .16</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.60</td>
<td>.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Status -</td>
<td>degraded</td>
<td>5.28</td>
<td>.71</td>
<td>.74</td>
<td>.23</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>5.06</td>
<td>.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affection</td>
<td>degraded</td>
<td>3.10</td>
<td>.83</td>
<td>- .83</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>3.42</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Originality</td>
<td>degraded</td>
<td>4.54</td>
<td>.96</td>
<td>.17</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>4.48</td>
<td>.81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.12 Mean scores on the SMB-scale for the corridor of Carré (N = 17). The lines between the measuring points are only added to facilitate comparison of the results.
Figure 4.13 Mean scores on the SMB-scale for the hall of Carré (N = 17). The lines between the measuring points are only added to facilitate comparison of the results

**Comparison with the real theater**

The results of the experiment in the real theater must be interpreted with care, because of the small number of participants (n=9), and the circumstances mentioned above.

An ANOVA’s analysis and post-hoc analysis showed a significant difference (p<.05) in the appraisal of the corridor, between the degraded model and reality on the factor Pleasantness; between the degraded model and the enhanced model, and between the degraded model and reality, on the factor Originality. A trend was noted in the appraisal of the hall, between the enhanced model and reality on the factors Pleasantness (p=.055) and Originality (p=.084). In every case, the real theater scored higher, so participants in general found the real theater more pleasant and more original than the model.

**Materials, decorations, and atmosphere**

The answers to the questions on the subjective impression of the corridor and hall, and the quality of the representation, varied much. They only showed a few marked differences between the two versions of the model, mainly in the identification of materials. In the degraded version, most participants interpreted the material of the floor as marble, stone or concrete. In the enhanced version, most interpreted the surface as lino or a similar synthetic material; some also mentioned carpet and wood. Materials mentioned for the walls, are concrete, stone, wood, metal, wallpaper, stuc, paint. Appreciation of the quality of the models in general was high: for the degraded version mean 7.71, for the enhanced version 7.88, on a scale from 1 – 10.

For both model versions, and the real theater, the chandeliers, the illumination, the predominance of the colour red, and the decorations were reported as especially characteristic. In all three conditions, participants described their affective response as festive, anticipative, impressed, and excited. About half of the participants referred to a feeling of recognition in their answers, to this or a similar theater, induced by the model.

After assessment of one version of the model, participants were presented with the other version. Many noticed differences in the representation as well as atmosphere (varying from 71 – 79% of the participants) but were not able to indicate in which respect both versions differed, until this was pointed out to them after the experiment. In the verbal comments, the enhanced version is
appreciated slightly more than the degraded model, and elicited statements such as “more realistic”, “felt warmer”, and was “nicer to visit”. On the other hand, some of the participants appreciated the colours, and spaciousness, of the degraded version more.

**Spatial Dimensions**

Participants were asked to estimate height, width and length of the corridor and the theater hall. Many of them remarked they found it very difficult to estimate spatial dimensions in reality and expected to perform badly in virtual environments as well. The answers generally showed a high variation, with some outliers. ANOVA’s over the three conditions did not show significant differences. Estimations of the dimensions of both the corridor (except the height) and the theater hall were higher in the enhanced version. The judgments of the view from the upper balcony in the enhanced model were more negative, stating it was rather high and far from the stage. This indicates that the perception of depth and distances is influenced by the manipulations in the models. However, the estimated distances in the enhanced version were not closer to the actual dimensions. Table 4.10 lists the results of the participants’ estimates.

**Table 4.10** Estimates of spatial dimensions in meters in the degraded and enhanced versions of the models, and in the real theater.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Condition</th>
<th>N</th>
<th>Mean (meters)</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>length of corridor</td>
<td>degraded</td>
<td>12</td>
<td>36.42</td>
<td>23.22</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>12</td>
<td>43.17</td>
<td>51.18</td>
</tr>
<tr>
<td></td>
<td>reality</td>
<td>9</td>
<td>27.56</td>
<td>10.49</td>
</tr>
<tr>
<td>width of corridor</td>
<td>degraded</td>
<td>12</td>
<td>3.13</td>
<td>.86</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>12</td>
<td>3.17</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>reality</td>
<td>8</td>
<td>3.31</td>
<td>.53</td>
</tr>
<tr>
<td>height of corridor</td>
<td>degraded</td>
<td>12</td>
<td>4.11</td>
<td>1.68</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>12</td>
<td>3.54</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>reality</td>
<td>9</td>
<td>3.5</td>
<td>.56</td>
</tr>
<tr>
<td>length of theater hall</td>
<td>degraded</td>
<td>12</td>
<td>23.42</td>
<td>9.30</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>12</td>
<td>28.75</td>
<td>16.53</td>
</tr>
<tr>
<td></td>
<td>reality</td>
<td>9</td>
<td>18.17</td>
<td>18.17</td>
</tr>
<tr>
<td>width of theater hall</td>
<td>degraded</td>
<td>12</td>
<td>41.25</td>
<td>17.47</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>12</td>
<td>52.50</td>
<td>32.30</td>
</tr>
<tr>
<td></td>
<td>reality</td>
<td>9</td>
<td>37.06</td>
<td>20.54</td>
</tr>
<tr>
<td>height of theater hall</td>
<td>degraded</td>
<td>12</td>
<td>23.33</td>
<td>10.30</td>
</tr>
<tr>
<td></td>
<td>enhanced</td>
<td>12</td>
<td>26.17</td>
<td>8.90</td>
</tr>
<tr>
<td></td>
<td>reality</td>
<td>9</td>
<td>26.78</td>
<td>10.13</td>
</tr>
</tbody>
</table>

**4.3.5 Conclusions and Discussion**

The lighting conditions in our experiment only weakly determined the affective qualities of the 3D model. However, low Cronbach’s alpha’s for some of the factors indicate that the SMB-scale is perhaps not the optimal instrument for this research. The enhanced version appeared to give a more vivid sense of depth, but shadows and lighting did not improve estimates of the spatial factors.

The results of this study suggest that although viewers are susceptible to differences in lighting conditions when comparing scenes, they do not need specific lighting information when inferring affective qualities and dimensions of an environment. The models used in this study contain so many typical elements of a baroque style theater that they may invoke impressions of earlier visits to, or descriptions and images of, a similar theater. Consequently, they may induce subjective responses that are based on these mental representations and not only on the actual represented environment (Danford & Willems, 1975). Theaters worldwide have certain characteristics that help imagine a “prototypical” theater. Some evidence for this is found in the SMB-scores of the concert hall of the Sydney Opera House, evaluated by 20 viewers of colour slides (Küller, 1971). They resemble the scores of the hall of the Carré Theater, indicating a similar “ambience” (Figure 4.14).
Figure 4.14 Mean scores on the SMB-scale of the concert hall of the Sydney Opera House (Küller, 1971).

In this experiment, we noticed how easily participants interpreted the reduced information on the environment, provided in the model. The textures of the materials of the walls and floors were simple, but participants were in general convinced they identified the materials correctly. Their interpretations however differed widely. In the models of the urban area described in Chapter 4.2, the time-of-day and season were also easily determined, even though the models offered few cues.

Returning to the framework for our experiments (Figure 3.2), it seems that illumination is one of the features contributing to the ambience of a virtual environment, but not a predominant one. This is unexpected, regarding the attention given to illumination techniques by developers. Our participants however had little experience with 3D models, and it is possible that gamers or others who are experienced in computer graphics would react differently to the 3D model.

Further research is required to fully establish the effects of lighting and pre-existing mental representations on the affective appraisal of 3D environments, and to develop appropriate instruments to measure these effects. A series of experiments may for example compare the effect of lighting on the affective appraisal of models of directly recognizable and familiar buildings and environments such as theaters, churches, and campuses, with the effect when models of new and original types of environments are offered to the viewer. A new semantic differential scale has recently been developed by Vogels (2008) especially to investigate the influence of light on mood and perceived atmosphere of an environment, which may reveal more subtle differences in the appraisals.

Acknowledgements

The original model was created by VSTEP BV for Kopla Media BV. We are grateful to VSTEP BV for providing the models and to Greiner, van Goor, Huijten Architects for their advice, and for help in organizing the experiments in the Royal Carré Theater.
4.4 Experiment: Comparing affective appraisals of real and virtual urban environments


4.4.1 Introduction

The graphics quality of games and simulations has increased rapidly over the last years, and has raised the expectations of users of visualisations of urban planning projects. In the development of 3D models for visualisation, an underlying assumption is generally that the highest possible level of photorealism, which means realistic and dynamic lighting, perfectly anti-aliased graphics, and movie-quality motion would ensure a valid representation; that is, it would produce a cognitive, affective and behavioural response in the observer equivalent to the response produced by the real environment. However, for communication of design plans, city councils and other local governments require fast development of models. They also expect that the models are made available online, which puts severe restraints on the size of the files. Low-end (affordable) models are therefore developed to quickly visualise proposals, with an “acceptable” level of realism, accuracy and abstraction. Hence, the question of the validity of models for assessment of urban development plans is highly relevant (Wergles & Mugar, 2009). According to Pietsch (2000), the attention of researchers has focused on achieving highly realistic models and the potential of simpler forms of representation are not receiving the attention they deserve.

In this empirical field study we compared the affective appraisals of a series of models designed for online assessment and communication of urban development, with those of the urban environment in reality. The aim is to examine the comprehensive effects of modeling and of representation on a desktop system, on the affective appraisal of an urban environment. In the framework presented in Chapter 3 these determinants are included in the category Representational effects and modifiers.

The 3D model contains fewer details than the real environment, is cleaner, more static, and shows fewer differences in styles and materials. We expected that viewers would judge the modeled environment as more visually coherent, more pleasant, more modern, and of a higher social status.

4.4.2 Related Work

The validity of visualisations as representations of landscapes and urban environments has been topic of research since over 40 years (e.g. Lange, 2001; Bishop & Lange, 2005). Most empirical studies either compared the response to real environments with different types of static visualisations, such as photographs, slides, and screenshots of 3D models, or compared different types of visualisations, such as different levels of detail, or presentation modes (Huang, 2009). Only few studies have compared viewers’ responses between a real (urban) environment and a 3D model. Most often cited is the study by Bishop and Rohrmann (2003), who have used an animated (non-interactive) walk through a modelled urban park to compare perceptual responses to the model and the real environment. They concluded that (for that time) detailed and time-consuming computer simulations did not necessarily generate the same responses as the corresponding real environment,
although differences between day and night conditions were mostly the same in the simulated as in the real environment.

Wergles and Muhar (2009) also conducted an experiment comparing responses to a real environment and a visualisation of the same area in Vienna. They adopted a different research technique, and applied qualitative methods for eliciting viewers’ responses. They raise an objection to previous studies, which are restricted to evaluating preference or other visual perceptual responses to the environment and its representation. Their argument is that similar ratings for an environment or its surrogate do not necessarily mean that the motivation for the assessment is the same. Different determinants may result in similar responses, especially if the measurement scale used is little differentiated, for instance when preference is measured. In their study, Wergles and Muhar therefore focus on revealing which elements in the environment or model determine the observers’ evaluation; “whether it is indeed the same that they like or dislike in the depicted and the real environment”. In their study, participants either walked through a renovated urban area in Vienna, or were shown a series of (static) visualisations of the project proposal. Qualitative questionnaires were used to elicit responses, which were content analyzed for similarities and differences in their cognitive, affective and evaluative aspects. They found for instance that participants commented on the quality of the visualisations, although this was not asked for; and that participants who walked through the real environment gave more extensive answers than those who saw the visualisation.

In the field experiment discussed here, we compared the affective appraisals of an interactive visualisation of an area in Venlo, with that of the area in reality. Despite the objections mentioned above, a scale was used to quantitatively assess the affective appraisals of the environment. The SMB-scale by Küller (1975), used in the experiments described in Chapter 4.2, 4.3 and 4.4, contains many adjectives that relate directly to characteristics of the environment, so should reveal more of the motives for the appraisals than for instance the scale of Russell and Pratt (1980) used in other experiments (Chapter 4.5, 4.8, 4.9). Open questions were added to the questionnaire for this purpose as well.

A well known series of online models for communication and assessment of (future) urban environments in the Netherlands, are the multi-user 3D environments using the VirtuoCityEngine developed by Cebra. Several cities have projects online, for instance Helmond, Apeldoorn, Tilburg, and Venlo. VirtuoCity is based on central client-server architecture, with a client of limited size, requiring only a small download. In a VirtuoCity environment users navigate by using their mouse and/or keyboard as input device. The controls are simple, to allow users with little experience in 3D applications to use the application without instruction.

The 3D models in the example used in this study are created for VirtuoCity by engineering agency Movares Nederland BV, based on a large-scale base map of the Netherlands in combination with aerial photographs. Both are used to define the project outlines and boundaries. The base plate, containing basic information such as curbs, housing and, roads, and the models placed on it, are created in 3DStudioMax. Facades are photographed with as few visual disturbances as possible, as are other surrounding features such as lampposts, traffic signs, street furniture etc. CycloMedia 360° panoramic pictures are used as textures where possible. Although the geometric level of detail is low, the photographic textures provide a detailed impression. A small number of parked cars and persons (avatars) are included in the environment.

In this experiment, we compared the affective appraisals of a model of this type covering an area in Venlo, with the affective appraisals of the area in reality.

4.4.3 Hypotheses

The visualisation of Venlo is a typical example of a model developed for online interactivity and within a constrained budget. It presents a more coherent, and comprehensible view of the area as a result of the simplification of geometric shapes of the buildings, the photo textures applied on the facades, the similarity of the textures for the other surfaces (pavement, road etc.), and the reduction
of elements in the street, such as people, traffic and street furniture. The absence of people and sounds makes the environment less lively, and more static. We therefore hypothesize that (1) the appraisals of the 3D model show higher scores for the SMB-factor Unity, and lower for Complexity.

Negative elements like dirt, litter and indications of dilapidation and vandalism are not modeled. We therefore also expect (2) a higher score on the factor Pleasantness and Social Status. Because the look-and-feel of the representation is quite modern, we expect (3) a higher score for the factor Affection (an indication of how modern the area is perceived).

Prior knowledge of an environment and associated appraisals may interfere with judgments of this environment, or of the representation of the environment, in an experimental setup. In Chapter 2.5, the role of mental representations in appraisal was explained. In the design of this study, participants with prior knowledge of Venlo were equally divided over the conditions. No hypothesis was composed on the influence of prior knowledge on the appraisal of the environment. We assumed prior knowledge would affect the assessment of the visualisation, but on which factors, and in which direction, is determined by individual appraisals of the area in reality, which were unknown before the experiment.

4.4.4 Method

Participants

32 persons participated in the experiment, 20 male and 12 female. The average age was 32.5 years (minimum 18, maximum 59, SD= 11.8). All participants but three were students in higher education institutions or were local government officials, with at least a bachelor degree. Half of them had prior knowledge of Venlo. Participants from both groups were randomly assigned over the two conditions.

Materials

Model and selected area

An area of Venlo was selected that met the requirements for a tour in real life as well as in the model, for instance, not subject to changes or building activities at the time of experiments.

The area is known as the “Keulse Poort”, a gateway to the centre of Venlo, which comprises a number of landmark buildings like the modern “Limburgs Museum”, a monumental filling station and a large post-office, both about 70 years old (Figure 4.15). Visitors are attracted by shops, cafes, and small companies. The route to be traversed by participants was chosen to show a diversity of views available in the model.
Figure 4.15 Scenes in real Venlo and Virtual Venlo: a. the post office, b. the filling station and c. the Limburg Museum. Left: the virtual environment. Right: the environment in reality

The 3D model was installed at two locations, in similar settings, and on similar computers, with 19” flat screen monitors. Participants who made the walk through real Venlo, were welcomed in a separate room in the Limburg museum.

Procedure, questionnaires and SMB-scale

Participants in both conditions first received a short instruction and a questionnaire with general questions on education, computer experience, etc. All participants were then provided with a map (Figure 4.16) of the area on which the required route was indicated. At each of the 5 numbered locations the participant was asked to estimate a distance to or between buildings, or height or width of a building. Also the traversed distance from starting point to finish, was asked. Visual scale and openness are important physical attributes of an environment, and influence its affective qualities (Chapter 3). It is therefore important to determine whether deviations in estimations of dimensions and distances influence the affective qualities of an environment.

After the (virtual) walk a set of questionnaires was presented, adapted to the conditions. To assess the affective appraisal of Venlo and Virtual Venlo the SMB-scale (see Chapter 3.5) was used. The SMB-scale does not contain items that reflect the use of an area, and the ambience as a result of
presence of people. Three adjectives were added for this purpose (busy, quiet, cosy/lively). Participants were then asked what they found remarkable in the environment, such as certain buildings or the ambience. They were asked to rate on a visual analogue scale (a continuous line between two end points) whether they felt the colours were ugly or beautiful, and somber or cheerful.

For real Venlo, participants were asked about the sounds they heard during their walk, and the weather conditions. For Virtual Venlo, participants were asked if they had missed sounds, the time and season they thought the model represented, and whether they experienced problems with navigation.

![Figure 4.16 Map with route used by participants in both conditions](image)

### 4.4.5 Results

**SMB-scale**

First the internal consistency of the 8 factors of the SMB-scale was checked. Cronbach’s alpha for the factor Complexity was improved to \( \alpha = .53 \) by removing the item “subdued”; for the factor Affection to \( \alpha = .77 \), the adjective removed was “timeless”. The reliability for the factor Potency and Social Status was adjusted by removing one single item (fragile, resp. well-kept), but Cronbach’s alpha remained too low (resp. \( \alpha = .38 \) and \( \alpha = .4 \)). Results should be considered with some care.

A two-way ANOVA with the factors environment (virtual vs. real) and prior knowledge (high vs. low) resulted in a significant main effect of environment, for Pleasantness, Unity and Affection, that received higher scores in the virtual condition. The factor Affection is used to indicate how modern or aged an environment is judged; a higher score means more modern. Results for the versions are listed in Table 4.11, and shown in a graph in Figure 4.17.
Table 4.11 Scores on the SMB-factors.

<table>
<thead>
<tr>
<th>Virtual Venlo</th>
<th>Mean</th>
<th>SD</th>
<th>Real Venlo</th>
<th>Mean</th>
<th>SD</th>
<th>ANOVA</th>
<th>Main effect for environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasantness</td>
<td>4.59</td>
<td>.56</td>
<td>3.30</td>
<td>.74</td>
<td>31.27</td>
<td>&lt;0.01*</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>3.98</td>
<td>1.13</td>
<td>3.50</td>
<td>.82</td>
<td>1.86</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>Unity</td>
<td>4.63</td>
<td>.81</td>
<td>3.19</td>
<td>.83</td>
<td>24.16</td>
<td>&lt;0.01*</td>
<td></td>
</tr>
<tr>
<td>Enclosedness</td>
<td>3.47</td>
<td>1.06</td>
<td>3.56</td>
<td>.75</td>
<td>.09</td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td>Potency</td>
<td>4.15</td>
<td>.83</td>
<td>4.58</td>
<td>.59</td>
<td>3.04</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>Social Status</td>
<td>3.54</td>
<td>.71</td>
<td>3.75</td>
<td>.91</td>
<td>.59</td>
<td>.45</td>
<td></td>
</tr>
<tr>
<td>Affection</td>
<td>4.46</td>
<td>1.02</td>
<td>2.80</td>
<td>1.05</td>
<td>19.34</td>
<td>&lt;0.01*</td>
<td></td>
</tr>
<tr>
<td>Originality</td>
<td>3.72</td>
<td>.84</td>
<td>3.05</td>
<td>1.05</td>
<td>4.02</td>
<td>0.055</td>
<td></td>
</tr>
</tbody>
</table>

Note: A significant main effect was found for Pleasantness, Unity and Affection

No main effect of prior knowledge was found. An interaction effect was found for factors Enclosedness (F (1, 28) = 6.12, p=0.020) and Social Status (F (1, 28) = 4.69, p=0.039):

- Participants without prior knowledge judged the real environment as more enclosed than the virtual environment, whereas the participants who were familiar with Venlo, judged the virtual environment as more enclosed (so less open).
- Participants without prior knowledge gave higher scores on Social Status to the virtual environment than to the real environment; for the participants who were familiar with Venlo this was opposite.

![Graph with the mean scores on the SMB-factors for Virtual Venlo and real Venlo. The lines between the measuring points are only added to facilitate comparison of the results.](image-url)
**Other appraisals of the environment**

The virtual environment was considered significantly more quiet and less busy than the real environment, as expected. More interestingly, the virtual area was considered more “cosy” than the real environment, despite the absence of people.

*Table 4.12 Scores for items quiet, busy and cosy, scale 1-7*

<table>
<thead>
<tr>
<th></th>
<th>Virtual environment</th>
<th>Real environment</th>
<th>Main effect for environment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Quiet</td>
<td>4.25</td>
<td>1.53</td>
<td>2.69</td>
</tr>
<tr>
<td>Busy</td>
<td>3.19</td>
<td>1.33</td>
<td>4.56</td>
</tr>
<tr>
<td>Cosy</td>
<td>4.63</td>
<td>1.31</td>
<td>3.19</td>
</tr>
</tbody>
</table>

On the item “busy” a significant main effect was found for prior knowledge (F (1, 28) = 4.24, p= .05), and a significant interaction effect (F1, 28) = 5.77, p=0.023). Participants with prior knowledge judged the virtual environment as busier than the other groups.

Participants considered the colours in the virtual environment more beautiful and lively than those in the real environment (Table 4.13).

*Table 4.13 Scores for items ugly-beautiful colours and gloomy-lively colours*

<table>
<thead>
<tr>
<th>Colours</th>
<th>Virtual environment</th>
<th>Real environment</th>
<th>Main effect for environment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>ugly-beautiful</td>
<td>5.38</td>
<td>1.81</td>
<td>2.73</td>
</tr>
<tr>
<td>gloomy-lively</td>
<td>4.01</td>
<td>1.79</td>
<td>2.36</td>
</tr>
</tbody>
</table>

*Note: Visual Analogue Scale 1-10.*

Also a significant main effect of prior knowledge was found: participants with prior knowledge thought the colours were more beautiful, than other participants (mean 4.54, SD 1.89, versus 3.57, SD 1.89, F (1, 28) = 4.16, p=0.05.)

**Distances**

Estimates of dimensions of buildings (height and width) were very similar in both conditions, and close to the real dimensions. Heights of two buildings were asked. For Virtual Venlo they were estimated at resp. 18.63 m. (SD 2.03) and 12.50 m. (SD 2.61); for the real environment 19.13 m. (SD 7.37) and 14.53 m. (SD 5.97). The width of a building was estimated at 35.69 m. (SD 11.11) for Virtual Venlo, and 40.50 m. (SD 6.00) for real Venlo. In reality the width is 36 m. As in the experiments described in Chapter 4.2 and 4.3, people were able to judge (or calculate) dimensions using objects of familiar sizes as reference.

Estimates of egocentric distances and the traversed distance in both conditions were higher than the actual distances. In the virtual environment, they were overestimated by more than 50%. In this and other experiments we conducted, participants often mention they find it very hard to accurately estimate distances, both in reality, as in virtual environments. As a result, the range in answers is large. Table 4.14 lists the results of the estimations of egocentric distances and traversed distance in both conditions. The results do not provide an explanation for the interaction effect for the factor Enclosedness of the SMB-scale, since no similar interaction effect was found in the distance estimations.
Other results were positive; pleasantness was 4.4.6 Immersion

The participants in the virtual environment were asked whether they had felt immersed, felt they were walking through real Venlo, and were unaware of their surroundings, on a 7 point bipolar scale (from “do not agree at all” to “agree completely”). The first two questions were answered positively; mean 5.44, SD .63, mean 4.94, SD 1.18. The score for the last question was mean 3.38, SD 1.36. A score of 4 was neutral. From this we conclude that the environment was convincing at a satisfactory level, and that the quality of the representation and interaction was sufficient.

Other items, and open questions

Open questions were added to elicit response from participants on elements of the environment that they had noticed, and to find reasons for their appraisals. The number of participants was too small to perform a full content analysis but the unambiguous and relevant results are mentioned here.

When prompted, 12 of the 16 participants in the virtual environment answered that they had missed sounds, especially common sounds such as traffic, people talking, or other ambient sounds, to make the environment “more realistic”. The sounds in real Venlo were considered as normal to pleasant by the participants who walked through real Venlo. To the question what was remarkable in the environment, 8 participants answered with negative comments on the impression of the real area, such as unpleasant, old-fashioned, incoherent, and unclean. The virtual environment did not receive any negative comments, only positive or neutral. To the question what could be improved in the virtual environment, 10 participants answered more liveliness, people or movement in the streets.

The experiments were conducted in spring, during dry sunny to cloudy weather. The virtual environment shows blue skies with light clouds, and green leaved trees; 10 participants concluded it was spring. As we noticed before in experiments (Chapter 4.2), participants very easily draw conclusions on the time of day in the virtual environment: 11 participants thought it was afternoon, although the model does not contain evidence for this.

4.4.6 Conclusions and discussion

The aim of this experiment was to examine the comprehensive effects of modeling, and representation on a desktop system, on the affective appraisals of an urban area in Venlo. The area of Venlo represented in the 3D model, was considered more pleasant, more uniform, and more modern than the area in reality, in agreement with our hypotheses. The higher scores on the factors Pleasantness and Affect are probably determined by the cleanliness of the area, the absence of negative elements like litter and vandalism, and the colours, which were considered prettier and

<table>
<thead>
<tr>
<th></th>
<th>Virtual environment</th>
<th>Real environment</th>
<th>Actual Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>p</td>
<td>Mean</td>
</tr>
<tr>
<td>Egocentric distances</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance 1</td>
<td>124.69</td>
<td>77.64</td>
<td>.022</td>
</tr>
<tr>
<td>Distance 2</td>
<td>95.31</td>
<td>60.76</td>
<td>.064</td>
</tr>
<tr>
<td>Distance 3</td>
<td>175.63</td>
<td>104.43</td>
<td>.024</td>
</tr>
<tr>
<td>Traversed distance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance 4</td>
<td>1037.50</td>
<td>673.18</td>
<td>.020</td>
</tr>
</tbody>
</table>

Note: The traversed distance is measured from the starting point of the route to the finish. The estimated distances in the virtual environment and those in the real environment are compared with the actual distances. Whether this difference is significant is shown by the p-value.
livelier in the model. The environment was considered more uniform (apparent in the factor Unity), which can be ascribed to the photo textures. These conceal differences in materials and relief, and offer similar colour hues. Egocentric distances were overestimated in the virtual environment, providing an open view of the environment.

The factor Complexity, that reflects the perceived vivaciousness and colourfulness of an area, was not significantly different between the two groups, contrary to our hypothesis (p>.05). Although the colours in the model were considered lively, the area is emptier and contains less movement and visual variation, and no sounds, which may have counteracted the effect. Here a problem occurs that is noted by Wergles and Muhar (2009), and discussed in the introduction of this chapter: although the appraisal is similar, the reasons may be very different.

An aspect often mentioned by viewers of 3D models, for instance in citizen participation projects, is the (too) quiet, static or lifeless ambience of 3D models, because of the absence of traffic and people, and dynamic elements in general. Also in this experiment a significant difference was noted between the conditions in the perceived quietness of the area. Nevertheless the participants considered the real environment more “cosy” (pleasantly lively) than the virtual environment. An obvious difference between the environments is the terrace with chairs and tables on the otherwise rather empty square near the end of the walk, only in the virtual environment, suggesting the use of the area for social activities. Again the colours, that in the real environment are mostly grey and brown, may have given the area a livelier ambience. The absence of people and movement may thus, to some extent, be compensated by lively colours and indications of pleasant activities.

To our knowledge, no attempts have been made to assess the affective appraisals of the other VirtuoCity models. They all share similar characteristics, such as simplification of geometric shapes of the buildings, the photo textures applied on the facades, the similarity of the textures for the other surfaces (pavement, road etc.), the reduction of elements in the street, and the pleasant weather conditions. If results of similar experiments were combined, it would allow general conclusions and predictions on the effect of this particular style of these representations on the affective appraisals of an urban environment. In the absence of these results, a comparison will be made with experiments using other 3D models than the VirtuoCity models.

In a study that is similar in design, Bishop and Rohrmann (2003) assessed cognitive and affective reactions of participants to a suburban park, and its 3D representation. In contrast to our results, the virtual environment was considered less pleasant than the real environment. It is likely that this is related to the type of environment, as well as the quality of the representation. The environment chosen by Bishop and Rohrmann was a relatively quiet and green area with trees and shrubs; a stroll in such an area in reality is considered as a quite pleasant experience, and offers more sources for pleasure than only the visual information (Mambretti, 2007). The area chosen in Venlo contained very little vegetation. Secondly, the visualisation of the vegetation in the model used by Bishop and Rohrmann was not very detailed, and participants’ scores on perceived naturalness were quite low. This indicates that the virtual vegetation did not offer the same pleasing qualities as the real vegetation and negatively influenced the viewers’ response (Paar, 2006). So, whereas some typical features of 3D visualisations have a flattering effect on urban areas, they diminish the pleasing qualities of a natural environment.

Wergles and Muhar found that a real environment in Vienna elicited more negative comments than its virtual representation, in agreement with the negative comments in this study, and the lower score on Pleasantness in real Venlo. In the real Vienna environment negative elements were traffic and noise, and the predominance of asphalt and concrete; this is comparable with our findings that colours are strong determinants of the pleasantness of an area.

In the Vienna visualisation, negative aspects were the monotony and dullness of the area, directly related to restraints of the representation. Also the viewers of Virtual Venlo commented on the absence of traffic and people which makes the environment static; however, the use of photographic textures and the lively colours increased the visual variety of this area, and presence of people is suggested by a terrace.
Comparing the results of this study and the one of Vienna, we can conclude that the type of 3D visualisations as studied here makes the urban environment more pleasant, but also more dull. In the experiment “Visual dynamics and sounds in an outdoor environment” (Chapter 4.6), we examine whether dynamic elements make virtual environments livelier.
4.5 Experiment: Visual dynamics and sounds in an outdoor environment

This section is based on:


4.5.1 Introduction

3D models of environments are not only used for architecture and landscape visualisation, but also for (serious) gaming, where the environment provides the stage setting for incidents and actions of trainees. In the experiment discussed here, we studied the effects of dynamic visual and auditory elements in a serious game, the Levee Patroller, on the appraisal of an environment, and on the affective/emotional response of users. These dynamic elements are often sparingly, or not at all, applied in 3D models due to costs of implementation and unfamiliarity with the effects on the viewer. Human environmental experience is however multi-sensory, so auditory and visual dynamic attributes of a setting, such as wind and waves, are environmental elements which should be included in the representation (Huang, 2004; Huang 2009). The increase in the use of game engines to build virtual environments, offering easy implementation of dynamics and sounds, raises questions about the effect of these visual and auditory elements on the perception and appraisal of the user.

Our study in the first place concerns the fidelity and so-called incremental validity of virtual environments, by assessing to what extent adding an extra sensory channel to the environment and improvements in the visual information enhance the representational validity of the environment (Hetherington et al., 1993; Daniel & Meitner, 2001). This may increase the convincingly or perceived realism of the virtual environment. Events in a scenario that may affect the actions of a trainee in a serious game must be modeled in a convincing and verisimilar way: weather conditions may for instance influence the decisions of levee patrollers, but also of fire commanders.

Sounds and dynamic elements also increase the liveliness and attractiveness of the game environment or the interface to which the user may respond with for instance surprise or joy (Malone & Lepper, 1987). So, in the second place, we examine whether the enhancements influence the emotional response of the user and especially their engagement, which is expected to improve the learning process (Garris, Ahlers, & Driskell, 2002). Influencing the emotional response of the trainee in serious games and training applications is a very valuable means to motivate trainees and to prepare the trainee to perform in a diversity of circumstances.

In this study we added dynamic visual and auditory elements to a 3D outdoor environment. We hypothesized that this would increase the perceived severity of the weather circumstances and make the environment less pleasant, more distressing and more arousing. Sound was expected to have a stronger effect than visual dynamics, and the interaction effect to be the strongest. Visual
dynamics and sounds were expected to increase the arousal of the participants and their appreciation of both the representation and the game, and thus their engagement.

The manipulations in this experiment concern features of the environment, namely ambient conditions, and features of the representation. These are mainly related to the medium itself, such as loss of multimodal information, and to choices in the modeling process, as specified in the framework presented in Chapter 3. We assessed the affective appraisal of the environment, and the emotional response of users, especially their engagement. Besides visual and verbal scales, and questionnaires, we used a psychophysiological measure, facial EMG (electromyography) to assess emotional activation of participants.

4.5.2 Related Work

**Sound and visual dynamics in the representation of a training environment**

3D models used for training and serious games vary from simple and generic environments that supply a background when the focus of the training is for instance on complex cognitive or communicative tasks; to detailed verisimilar environments when elements in the environment are of importance for the goal of the training. Levee Patroller is an example of this second type. The serious game is developed by GeoDelft (now Deltares) and Delft University of Technology using the Unreal Engine 2 Runtime game engine, as training instrument for levee patrollers (Harteveld, Guimarães, Mayer, & Bidarra, 2007). Levee (or dike) patrollers inspect levees at regular times, but especially in situations when danger or damage may occur. Their task is to recognize failure mechanisms in an early stage and communicate relevant findings to the central field office. Practicing these skills in reality is difficult, as levee failures are quite rare. A serious game not only facilitates this but can also show failures of different kinds and severity, and in a variety of adverse weather circumstances which may for instance limit visibility.

The validity of the virtual environment is presumed to be important for the training effectiveness. In the Levee Patroller, firstly, the lifelike representation of the failure mechanisms is essential. Secondly, the game environment with low-lying land, canals, and levees, as well as the weather conditions, must be valid to inform the trainee of the relevant geographical and meteorological circumstances. In the game, strong wind and heavy rain are suggested, which contribute to the danger of the situation. Thirdly, in reality the environment and events that occur, may affect the emotional state of the levee patrollers and consequently their behaviour and performance.

The representation of these atmospheric and geographical conditions in the game is restricted. In 3D environments used in training applications, sound is often lacking, although its importance for the user is obvious from a number of studies. Auditory background and sound events are essential to establish a feeling of contact with and presence in an environment (Dinh et al., 1999; Larsson & Västfjäll, 2001; Sanders & Scorgie, 2002; Whitelock et al., 2000; Serafin & Serafin, 2004). Sound makes the environment come alive, and envelops the user, thereby enhancing her involvement (Blesser & Salter, 2007).

The static appearance of most 3D environments, where the only movement is the consequence of navigation by the user, is another obvious characteristic imposed by restrictions of hardware, software and by the required effort in modeling. Sometimes dynamic textures are used to represent water and clouds, thus increasing the temporal variation in the scene. However, in most cases environmental elements are depicted motionless. The environments thus lack the subtle movements that are characteristic for a real environment, such as branches moving in the wind, birds, cars and trains passing in the distance. In reality, dynamic features in a polder landscape are very important indications for the weather circumstances and the assessment of the risk of levee failures.

Movement is often accompanied by sound, and may be suggested by sound. Adding sounds to a 3D environment may enhance the effect of the visual dynamic elements, or compensate for their absence or perhaps unsophisticated representation.
Sound and visual dynamics as interesting elements in the interface

Abundant evidence from research on films and games indicates that visual dynamic features (El-Nasr, 2006; Gomez, Zimmermann, Guttermosen-Schär, & Danuser, 2005) and sound (Larsson et al., 2001; Morinaga, Aono, Kuwano, & Kato, 2003) can be used to influence the emotional response of a user. Sound is known to promote attention and recognition (Rohrmann & Bishop, 2002), to increase enjoyment, the experience of presence, and to improve performance in memory tasks (Larsson et al., 2001) in virtual environments.

Visual dynamics and sounds increase the sensory richness of a game, and can thus increase engagement (Rozendaal et al., 2009; Sutcliffe, 2002). In a well designed serious game the realism, didactics and engagement factors are in balance to accomplish the purpose of the serious game: to learn in a safe, engaging setting (Harteveld et al., 2007). The sensory richness of a virtual environment is often also related to a higher level of immersion or presence, which is considered beneficial for the learning process (Garris et al., 2002). According to Garris et al. (2002) engagement in repetitive play, or being occupied with the game play, leads to better learning results. Learning is thought to improve as the quality of cognitive engagement increases, and greater engagement during learning is expected to lead to longer retention of information (Hannafin & Hooper, 1993).

By adding dynamic and auditory elements to Levee Patroller we may thus increase the validity of the environment, influence the emotional response, and increase the engagement of the trainee.

4.5.3 Method

Participants

Participants were 55 Dutch residents, 27 males, 28 females. Eight participants (14.5 %) had no or little experience with personal computers, all others “average”or “much”. Forty-one participants (74.5%) had no or little experience with using 3D environments, for instance in games; and only 14 “average” or “much”. None had experience in developing 3D environments, and only one participant knew a little of levee patrolling. After removal of 5 participants who experienced problems with navigation, the average age was 44.7 years, with a standard deviation of 18.5. Participants were randomly assigned to the conditions.

Manipulations

For the experiment a small area of the polder landscape of the Levee Patroller was selected. Two variables were manipulated: the visual environment, which was either static or dynamic, and sound, which was absent or present. Four versions (experimental conditions) were created:

1. static, no sound
2. static, with sound
3. dynamic, no sound
4. dynamic, with sound.

In the first and second versions, some dynamic elements that were already in the game were retained: raindrops, clouds, water surfaces, the failure mechanisms, and a meadow gate that opens on approach.

The dynamic visual effects that were added in versions 3 and 4 were:
- rain is positioned more obliquely (Figure 4.18)
- clouds are darker and are passing faster (Figure 4.18)
- frequent lightning
- branches of trees are moving
- water surfaces are more rough, waves are stronger and more frequent (Figure 4.18)
- a large boat is passing (Figure 4.18)
- cars and a train are passing
- ducks swim in a small canal
- sheep are moving in a meadow (Figure 4.18)
- seagulls are passing
- a small boat is sinking
- a street lamp pole falls over in the storm.

To versions 2 and 4 sounds were added:
- a continuous soundtrack of rain, strong wind and thunder
- positioned sounds for the animals, and the events pertaining to the failure mechanisms.

![Figure 4.18 Screenshots of the dynamic versions (3 and 4)](image)

**Equipment**

The experiments were performed on a Dell Dimension XPS-710 Dual Core computer, equipped with a Dell 19” monitor. A standard keyboard and mouse were used for navigation. For the versions with sound, a Sennheiser EH 150 headphone was provided.

**Materials**

We applied verbal measures to assess the affective appraisal of the environment and weather conditions, and affective/emotional responses of the participants. Additionally, by measuring facial EMG (electromyography) we wanted to gather additional, objective and continuous data on the affective/emotional response of the participants during the task. More specifically, we used the following instruments. After the task in the VE, the participants were asked to fill in a questionnaire. This included the following:
1. The assessment of weather conditions was assessed by nine adjectives (for instance “stormy”, or foggy), based on Stewart (2007). These were rated on a visual analogue scale allowing participants to indicate a point on a line with a range from zero to ten.
2. The convincingness of the representation of the rain and wind, whether they were tiring an annoying, and the congruence of sound and visual representation were summarized in 5 statements. They were rated on a 5-point scale from “totally disagree” to “totally agree”.
3. The affective qualities of the environment were measured with the scale of Russell and Pratt (1980) (Chapter 3). The 40 adjectives were translated into Dutch and checked by a native speaker.
4. For engagement 15 questions were included, relating to the dimensions that Mallon and Webb (2000) proposed for engagement. Five questions concerned the “structure” dimension (spatial characteristics and causal connections in the game play), and five the interaction with the virtual environment. These questions concern elements in the game and environment that are important for engagement, such as the layout of the area and cues to assist the user. Five other questions were drawn from the ITC-SOPI physical presence questionnaire, concerning the actual experience of engagement (Lessiter et al., 2001).
5. The valence (pleasure) and arousal score on the Self-Assessment Manikin (SAM; Lang, 1980) was obtained before and directly after the game play. The scores are from 1-9. Engagement with games is assumed to be related to high scores on positive valence and arousal.
6. During the task in the VE two facial muscles were monitored through electromyography. The tension of the corrugator (frown muscle) and the zygomatic muscle response (smile muscle) was adopted as an indicator for the experience of negative and positive valence respectively (Lang, Greenwald, Bradley & Hamm, 1993; Bolls et al., 2001; Fujimura, Sato, & Suzuki, 2010) and measured with the Mobi and Kendall ECG electrodes (H124 SG).

Procedure
On arrival, participants read and signed an informed consent and had a moment to relax. The pre-test questionnaire for general information was presented next, after which the participants rated the SAM for the first time. The experimenter explained the navigation and the task and placed the electrodes.

For the versions with sound participants were asked to put on the headphone. They were allowed to place the headphone further to the back of the head if they found the sound volume uncomfortable (which only a few did). The assignment was to find and mark three failure mechanisms and to draw a map of the inspected area after the task.

To increase the comprehension of the layout of the area during the experiment, we informed participants that they had to draw a map of the game environment and the important landmarks, after completing the assignment.

The duration of the experiment was 8.40 minutes. Finally, the second series of questionnaires was completed, including the SAM, questions on engagement and the appraisal of the environment and representation.

Hypotheses
The visual dynamic elements and sounds were added to compensate for information lacking in the representation of severe weather conditions. We therefore hypothesized an increase in the perceived severity of the weather circumstances. We also expected to find a decrease in the perceived pleasantness and an increase in the distressing and arousing qualities of the environment, in the manipulated versions. Sound was expected to have a stronger effect than visual dynamics, and the interaction effect to be the strongest.

Visual dynamics and sounds were expected to increase the arousal of the participants and their appreciation of both the representation and the game, and thus their engagement. As for the performance in locating the failure mechanisms, the hypothesis was more tentative. On the one hand, the manipulations might increase engagement and therefore concentration and performance. On the other hand, visual and auditory effects could distract or even irritate participants and have a
negative influence on performance, or be perceived as unpleasant (Wolfson & Case, 2000; Morinaga, Aono, Kuwano, & Kato, 2003), and obscure the indications of failure mechanisms, as in real life.

4.5.4 Results

1. Assessment of weather conditions

The items listed in the questionnaire were stormy wind, somber, cold, wet, cloudy, calm, foggy, threatening and thunderstorm. The scores indicate the extent to which a weather condition was considered appropriate by the participants. The mean scores for the conditions show that the manipulations were very effective in influencing the perception of certain weather conditions, especially “stormy wind” and “thunderstorm”, with scores of ca. 8 on a scale from 0 to 10. Thunderstorm in the simulation includes lightning and thunder (Table 4.15).

A two-way ANOVA with factors visual dynamics and sound, shows a significant main effect for sound on the items “calm” (F(1,46)=5.05, p=.029), “threatening” (F(1,46)= 4.11, p=.048); and thunderstorm (F(1,46)=18.47, p=.00). So, as expected, sounds are effective in indicating stormy weather conditions. A significant main effect for visual dynamics was found for thunderstorm (F(1,46)=7.08, p=.011). An interaction effect was found for stormy wind: post-hoc Tukey, 2-tailed p< .01, and thunderstorm: post-hoc Tukey, 2-tailed p< .01. The effect of the combination of sound and dynamics was also noticed in the item “threatening” (ANOVA analysis, post-hoc Tukey, 1-tailed p< .05). On the other items (somber, cold, wet, cloudy, calm, and foggy) no significant differences were found.

<table>
<thead>
<tr>
<th>Weather</th>
<th>condition</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>stormy wind</td>
<td>1</td>
<td>12</td>
<td>4.21</td>
<td>2.81</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13</td>
<td>7.20</td>
<td>2.28</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>13</td>
<td>6.22</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>12</td>
<td>8.06</td>
<td>1.38</td>
</tr>
<tr>
<td>calm</td>
<td>1</td>
<td>12</td>
<td>4.86</td>
<td>2.58</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13</td>
<td>2.93</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>13</td>
<td>4.01</td>
<td>3.38</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>12</td>
<td>2.33</td>
<td>2.52</td>
</tr>
<tr>
<td>threatening</td>
<td>1</td>
<td>12</td>
<td>4.81</td>
<td>2.84</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13</td>
<td>6.21</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>13</td>
<td>5.92</td>
<td>2.69</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>12</td>
<td>7.60</td>
<td>1.15</td>
</tr>
<tr>
<td>thunderstorm</td>
<td>1</td>
<td>12</td>
<td>.95</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13</td>
<td>6.45</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>13</td>
<td>4.85</td>
<td>4.61</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>12</td>
<td>7.78</td>
<td>2.18</td>
</tr>
</tbody>
</table>

Note: The score runs from 0 – 10.

2. Judgment of the representation of the rain and wind

Five statements were included on the perceived quality and convincingness of the representation of rain and wind. A two-way ANOVA with factors visual dynamics and sound results in a significant interaction effect for the perceived realism of rain (F (1, 46) =7.19, p=.01). Contrary to what might be expected, the version with sound and visual dynamics had the lowest score. The version with visual dynamics but without sound showed a significantly higher score compared to the full version indicating that in this experiment sound did not contribute and even hindered the perceived quality of rain (ANOVA, and post-hoc Tukey p=.024). A significant main effect was found for sound and for visual dynamics, on the perceived realism of wind (F(1, 46)= 11.05, p=.002, and F(1,46)=6.66, p=.013)). Results are shown in Figure 4.19.
A significant main effect was found for sound for one other aspect of the representation of weather conditions: the conditions with sound were considered more tiring and annoying \( F(1, 46) = 8.075, p=.007, F(1, 46) = 6.012, p=.018 \). However, all scores remained just below the score of 3 on the 5-point scale; therefore the participants did not consider the sounds too annoying.

Figure 4.19 Mean scores for the perceived convincingness of rain (left) and wind (right). The Y-axis represents the values on a 5-point scale (1 = “not at all”, 5 = “very much”).

3. Appraisal of the environment

We again performed two-way ANOVA’s with visual dynamics and sound as factors, for the 8 unipolar scales to assess the affective quality of the environment developed by Russell and Pratt (1981). No significant interaction effects between the factors were found. There is a significant main effect for sound on 4 scales: Arousing \( F(1,46)=5.31, p=.026 \), Distressing \( F(1,46)=9.31, p=.004 \), Gloomy \( F(1,46)=5.98, p=.018 \), and Relaxing \( F(1,46)=11.94, p=.001 \). Sound increases the arousing and distressing qualities, and decreases the gloomy and relaxing qualities of the environment.

For visual dynamics we found a significant main effect on the scales Pleasant \( F(1,46)=7.29, p=.010 \), Gloomy \( F(1,46)=5.79, p=.020 \) and Unpleasant \( F(1,46)=9.53, p=.003 \). Visual dynamics increase the pleasant qualities, and decrease the gloomy and unpleasant qualities of the environment. Table 4.16 shows the scores on the Russell and Pratt scale. Post-hoc Tukey tests resulted in a significant difference on the scale Gloomy between versions 1 and 4 \( p<.01 \), Relaxing (versions 1 and 2, and 2 and 3, both \( p<.05 \)), and Pleasant as well as Unpleasant between versions 2 and 3 (both \( p<.05 \)). The condition with sound only was considered more unpleasant, and less pleasant, than the condition with only visual dynamics.

Contrary to our expectations the enhanced environments, containing stronger indications of severe weather circumstances, are not considered less pleasant. The dynamic versions 3 and 4 are even considered more pleasant than the static versions. The addition of sound without visual dynamics (version 2) seems to have a negative effect on the appreciation of the environment.
Figure 4.20 Mean scores on the scales of the Russell and Pratt scale. The lines between the measuring points are only added to facilitate comparison of the results.

Table 4.16 Mean scores on the scales of the Russell and Pratt scale.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Version</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arousing</td>
<td>1</td>
<td>3.88</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.68</td>
<td>.86</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.25</td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.93</td>
<td>.97</td>
</tr>
<tr>
<td>Exciting</td>
<td>1</td>
<td>4.30</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.28</td>
<td>.90</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.51</td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5.25</td>
<td>.92</td>
</tr>
<tr>
<td>Pleasant</td>
<td>1</td>
<td>4.52</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.85</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5.18</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5.05</td>
<td>1.18</td>
</tr>
<tr>
<td>Distressing</td>
<td>1</td>
<td>3.02</td>
<td>.85</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.99</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.05</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.87</td>
<td>.87</td>
</tr>
<tr>
<td>Sleepy</td>
<td>1</td>
<td>4.32</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.74</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.89</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.50</td>
<td>1.25</td>
</tr>
<tr>
<td>Gloomy</td>
<td>1</td>
<td>4.60</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.97</td>
<td>.93</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.98</td>
<td>1.55</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.02</td>
<td>.84</td>
</tr>
<tr>
<td>Unpleasant</td>
<td>1</td>
<td>3.75</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.97</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.68</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.97</td>
<td>1.32</td>
</tr>
<tr>
<td>Relaxing</td>
<td>1</td>
<td>5.13</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.57</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5.08</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.13</td>
<td>1.50</td>
</tr>
</tbody>
</table>
Table 4.17 Post-hoc Tukey results of the ANOVA-analysis of the Russell and Pratt scale.

<table>
<thead>
<tr>
<th>Scale</th>
<th>condition</th>
<th>condition</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasant</td>
<td>2</td>
<td>3</td>
<td>.037</td>
</tr>
<tr>
<td>Unpleasant</td>
<td>2</td>
<td>3</td>
<td>.039</td>
</tr>
<tr>
<td>Gloomy</td>
<td>1</td>
<td>4</td>
<td>.008</td>
</tr>
<tr>
<td>Relaxing</td>
<td>1</td>
<td>2</td>
<td>.019</td>
</tr>
<tr>
<td>Relaxing</td>
<td>2</td>
<td>3</td>
<td>.022</td>
</tr>
</tbody>
</table>

Russel and Pratt (1980) have combined the eight unipolar scales into four bipolar dimensions, and computed the correlations among the scales. This results in an “affective quality space” as shown in Figure 2.3, in which the x and y axes are respectively, the horizontal (pleasant-unpleasant) axis and the vertical (arousing-sleepy) axis. The affective quality attributed to an environment by an individual can be represented by a single point in this circumplex, (Figure 4.21). We notice a shift in appraisal between the conditions: the environment without sound and dynamics (1) is considered relaxing (least arousing and slightly pleasant); the environment in the sound-only condition (2) was experienced as slightly unpleasant and arousing, so somewhat distressing; the environment with dynamic elements (3) pleasant and minimally sleepy; the environment with sound and dynamic elements (4) exciting (pleasant and arousing).

![Figure 4.21 Plot of the affective quality of the environment measured with the Russell and Pratt scale, for the means of each condition.](image)

4. Affective/emotional response: engagement questionnaire

The first five questions of the engagement questionnaire concerned the structure dimension distinguished by Mallon and Web (2000), and the next five questions the interaction dimension. A two-way ANOVA with visual dynamics and sound as factors found no significant differences (p>.05) between the conditions for structure; this means that sound or dynamics did not make the spatial characteristics of the environment more engaging for the gameplay. Also for the dimension
interaction, no effects of sounds or visual dynamics were found: the added factors did not make interaction and control of the environment more engaging.

The analysis of the questions on engagement derived from the ITC-SOPI questionnaire showed only a main effect of visual dynamics ($F(1, 46) = 5.65, p= 0.02$). On a 5-point scale, the mean for the condition with only visual dynamics (3) was 3.80 (st. dev. 0.55); for the other conditions: condition 1, mean 3.22 (st.dev. 0.52); condition 2 (sound only) mean 3.52, (st.dev. 0.38) and condition 4 (sound and dynamics) mean 3.62, (st. dev. 0.54). This indicates that visual dynamics increased the engagement of the participants during their task in the virtual environment.

5. **Valence and arousal on the SAM (Self Assesment Manikin)**

Averaged over all conditions, valence (pleasure) was rated lower on the SAM-scale after game play than before ($t(49)= -3.21, p<0.01$; Figure 4.22). Although the decrease was largest in the sound conditions (2 and 4), no interaction effects were found due to large variations between participants. An increase in the evaluation of arousal was found over the course of the experiment ($t (49) = 3.32, p<0.01$; Figure 4.23).

![Figure 4.22 Valence (pleasure) before and after game play. The scale runs from 1 (very high valence) to 9 (very low valence).](image)

![Figure 4.23 Arousal before and after game play. The scale runs from 1 (very high arousal) to 9 (very low arousal).](image)

6. **Facial electromyographic (EMG) reactions**

Figure 4.24 shows the facial muscle tension of one typical participant. The blue area (X) shows the tension of the zygomatic region and the green area (Y) the tension of the corrugator region. The red lines present the triggers which were marked for the actions that followed the discovery of the failure mechanisms. We looked for reactions of the facial muscles, as an indication of increase or decrease in valence.
Figure 4.24 Example of EMG data acquired during game play for one participant.

The muscle tension measured in the zygomatic region did not show action potentials shortly before or after these triggers. However, a gradual increase of zygomatic muscle tension was found over all participants, indicating an increase in pleasure. Indeed, the median of the tension before the game differs significantly from the median during the game play ($p<0.01$). However, no difference of increased tension was found between the conditions (Table 4.18).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average difference in median in μV</th>
<th>Percentage of change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.51</td>
<td>64%</td>
</tr>
<tr>
<td>2</td>
<td>12.01</td>
<td>46%</td>
</tr>
<tr>
<td>3</td>
<td>4.73</td>
<td>69%</td>
</tr>
<tr>
<td>4</td>
<td>7.03</td>
<td>76%</td>
</tr>
</tbody>
</table>

Table 4.18 EMG zygomatic region

In all conditions, the corrugator region showed clear action potentials during the game. These could not be matched to the triggers, and they probably represent response to other events. Trouble with navigation, or the insecurity about the assignment could have contributed to the amount of increased tension in the corrugator region. No significant difference was found between the conditions. Perhaps effects were obscured by large individual differences that were noticed (Table 4.19).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average frequency of action potentials</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.9</td>
<td>11.2</td>
</tr>
<tr>
<td>2</td>
<td>11.9</td>
<td>8.9</td>
</tr>
<tr>
<td>3</td>
<td>16.5</td>
<td>14.1</td>
</tr>
<tr>
<td>4</td>
<td>8.0</td>
<td>9.7</td>
</tr>
</tbody>
</table>

Table 4.19 EMG corrugator region.

Performance
Participants did not perform better in the enhanced versions, possibly because the visual conspicuity of the failure mechanisms was similar and dominant in all versions.

4.5.5 Conclusions and discussion

The results of this experiment show that sound and visual dynamic features can effectively be used to convey the impression of certain severe weather conditions in a 3D environment. Both the severity of storm and thunderstorm, as well as the perceived quality of the representation of windy weather conditions was rated higher in the manipulated conditions. The sounds that were added
were more effective than the visual dynamics to indicate windy conditions. For rain, the visual dynamics were judged more realistic than the sound effects.

The severity of the weather circumstances would in real life negatively influence the affective appraisal of the environment and increase the urgency of the situation. In this experiment, we found that sound increased the arousing qualities of the environment, so made the environment less “dull”; however sound only made the environment less pleasant when not accompanied by visual dynamics. Visual dynamics increased the pleasant qualities of the environment. Together they made the environment livelier.

Apparently the enhanced conditions offered more variety and were more interesting for users, compared to the rather lifeless environment of the first condition. This is an important conclusion regarding the validity of the appraisal of virtual environments, especially those with visual and auditory dynamic features: they do matter and cannot be omitted or simplified without consequences for the affective appraisal for an area.

The lower scores on pleasantness for the sound-only condition seem to confirm earlier findings that auditory events, originating from within the field-of-view but not accompanied by visual dynamics, may confuse the user (Hetherington et al., 1993). This can be considered as an incongruity in the representation, where movement suggested by auditory effects is not corroborated visually and the environment is thus not ecologically valid. On the other hand, the participants did not indicate that they disliked the presence of sounds in the sound only condition. The negative appraisal may result from a more subtle effect, sometimes used in cinematography, of a contrast between sounds and visuals that results in an eerie atmosphere.

The subjective evaluation through the engagement questionnaire shows that only the visual dynamics had a positive influence on the engagement of the participants. In contrast to what we expected, sound did not contribute to the user engagement of this game, at least in this experimental setup, compared to other game design elements like challenge, discovery, feedback, and control as defined by several authors.

The evaluation of arousal, acquired with the SAM, shows an increase between the measurement before and after the task.

In all conditions the zygomatic muscle tension increased during the game, which is considered as an indication of higher valence (pleasure). Again, no difference was found between the conditions. In contrast, the subjective (SAM) measures of valence (pleasure) were lower after the game than before, over all conditions. Many participants reported that during their exploration of the virtual environment they worried about the task afterwards: this task included drawing a map of the environment, which they considered difficult. We suggest that the facial reactions in the zygomatic region were direct and fleeting responses to pleasing or surprising elements in the environment, but that these elements did not compensate for the apprehension about the task.

The corrugator region shows a frequency of tension that cannot be related to the triggers in the game. Instead, it seems to react to events in the environment which were not under control of the participant. In addition to a reaction to these events, frustration due to navigational problems could cause the polarization. This cannot be related with the interaction questions of the engagement questionnaire. The influence of sound and dynamic elements on the facial muscle tension cannot be deduced from the data, due to large individual differences. A more controlled environment, for example a predefined route through the environment, may give more insight in the relation between physical reactions and environmental events. It has been noted often that facial EMG is known to be susceptible to noise, for example, evoked by sudden movements of the subject which reduces reliability (e.g. Poels & Dewitte, 2006); this may have contributed to the difficulties in interpreting the data.

The results of this experiment confirm the importance of elements in the first and second column of the framework, presented in Chapter 3: environmental dimensions (here ambient conditions) and representational modifiers and determinants. Dynamic features, visual and auditory,
that are often added to 3D models of environments in a somewhat arbitrary manner in the modeling process, are relevant for the validity of the virtual environment. However, their effects on the appraisal of a virtual environment may diverge from expectations. On the one hand, severe weather conditions can be suggested effectively by these cues; users recognize them easily. On the other hand the unpleasant weather conditions do not have the expected effect on the affective appraisal of the environment; they make the environment more pleasant. Users appreciate diversity in the virtual environment, even when the features in reality are considered unpleasant.

**Acknowledgements**

We are grateful to Deltares for providing us with the original model and performing the manipulations, and to the KNMI (Royal Netherlands Meteorological Institute) for advice on the representation of the weather circumstances.
4.6 Experiment: Experiencing danger on a virtual highway

4.6.1 Introduction

In the experiment described in this section, the effect of manipulated depth and speed cues on the affective/emotional response of the user, especially the feeling of risk, was assessed. Visual and auditory depth and speed cues provide the user with essential information about the simulated environment and the events occurring therein. The role of these factors in tasks like navigation and distance assessment has been extensively investigated. However, little is known of their effects on the affective/emotional response of the user.

We selected these factors because they are modified by the representation on a conventional 2D display (which does not for instance, afford stereopsis) and are expected to affect the emotional response of the user (see Chapter 3). The manipulations were performed in a virtual environment representing a highway with traffic and incidents, which is used as examination program to assess procedural knowledge of highway patrol officers. In the examination the affective/emotional response of the user to the environment, the awareness of a dangerous situation, and especially the feeling of risk, play a role in the actions and decisions of the exam candidates, and thus relate to the validity of the environment. Previous evaluation of the application had shown that the environment did not convey the danger that in reality is involved in the simulated events (E-semble, personal communication). As a result, users did not experience a feeling of risk, although the incidents were modeled accurately. Exam candidates and instructors were convinced this influenced their decisions and thus their performance.

We selected visual and auditory cues that (1) are known to contribute significantly to depth and speed perception (Blakemore & Snowden, 2000; Pfautz, 2002; Sherman & Craig, 2003; Pretto & Chatziastros, 2006; Pretto, Vidal, & Chatziastros, 2008; Pretto, Ogier, Bülthoff, & Bresciani, 2009; Friedman, 2007), (2) can be enhanced at low costs in time and budget, and (3) fit within the context of the application and exam. The manipulations concern factors of the representation that are affected by the medium, and by characteristics of the platform, as specified in the framework presented in Chapter 3. We assessed the affective/emotional response of the participants, and in one part of the experiment, the effect on behavioural response, namely their decision to cross the road, and actually crossing the road.

Our hypothesis is that the introduction of additional visual and auditory depth and speed cues would increase the feeling of risk experienced by the users.

4.6.2 Related Work

The world surrounding us is generally full of objects of different dimensions and various velocities, while we are moving through it in different directions and with variable speeds. In real cityscapes for instance we are surrounded by a multitude of moving objects like cars, cyclists, pedestrians, flags and branches of trees moving in the wind, birds, insects, and clouds. Desktop simulations used for visualisation or training typically contain a small number of moving objects, and when navigating on desktop monitors the world often seems to move around the viewpoint of the user in a restricted window. The static appearance and lifelessness of virtual environments is one of
the characteristics most often remarked on by users. Adding visual and auditory dynamic elements to a virtual environment can increase the pleasing and exciting qualities of the environment, as shown in the experiment described in Chapter 4.5. In this experiment we focused on the experience of being located in a dynamic situation with many moving elements, and the effects on the user’s emotional response.

The application chosen for the experiment was a virtual environment used in the examination of highway patrol officers. Highway patrol officers are appointed by the Dutch Road and Traffic Authority, Rijkswaterstaat (RWS) and survey Dutch highways. They monitor the traffic situation, ensure safety during incidents, and clear the way for other emergency services like the police. Examinations to assess procedural knowledge of candidate highway patrol officers currently use a VE which represents a highway environment with traffic incidents. Confronted with a given simulated highway incident, trainees navigate through the VE using a joystick, and decide which actions they should take to resolve the problem and to restore normal traffic flow, while minimizing the risk for other road users and themselves, in accordance with the rules and prescribed procedures. Previous evaluations showed that this VE does not convey the danger involved in the simulated events (E-Semble, personal communication). As a result, users do not experience a feeling of risk, even though important elements in the virtual scenario which are known to contribute to perceived danger in reality, such as the speed and proximity of vehicles, are accurately modeled. Users and instructors are convinced that this simulation deficiency leads to incorrect decisions and consequently degrades their performance during exams.

**Feeling of risk**

The feeling of risk is a fast affective response to a threat, mediated by an intuitive and mostly automatic, unconscious experiential system (Slovic, Finucane, Peters, & MacGregor, 2004). This experiential system is believed to rely on images and associations, which are linked by experience to emotions and affective feelings. In this perspective, the vividness with which future outcomes of a situation are described or represented mentally is thought to be one of the most important determinants of emotional reactions to future outcomes.

The experiential system is thought to work in parallel with a rational system, which relies on probability calculus and risk assessment. Because their determinants are different, affective reactions to risks often diverge from cognitive evaluations of the same risks. According to the so-called risk-as-feelings hypothesis (Loewenstein, Weber, Hsee, & Welch, 2001), emotions often produce behavioural responses that depart from what individuals view as the best course of action (Figure 4.25).

![Figure 4.25 Risk–as-feelings perspective, Loewenstein et al. 2001.](image)

The vividness of the mental representation of future outcomes is an important determinant of emotional reactions to those outcomes (Loewenstein et al., 2001). In the case of the highway patrol officer exams, the anticipation of accidents, crashes and their consequences would influence cognitive evaluations and feelings (affect) of the user; and vividness of the anticipated outcomes,
induced by the VE, especially the feelings of the user. When used as an examination tool, the VE should therefore convey a valid impression of the risk involved in a simulated incident (Figure 4.25).

In real-life critical situations, visual and auditory depth and speed cues are important factors that determine risk perception by highway patrol officers, since these factors relate to danger for themselves and others. When estimating depth and motion from a VE represented on a 2D display, viewers have to rely on a restricted set of pictorial cues, which are automatically generated by the 3D rendering software, and on elements added or enhanced by the developers. Generally, distances and speed are underestimated in virtual environments, resulting in what is often called a “compressed” view of the world (Campos, Nusseck, Wallraven, Mohler, & Bülthoff, 2007).

In an attempt to provide users a more compelling impression, we enhanced and added several visual and auditory depth and speed cues in the VE, and added variability to traffic behaviour. From the cues that are known to determine the perception of depth and speed in virtual environments we selected those that can easily be implemented in standard desktop environments. Visual cues were added to or enhanced in this VE by adding extra landmarks to support motion parallax (Pfautz, 2002; Gogel & McNulty, 1983), using a 3rd person perspective (Salamin et al., 2006), adapting the viewing angle (Kemeny & Panerai 2003; Ooi, Wu, & He, 2001), improve shadowing (Slater et al., 2002), adding an atmospheric haze effect (Cutting, 1997; Sherman & Craig, 2003) and by varying traffic behaviour. The soundtrack added to the VE includes ambient highway sounds, and the sounds of passing vehicles and truck horns.

The goal of the experiment presented here was to assess whether enhancement of the representation of the highway patrol officers’ exam VE with these cues induces a feeling of risk; and whether they convey the danger of physical harm for the highway patrol officer and the other road users as convincing and urgent. The feeling of risk in a VE has previously been investigated in the context of racing games (Fischer, Kubitzki, Guter & Frey, 2007) and the validity of driving simulators. The possibility to enhance the feeling of risk by manipulating visual and auditory depth and speed cues in a non-immersive VE has, to our knowledge, not been investigated before.

Enhancing visual and auditory depth and speed cues in a VE may affect the user’s feeling of risk in several ways. The additional visual cues may make the VE more convincing, so that it evokes more easily images of past experiences and their associated affective reactions (Slovic et al., 2004). The additional sound cues may improve the sense of location (Hawkins, 2005), and may effectively evoke memories, associations and provoke images, thus enhancing the emotional experience (Pressing, 1997, Forrester 2002). Finally, arousal is believed to be an important component of the complex feeling of risk. Both image motion (Detenber, Simons & Reiss, 2000) and sounds (Dekker and Champion, 2007; Krebs, 2002; Sanders and Scorgie, 2002) are known to effectively induce arousal in the user.

4.6.3 Method

Participants
Participants were 28 highway patrol officers, 2 of which were seniors; 26 males and 2 females. Their age varied between 23 and 60 (M=44.6, SD=9.5) years old. All but one had previously participated in a session with a version of the highway patrol officers’ exam environment. Most participants had little or no experience with other gaming and/or simulation environments.

Participants were informed that the purpose of the experiment was to improve simulation quality for future highway patrol officers’ examinations, and to gain more insight into serious games for training applications.

The participants were randomly divided into two groups of 14 people each. One group was presented with the original version of the simulation, the other with the enhanced version.

Virtual Environment and Scenario
The simulations used in this study were both the original and an enhanced version of the “RWS Weginspecteur Callandtunnel 1.12” examination program for highway patrol officers, developed by E-Semble (www.e-semble.com) for the Dutch Directorate-General for Public Works and Water
Management. The original application is used both as an examination tool for candidate highway patrol officers, and as a knowledge refreshment tool for practicing ones. The program simulates a highway environment and several types of incidents. In the exams, the exam candidate stands about 3 meter from the projection screen on which the environment is shown, and uses a joystick to navigate through the environment. The trainee has the role of the highway patrol officer. An instructor reacts verbally and impersonates all other persons that occur in the scenario (traffic control center, drivers involved in an accident, the police, etc.). An operator, using a separate module, responds to changes in the situation or requests of the trainee, by performing appropriate actions in the virtual environment, such as making police cars arrive at the scene, or moving objects according to the instructions of the candidates.

The incident selected for this research represents a situation in which a van has come to a standstill on the left lane of the highway, after colliding with a big bag of sand that has fallen off the back of a truck. The driver of the van stands next to the van, apparently not knowing what to do next.

In a regular course of actions, trainees should drive up to the incident, select a safe spot to park their car, take actions to ensure the safety of all those involved in the incident (including their own), and restore traffic flow as quickly as possible. After communicating the situation to the traffic control center, trainees should assess which traffic measures (e.g. blocking all traffic on a single lane), safety measures (e.g. instructing the van driver to move to a safer position) and problem handling measures (e.g. calling assistance of a tow-truck) are required. To perform their task, trainees should at any time be able to rapidly assess

- their own (egocentric) distance and speed relative to moving or static vehicles and objects,
- the (exocentric) distances and speed between other vehicles and objects, and
- the path of the moving vehicles relative to themselves.

![Figure 4.26 Annotated screen shot of the enhanced version of the simulation.](image)

**Visual Enhancements**

In the enhanced (experimental) version of the environment 8 major visual cues were manipulated (Figure 4.26). Six of these affect depth and speed perception, and two are changes in traffic behaviour.

The original simulation presents a 1st person perspective, both in- and outside the car. The enhanced simulation uses a 3rd person perspective when the candidate is outside the car. The scene is seen from 0.5m behind the user’s avatar and from a height of 1.80m, looking downwards at an
angle of 36 degrees (item 1 in Figure 4.26). The larger simulated field of view (the FOV used by the graphics generator to render its images) used in the 3rd person perspective, in combination with the reference provided by the avatar’s head, helps users to better estimate distances, and to better anticipate and extrapolate the speed and trajectory of vehicles (Salamin et al., 2006). A larger simulated FOV also allows the representation of more motion cues (vehicles) in the peripheral part of the display, thereby enhancing the perception of speed (Pretto et al., 2009) and increasing perceived speed (Gogel & McNulty, 1983). The downwards tilted viewing angle slightly increases the visual angle between objects, thereby increasing their perceived separations. Also, objects on the ground appear closer because their angle with the horizon increases (Ooi et al., 2001). Furthermore, observers can use the additional near ground information available in this camera view to scale far ground details, thereby obtaining more accurate distance estimates (Wu et al., 2004).

By adding trees to both sides of the road (item 2 in Figure 4.26), an extra depth layer is created, which provides extra size and motion parallax cues, and partly occludes the horizon. The trees also constitute a background texture for the cars on the highway, which is likely to increase their perceived speed (Blakemore & Snowden, 2000; Gogel & McNulty, 1983).

In the enhanced version trucks (item 3 in Figure 4.26) also appeared in the user’s lane, in an attempt to:

- improve distance estimates (trucks provide additional occlusion and relative size cues),
- increase the sense of danger (trucks have a larger visual impact than cars when passing close by, which may increase the looming sensation of impending danger (Reeves & Nass, 1996),
- increase the perceived risk (objectively, a collision with a truck has more serious consequences than one with a car (Sjöberg, 2000),
- increase the difficulty of predicting traffic behaviour (trucks occlude other vehicles).

In the original simulation the shadows beneath all vehicles disappeared slightly into the ground. This may have presented the viewer false information, leading to incorrect distance estimates (Goldstein, 2006). The shadows were corrected in the enhanced version (item 4 in Figure 4.26).

Aerial perspective was introduced in the simulation by adding a distant haze (item 5 in Figure 4.38), which gradually reduces environmental contrast at distances larger than about 400m. To maintain atmospheric consistency, the sunny sky was replaced by a texture representing light clouds near the horizon (item 6 in Figure 4.26).

Predictability of traffic behaviour was decreased by introducing random variations in the distance of the vehicles to the edge of the lane (item 7 in Figure 4.26), and in the distance between vehicles (item 8 in Figure 4.26). Increasing behavioural uncertainty is expected to increase the user’s perception of risk (Sjöberg, 2000). Irregular traffic behaviour also makes the simulation appear more natural. Figure 4.28 shows examples of the visual modifications for each of the 5 video clips.

**Auditory enhancements**

Highway traffic generates a typical and recognizable sound, consisting of engine sounds, wind and friction of tires, which is highly familiar to highway patrol officers and linked to their work experience. The soundtrack of the original simulation represents only general traffic noise, which is not specific for a highway environment. Since “ambient sound” (Pressing, 1997) significantly determines the perceived atmosphere of a setting (Hawkins, 2005), a soundtrack was made that more closely resembles ambient highway sounds. The sounds were recorded alongside a road with traffic passing at the same speed as simulated in the virtual environment (80 km/h). In addition, car engine sounds were recorded inside a Mitsubishi L200 pick-up truck travelling at 80km/h. These sounds were added to the clip in which the observer drives a company pickup truck, such that the pitch is related to the speed of the car. This type of non-spatialized engine sounds is known to increase experienced translational self-motion, probably by enhancing the perceived dynamics of the displayed scene (Väljamäe, Larsson, Västfjäll, & Kleiner, 2008).

In the experiment, one particular sound was used for all cars, and another one for all trucks. Sounds were synchronized with vehicles in the nearby lane. They included the Doppler-effect, which
may improve distance and speed perception since it provides a cue for both (e.g. Begault, 2000; Friedman, 2007). In the enhanced simulation, one in ten trucks passing sounds its horn. The horn’s high volume and unpredictability is expected to draw the attention of the observer. Its association with danger should also increase the user’s perception of risk (Hawkins, 2005).

Note that vehicles passing by on the far side of the road make no sound, in contrast to vehicles on the near side of the main road.

**Stimuli**

In an actual exam, candidates interact with the simulated environment, in which many events are effectuated by an operator, and simulated traffic behaviour shows random variations. To provide all participants with the same conditions in this experiment, the selected scenario was used to create five video clips, representing consecutive phases in the regular course of actions during incident support. The interactive simulation was only used in a final assignment. The length of the video clips is about 30 seconds. The first clip shows the approach of the site of the incident, seen from the driver’s position inside the company vehicle. The clip ends when the car has passed the location of the incident on the left lane. The second clip again shows the driver’s view from inside the car, after it has been parked on the safety lane next to the incident. The third, fourth and fifth clip all feature the participant standing on the safety lane, next to the incident (on the left lane of the main road). The viewing direction is respectively facing upcoming traffic, looking directly at the incident (perpendicular to the traffic direction) and looking in the direction of traffic movement.

**Location and equipment**

The experiment took place in a dimly lit room. A HP Compaq 8510p laptop with a Centrino Pro T7700 (2.4GHz) processor, 2 gigabytes of memory (RAM) and a ATI X2600 (256MB) video card was used to run the simulation and present the video clips. The virtual environment (video clips and real-time simulation) was projected on a wall using a Sony projector with a resolution of 1024x768 pixels at 2000 lumen. The projection was about 125cm (height) by 160cm (width), covering a field of view of about 23.5° horizontal x 30° vertical at the viewing distance of 3 meters (similar to the conditions in regular RWS examinations). Two 5 Watt speakers (Creative SBS260) were placed at the front left and right sides of the participant at a distance of about 40 centimeters, to simulate a stereo effect. They were set to one third of their maximum output, which was loud enough to make all sounds clearly audible. A Logitech Extreme 3D Pro joystick was used for maneuvering during the interactive part of the experiment.

**Procedure**

First the participants filled out a general questionnaire (referring to age, game experience etc.) and indicated their emotional state (arousal and valence) on a pictorial scale (the Self-Assessment Manikin or SAM: Bradley and Lang, 1994).

Then they watched five video clips, representing consecutive characteristic events and actions that occur in the simulated scenario during the course of a regular interactive exam. After watching each video clip the participants filled out a form with questions about

- perceived speed (e.g. “How fast was the traffic on the nearest lane going?”),
- perceived distances (e.g. “What is the distance to the matrix-signs?”),
- perceived danger of the situation (e.g. “How do you rate the danger for other road users?”), and
- their feeling of risk (e.g. “I felt I wanted to step backwards).

Some questions, such as perceived speed of traffic on the far side of the road, were relevant in only a few of the clips.

After the last video clip participants scored questions on how convinced they felt about their speed and distance estimates in the video clips. They indicated their emotional state on the SAM scales for the second time.

In the second part of the experiment participants were asked to move through the interactive VE using a joystick, and to cross one of the simulated main road lanes. Before crossing, they first had
to walk to the end of the safety lane, which allowed them to get used to the walking speed in the VE. At the end of the lane participants had to judge if and when it was safe to cross the road. The assignment ended when the participant had either crossed the lane, or did not cross within 3.5 minutes. Because traffic was randomly generated, each participant was confronted with a slightly different stimulus. The specified settings were such that there was at least one chance to cross the road every 30 seconds. The time interval between the moment of arrival at the crossing location and the moment they reached the opposite side of the road was adopted as a measure of their risk appraisal.

Directly after completing this task participants filled out a form with questions relating to

- experienced arousal and valence during the assignment, on the SAM scale (Bradley and Lang, 1994),
- experienced engagement and spatial presence during the assignment, including questions on the perceived convincingness and realism of the environment (derived from the ITC-Sense of Presence Inventory questionnaire: Lessiter et al., 2001), and
- their affective appraisal of the environment. We used a 7 point bipolar scale with the adjectives boring – exciting, somber – bright, unpleasant – pleasant, ugly – beautiful and enclosed – spacious, derived from the factors used by e.g. Russell and Pratt, 1980.

The entire experiment took about 15-25 minutes.

4.6.4 Results

*Distance perception*

For 3 of the 5 video clips participants provided distance estimates, both for the original and enhanced simulations (see Table 4.20). T-tests reveal that

- there is no significant difference in perceived distance between both simulation conditions for all three video clips (p = .26, p = .76 and p = .86 respectively),
- all average distance estimates for the original simulation differ significantly from the actual distances (Table 4.21), whereas in the enhanced condition only once the estimated distance was significantly different from the real distance.

*Table 4.20 Mean (SD) distance estimates versus actual distance, for the three video clips in both experimental conditions.*

<table>
<thead>
<tr>
<th>Clip nr.</th>
<th>Actual Distance</th>
<th>Original VE</th>
<th>Enhanced VE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>25.15 (13.78)</td>
<td>35.23 (28.28)</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>59.71 (27.67)</td>
<td>63.14 (31.86)</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>37.92 (13.70)</td>
<td>39.36 (24.63)</td>
</tr>
</tbody>
</table>

*Table 4.21 Probability (p) values of the t-test results (test value = actual distance) for all three distance estimates*
Participants shown the enhanced simulation were more convinced that their estimates were correct, than those presented with the original version (M= 4.07, SD=0.62, versus M=3.36, SD=0.84, p=0.02).

**Speed Perception**

Egocentric movement was only simulated in the first video clip, which shows the approach of the site of the incident, seen from the driver’s position inside the company vehicle. The egocentric speed estimates for the original and enhanced VE were respectively M=83.86, SD=17.77 km/h and M=82.50, SD=9.35 km/h, when the simulated speed was set at 80 km/h. A T-test shows no significant difference between both results (p = .77).

Table 4.22 lists the (exocentric) speed estimates of traffic on the user’s *near side* of the road, which in the clips was set at 80 km per hour. A repeated measures ANOVA shows no significant difference between the two VE conditions (p = .08, two subjects were removed from the statistical analysis due to missing data). The version*video clip interaction was however significant (F( 4,96) = 2.79, p<.05). Follow up independent t-tests result in a significant difference between the two versions of the environment for video clip 2 (p < .05) and video clip 3 (p = .04), with a lower estimated speed in the enhanced condition.

The estimates in the enhanced version were in all but one case closer to the actual speed of 80 km/h.

*Table 4.22 Mean (SD) exocentric speed estimates for vehicles on the near side of the road, for all 5 video clips and both experimental conditions.*

<table>
<thead>
<tr>
<th>Clip nr.</th>
<th>Original VE</th>
<th>Enhanced VE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>81.69 (18.74)</td>
<td>84.23 (10.58)</td>
</tr>
<tr>
<td>2*</td>
<td>97.86 (15.90)</td>
<td>86.07 (13.89)</td>
</tr>
<tr>
<td>3*</td>
<td>94.64 (11.17)</td>
<td>83.57 (14.86)</td>
</tr>
<tr>
<td>4</td>
<td>89.64 (17.04)</td>
<td>85.36 (13.22)</td>
</tr>
<tr>
<td>5</td>
<td>91.07 (13.04)</td>
<td>90.00 (8.99)</td>
</tr>
<tr>
<td>Average</td>
<td>90.74 (10.94)</td>
<td>85.75 (9.42)</td>
</tr>
</tbody>
</table>

*Note: The simulated exocentric speed was 80 km/h. Estimates differed significantly between versions for clips 2 and 3 (p < .05).*

Table 4.23 lists the (exocentric) speed estimates for traffic on the user’s *far side* of the road, where the simulated speed was set at 120 km/h. A repeated measures ANOVA shows a significant difference between the versions of the VE ((F1,23 = 4.15, p<.05); (three subjects were removed from the statistical analysis due to missing data). Follow up independent t-tests result in a significant difference between the two versions for video clip 4 (p < .05). Speed estimates were lower for the enhanced VE.

*Table 4.23 Mean (SD) exocentric speed estimates for vehicles on the far side of the road, for video clips 4 and 5, and both experimental conditions.*

<table>
<thead>
<tr>
<th>Clip nr.</th>
<th>Original VE</th>
<th>Enhanced VE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4*</td>
<td>107.50 (11.22)</td>
<td>96.67 (15.57)</td>
</tr>
<tr>
<td>5</td>
<td>105.36 (15.50)</td>
<td>95.00 (15.55)</td>
</tr>
<tr>
<td>Average</td>
<td>106.43 (12.62)</td>
<td>95.8 (15.25)</td>
</tr>
</tbody>
</table>

*Note: The simulated exocentric speed was 120km/h.*
Further t-tests comparing the estimates with the simulated speed, show that mean perceived speed is overestimated for traffic on the near side of the road (test value = 80), but underestimated for traffic on the far side of the road (test value = 120), for both VE conditions. Also, mean perceived speed is closer to actual speed for traffic on the near side of the road in the enhanced VE, but for traffic on the far side of the road in the original VE. Table 4.24 and Table 4.25 list the t-test results, when mean estimates are compared to the actual traffic speed on each side of the main road.

Participants were more convinced that their estimates were correct for the enhanced VE than for the original VE (on a 0-5 six point scale: M=3.86, SD=0.66, versus M=3.29, SD=0.91, p=.04 one-tailed).

Table 4.24 Probability (p) values of t-test results (test value = 80) for all exocentric speed estimates (near side of road)

<table>
<thead>
<tr>
<th>p values of T-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clip nr.</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>Mean</td>
</tr>
</tbody>
</table>

Table 4.25 Probability (p) values of t-test results (test value = 120) for all exocentric speed estimates (far side of road)

<table>
<thead>
<tr>
<th>p values of T-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clip nr.</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>Mean</td>
</tr>
</tbody>
</table>

Assessment of danger

After every video clip, participants were asked to judge the situation for themselves, for other road users, and for the person standing in the middle of the main road near the traffic incident on a 4 point scale, from very safe to very dangerous. Results were averaged over all five video clips.

The participants judged the situation more dangerous for themselves in the manipulated version than in the original version; a repeated measures ANOVA shows a significant difference between the two versions, F (1,26) = 4.82, p = .04; a significant main effect for video clip (F(4,104) = 27.86, p<0.001) and a video clip*version interaction effect (F(4,104) = 4.35, p =.003); particularly for video clip nr 2. There was a significant between conditions in the assessment of danger.

The scores for the perceived danger for other road users show no significant difference. However, a version*video clip interaction significant effect was determined, of (F(4,100) = 4.55, p=0.02. For video clip 2 the assessment of danger to others was significantly higher in a t-test (p < .001), in the enhanced condition. A graphical representation of the estimated marginal means of the scores for danger to self and other traffic in all video clips is shown in Figure 4.27a and Figure 4.27b.
Feeling of risk
To assess the subjective feeling of risk, participants were asked whether they felt they were in a dangerous location, that the car of the highway patrol officer would get hit, that they wanted to step backwards, that they wanted to move to a safer location, or felt the traffic was moving too close for safety. The questions were different for each clip. Since the answers has a high internal consistency (Cronbach’s alpha for the 10 items was .93), all responses could be pooled. Participants felt significantly more risk in the enhanced VE than in the original VE (on a 6-point scale: M=3.95, SD=1.05, versus M=4.53, SD=0.46; p=.035 (one-tailed t-test)).

Lane crossing task
The assignment in which the participant was asked to cross the road as soon as the traffic allowed, was used to obtain additional risk estimates.

Three participants using the enhanced VE said they would never cross the road because it was too dangerous, and they were assigned the maximum risk estimate. Participants took significantly longer to cross the lane in the enhanced VE (p=.032). Because of the random variation in traffic density, crossing opportunities occurred at different moments for each participant. To correct for this effect, a new variable was created, counting the number of crossing opportunities missed, assuming at least one opportunity occurs each thirty seconds. Comparing the results for the two groups using this new variable also shows a significant difference between both VE versions (p=.045).

Other measures
Engagement, presence & appraisal of visual elements
The enhanced VE induced a significantly stronger feeling of physical presence than the original VE (p=.038, one-tailed t-test). No significant effects were observed on engagement and (general) presence. The appraisal of the environment and traffic representation did not differ significantly between the two versions of the VE.

Sound appraisal
The participants considered the sound in the enhanced version as more realistic than in the original version. A repeated measures ANOVA shows a significant difference between the two versions (F(1,26)= 7.054, p = 0.013), as well as a significant main effect for video clip (F(3,78)= 6.06, p=0.001), and a significant video clip*version interaction effect (F(3,78) = 5.16, p=0.003).

Self-assessment Manikin scores (SAM)
The arousal and valence scores of both groups were similar at the start of the experiment, and did not change significantly during the experiment. The scale was probably not sensitive enough to discover the subtle variations in this experiment, and some participants said they found it hard to understand the meaning of the SAM arousal-scale.
4.6.5 Conclusion & Discussion

Our hypothesis was that the enhanced representation of speed and depth would increase the feeling of risk, by improving the convincingness of the simulation, and evoking past experiences.

The speed estimates, averaged over all the video clips, did not differ significantly between the two groups. However, there are indications that participants in the enhanced version were able to make more accurate estimates about the distances and about speeds of vehicles that were on the same part of the road as the participant. Moreover, participants in the enhanced version felt more certain of their estimates. The estimates were closer to the actual modeled speed of 80 km per hour, whereas the estimates in the original version were higher (85.75 versus 90.75).

Higher speed and smaller distances between cars increase the chance of serious incidents on a highway. Although the perceived speed of the vehicles near the user was not higher in the enhanced version, and objects were not estimated to be closer, both the assessment of danger (for the participant in the scenario) and the feeling of risk were significantly higher. No significant results were found for judgments of risk for others except in one video clip. Personal risk estimates are more prone to change due to negative affect changes (Helweg-Larsen & Sheppard, 2001). We took the behaviour of the participants in the final task, trying to cross the road, as an indication of their personal risk estimate. The fact that participants took significantly longer to cross the road in the enhanced version may thus also be explained by a stronger feeling of risk.

These results support the theory that emotional reactions to risks can diverge from cognitive evaluations of the same risks, in this experiment the perceived danger and feeling of risk versus the judgment of speed and proximity. The improved convincingness of the representation may have stimulated the vividness of the mental images and associations that induced a feeling of risk (Loewenstein et al., 2001).

This experiment confirms our viewpoint that validity of a virtual environments for cognitive tasks such as assessing speed and distances, does not automatically lead to validity for tasks that involve the affective/emotional response of users, and vice versa. The elements in this virtual environment were modeled accurately, but (because of limitations of the display) did not have the effect on the users’ experience of danger as we assume they would in reality. Non-immersive systems like the one used in this experiment lack some of the cues (e.g. stereoscopic cues) and the wide field of view offered by more immersive systems, for example head-mounted displays, but nevertheless these systems could be a valuable tool for training and evaluation of skills. The results have shown that small but carefully selected changes to a virtual environment can influence the vividness of the representation and the experience of the user. This is of importance for the validity of training environments.

The relative strength of the types of manipulated cues was not assessed in this experiment, for several reasons. First of all, it is difficult to separate the visual and auditory cues that in real life generally occur together. Secondly, the effect of each of the cue types is probably too small to assess.

An interesting question is whether the enhanced auditory cues, which are known to evoke a sense of location (Hawkins, 2005), and which were perceived as more realistic, have added substantially to the increased feeling of risk or were even predominant. However, the sound level was deliberately kept low (since this is also the case in actual exams), and there were no indications that sound dominated the simulated experience in the experiment. Verbal comments of the participants either concerned the simulation and scenario in general, or aspects of the visual representation, but not the soundtrack. Visual cues indeed contributed to the increased feeling of risk, since different video clips with similar soundtracks yielded different results.
4.6.6 Appendix: Images of the video clips (original – enhanced)

Video clip 1

Video clip 2

Video clip 3

Video clip 4

Video clip 5

Figure 4.28 Video clips
4.7 Experiment: The effect of sounds on engagement and arousal in a virtual training

This section is based on:

4.7.1 Introduction

The design of virtual environments (VEs) has traditionally focused on the quality of the visual representation. Auditory information is frequently lacking, incomplete, or merely added in afterthought as simple soundtracks. Sounds however are of obvious importance for the perception and appraisal of our environment, and for our behaviour in an environment.

It is difficult to identify and define in general the effects of the absence of sounds in virtual environments, because the effects are situation-dependent, determined by the goal, tasks, context of use, and characteristics of users.

In three experiments we studied the effects of sounds on the affective appraisals and response of users of virtual environments. In the experiment described in Chapter 4.5, the effects of sounds (and visual dynamic effects) on the affective qualities of an environment and the engagement of the user were studied. In the experiment on the effects of depth and speed cues on the feeling of risk (Chapter 4.6), sounds were added to support the visual cues, and to improve the convincingness of the scenario. This chapter describes two related empirical field studies using a virtual scenario for firefighters. We hypothesize that the inclusion of veridical sounds that provide information that is critical for a successful task performance in an instructor in-the-loop virtual training scenario will increase the convincingness of the simulation and the engagement and arousal of its users.

In this experiment, information is added that is often omitted from virtual environments; in the framework in Chapter 3 this is specified as loss of multimodal information.

4.7.2 Related Work: sound in virtual environments

Sound design for virtual environments and games often distinguishes between ambient sounds and sound events (Serafin & Serafin, 2004). Other types of sounds are those that augment or substitute perceptions made by the other senses, communication sounds, and sounds that are not part of the virtual world. These are so-called non-diegetic sounds, such as voice-overs or interface feedback (Duridanov & Sirnoff, 2005).

A more recent and comprehensive framework for sounds in games, the IEZA framework, categorizes sounds using their function, role and properties, and explains the role of sounds and music on user’s experience of the game (Huiberts & van Tol, 2008). The framework recognizes two dimensions in game sounds, the first indicating the origin of the sounds, diegetic (from within the game world) or non-diegetic; the second the so-called expression of the game audio, from setting to activity (Figure 4.29). In the IEZA framework, Zone sounds are ambient sounds. They are diegetic and belong to the setting of the user’s actions in the VE. Effects sounds are also diegetic and are directly or indirectly triggered by the user’s actions. Sounds that are triggered by the user’s actions but are non-diegetic are Interface sounds. Finally, Affect sounds are intended to set the mood, but are outside the game environment. They include music or other sounds that indicate the emotional
state the game intends to convey, and signals upcoming events in a game. This last category has a
strong effect on the suspense in a game, and on the player’s engagement and experience in the
game. Because they do not belong to the game world, they are generally not included in virtual
training environments.

Figure 4.29 The IEZA framework (Huiberts & van Tol, 2008)

Ambient or Zone sounds (e.g., bird songs, traffic noise) can be used to set the mood or ambience
of an environment, and to affect the emotions of users. They serve to identify an environment (e.g.
forest, city) and to make the environment feel alive without drawing explicit attention.

Sound events or Effect sounds (e.g., closing doors, explosions, screams) provide information on
discrete events, actions of other people in the environment, and feedback to actions of the user in
the virtual world (Shilling & Shinn-Cunningham, 2002). They can also indicate events outside the field
of view of the user. Sound events are designed to be consciously attended to because they
potentially contain important information (thus enhancing the user’s situational awareness (Shilling,
Zyda, & Wardynski, 2002). Ambient sounds and sound events are essential to establish a feeling of
contact with and presence in an environment (Blesser & Salter, 2007, Dinh et al., 1999; Larsson &
Västfjäll, 2001; Sanders & Scorgie, 2002; Serafin & Serafin, 2004; Whitelock, Romano, Jelfs, & Brna,
2000).

Classifications of sounds such as the IEZA framework help to outline functions and expected effects
of sound types in virtual environments. However, users also have different strategies for listening to
sounds, determined by their intentions and the context. Tuuri, Mustonen, and Pirhonen (2007)
identify eight so-called listening modes with associated focus on different aspects of a sound, for
instance reflexive (such as producing a startled response to an alarm), listening for establishing the
source or cause of a sound, listening for the meaning of a sound, for its appropriateness for a
situation, or its acoustic quality. Different listening modes can operate concurrently and may result
in different individual experiences.

Another important factor related to the user experience in the design of audio for applications
in virtual reality is expectation (Chuang & Marsden, 2003). Sounds that agree with expectation
enhance the user’s experience, whereas unexpected sounds degrade the perceived quality of the
representation (Anderson, Mulligan, Goodman, & Regen, 1983; Nunez, 2007; Violon, Lavandier, &
Drake, 2002). In principle, a minimal set of expected sounds may just be as effective to engage an
observer as a fully realistic soundscape (Back & Des, 1996; Chueng, 2002). This even holds for music
that is semantically or emotionally related to the expectations created by the contents of the virtual
environment (Nunez, 2007).

In spite of the fact that sounds determine the experience of a virtual environment to a large
extent, it is difficult to identify and define general effects of their absence, because their relevance is
context dependent, and determined by the goal and tasks of the user (Tuuri, Mustonen, & Pirhonen,
Studies on virtual environments have shown that the provision of sound enhances their acceptance and perceived realism, while the absence of sound is frequently judged as a deficiency in the simulation (Rohrmann & Bishop, 2002; Rohrmann et al., 2000). The absence of sounds in a virtual environment can be far more noticeable to a user than their presence, and may draw attention to the mediated nature of the environment, thereby diminishing its convincingness (Fencott, 1999). The perceived fidelity of a virtual environment increases with both the relevance of a sound source and its fidelity for the task at hand (Bormann, 2008). If the appropriate background sounds (Zone sounds) are not included, the user may feel detached from the simulation. Task relevant sounds (Effect sounds), which are the subject of this study, increase the involvement with the visual environment (Bormann, 2008).

**Sound in virtual training**

Film industry (Chion, 1994) and in its wake game industry (Collins, 2008), devote many of their resources to develop soundtracks that evoke emotional responses and allow the user or player to be engaged, or immersed in the represented world. In contrast, the developers of serious games still focus on the visual environment (Doornbusch, 2002). A reason may be that the development of a dedicated soundtrack for a virtual training requires a considerable budget and expert knowledge, while the benefits for the users (e.g. in transfer of training) are not immediately evident.

It is known that humans learn more efficiently if they have an emotional involvement with the training scenario (McGaugh, Ferry, Vazdarjanova, & Rozendaal, 2000; Ulate, 2002). This can be achieved through a simulation that “feels real” and that may therefore trigger memories of similar situations and related emotions. Sounds can improve the ecological validity and effectiveness of virtual training scenarios by increasing the naturalness and convincingness of the experience, by providing task-relevant information, and by affecting the users emotionally. Trainees expect that the visual events they perceive in the virtual training are accompanied by sounds, just like they are in reality. Moreover, they are increasingly accustomed to computer games with high quality soundtracks and expect an equivalent experience in a virtual training. The absence of sounds is therefore experienced as unnatural (i.e. it is a distracting factor), and thus degrades the perceived quality of the simulation (Pettey, Campanella Bracken, Rubenking, Bunche, & Gress, 2010). Sounds that present relevant information in the training make the scenario more convincing. As noted in Chapter 4.6, a convincing scenario is thought to trigger memories of similar situations and related emotions, and thus awareness of the importance of performing well.

Sounds in a virtual training can convey task-relevant information on significant events on which the trainee should act. Finally, sounds can be deployed to affect the emotional response of users (Bradley & Lang, 2000). Hence, scenarios can be designed to train professionals to perform in conditions in which they will experience negative emotions, stress, and arousal. This is called “affectively intense learning” (Wilfred et al., 2004) or “stress exposure training” (Tichon, 2007), and prepares trainees to perform effectively in stressful environments (Morris, Hancock, & Shirkey, 2004).

**The effects of sounds on engagement and arousal**

Two important aspects of the affective/emotional response that determine the effectiveness of a virtual training are engagement and arousal. Engagement is a widely used concept in studies of game experience and interaction design, and currently various definitions exist (Bianchi-Berthouze, Kim, & Darshak, 2007; Garris, Ahlers, & Driskell, 2002; Jennett et al., 2008; Mallon & Webb, 2000; Rozendaal et al., 2009). Here it is defined as a state of being involved in the training, without effort, and without extrinsic motivation. It has been observed that the addition of sound to a virtual environment can significantly enhance the level of presence and engagement (Darken, Bernatovich, Lawson, & Peterson, 1999; Dekker & Champion, 2007; Dinh et al., 1999; Hendrix & Barfield, 1996; Larsson, Väljamäe, Västfjäll, & Kleiner, 2008; Larsson & Västfjäll, 2001; Larsson, Västfjäll, & Kleiner, 2001; Larsson, Västfjäll, Olsson, & Kleiner, 2007; Lessiter & Freeman, 2001), and induce arousal in the user (Dekker & Champion, 2007; Krebs, 2002; Sanders & Scorgie, 2002).
The emotional experience of users of serious games and virtual trainings, and thus their reactions and judgments, should reflect their behaviour in similar real-life conditions. Sound design for entertainment games and film typically aims to enhance the emotional experience of the viewer, by using “Affect sounds” (Huiberts & van Tol 2008), consisting of sound effects and music (e.g. Ekman, 2008; Ekman & Kajastila, 2009), or other non-diegetic sounds (e.g. Hébert et al., 2005; Nunez, 2007). In contrast, instructors and trainees using virtual environments for training generally believe that only veridical sounds induce the desired sense of urgency and impending danger. However, the fidelity of sounds is a subjective concept: in films, recordings of real events are typically experienced as unconvincing. To create the desired experience, and to live up to the user’s expectations, sounds either need to be edited extensively, or they need to be replaced by artificially produced ones that exaggerate reality (hyper-real or Foley sounds; Back & Des, 1996; Ekman, 2008; Holman, 2002). Since sound design for serious games is only beginning to emerge, little is known about the effects of diegetic and non-diegetic sounds on the perceived fidelity of a computer mediated training.

The present study

The aim of the present study is to investigate if the addition of diegetic sounds to a virtual training scenario for fire fighter commanders can enhance the engagement and arousal of the users. In a previous study on this training application, the absence of sounds was mentioned as an important shortcoming (Houtkamp & Bos, 2007). Instructors even tried to compensate for the absence of sounds during training sessions by for instance imitating the cries of victims, and the sounds of sirens. We added realistic and task-relevant sounds to a virtual fire fighter training scenario, and we performed two field studies to assess how these sounds affect user experience.

We hypothesize that the inclusion of diegetic, veridical sounds that provide information that is critical for a successful task performance will increase the convincingness of the simulation and the engagement and arousal of its users.

4.7.3 Experiment 1: realistic sounds in a virtual training environment

In the first field study we used a soundtrack developed in close collaboration with experts in firefighting and training of first responders. Afterwards some adjustments were made to the soundtrack, and a second study was conducted.

Method

Participants

The participants in the first experiment with the virtual training environment were 21 crew commanders (all males), aged between 27 and 55 years (M=42.7 years, SD=8.6). All participants had 8 to 32 years of firefighting experience (M=17.8 years, SD=7.1) and had been crew commanders for 0 to 22 years (M=8.4 years, SD=6.2). Participants were randomly assigned to each of the two conditions: 11 to the sound condition, and 10 to the no-sound condition.

Design

Two conditions were compared in a between-subjects design: the original fire fighter training scenario without sounds, and the same scenario, enhanced with sound events and ambient sounds

Equipment and set-up

The experiment was performed in a dimly lit room at the Lieszout fire station (Lieszout, The Netherlands). The virtual environment was generated using a desktop computer with a Pentium 4 processor and an ATI RADEON 2600 graphics card, and was displayed by means of a beamer on a projection screen subtending 1 m wide by 1.5 m high. The trainee was seated on a high chair that was placed at a distance of 2.5m in front of the screen. The instructor, the operator and the experimenter were all seated behind tables that were placed perpendicular to the screen, at a distance of 1.5 m from its right edge (Figure 4.30). This set-up allowed them a clear view of both the
trainee and the screen, and is similar to the setup that is typically used in actual training sessions. The trainees used a gamepad to move their avatar through the virtual environment. A 5.1 Creative surround sound system, equipped with 5 speakers and a subwoofer was used to present the audio signals. The sound level was such that the sounds of the simulation were clearly audible, while the instructor and trainee could still talk to each other. Galvanic skin response (GSR) was measured by attaching a self-made device, consisting of Velcro straps equipped with small metal plates, to a finger of the participant. The signal was read out and stored by a laptop computer at a rate of 28 Hz.

The virtual environment

The virtual scenario training environment was developed by VSTEP B.V. (www.vstep.nl) and Artesis B.V. in Quest 3D version 3.6 (www.act-3d.com). This application provides procedural fire fighter training on virtual scenarios in several environments, with or without an instructor. It includes a range of tools, equipment, and other emergency services that can be used by the trainee.

During the virtual training, the trainees identify with a character that is initially located inside a fire engine (Figure 4.31a), and later in the vicinity of a house (Figure 4.31b) or inside a house (Figure 4.31c). At that stage in the training the virtual environment represents a typical modern Dutch suburb, and contains all elements that are relevant for the scenario, such as water sources. The buildings and objects are geometrically accurate and textured. The representation of the environment is neutral, and has not been designed to increase arousal or tension. Although the characters featuring in the environment (such as firemen, police and local residents or bystanders) are easily recognizable, their representation is rather schematic. The trainee has the option to toggle between a 1st person view and a tethered view (Figure 4.32b).
Training scenario

The training scenario selected for this study represents the response to a suburban house fire. The trainee adopts the role of the fire commander. The instructor reacts verbally to requests from the trainee and impersonates all other characters that occur in the scenario (dispatch center, crew members, local residents, police, etc.). Using a separate command module, the operator responds to changes in the situation or requests from the trainee, by performing the appropriate actions in the virtual environment, such as transporting and adding vehicles and avatars, and manipulating the fire. The operator controls the events, and logs the main events and actions of the trainee.

At the start of the session, the trainee’s avatar is situated in a fully equipped fire engine at the fire station. When a dispatch call reports a fire in a house, the fire engine leaves the station on its way to the given address. During transit, the trainee’s task is to request all relevant information, to instruct the other members of the team, and to decide which actions need to be taken upon arrival. The operator lets the vehicle arrive at the location of the fire and places the trainee’s avatar outside the vehicle. From that moment on, the trainee is free to navigate in all directions, using a gamepad.

The trainee then performs the required procedures, such as gathering information about possible victims, visually locating the fire from outside the house, saving victims, determining sources of danger in the immediate vicinity, inspecting the house, assessing the extent of the fire, and taking actions such as instructing the crew and ordering a 2\textsuperscript{nd} alarm. In the virtual training scenario the fire commander’s major task is to create a mental model of the location of the fire, and the areas or objects on all six sides of the fire. To construct this mental model, called “cube”, commanders use both the information they can directly perceive, and information provided by crew.
members and other human sources. In their own words: “crew members are the eyes and ears of the commander”. In the training scenario, the trainee performs a quick outside inspection of the house before sending the crew inside. Adopting the role of the other crew members, the instructor provides the trainee with the required information about the layout of the house, the location of the fire, and any changes in the situation. The commanders dynamically update and extend their mental model of the situation with incoming information and in response to events, for instance to include or update the location of the crew members, the equipment they are using, the position of the victim during the rescue process, and the size or the spread of the fire to another room.

**Soundtrack**

Since the original environment had no soundtrack, a dedicated soundtrack had to be developed for this experiment. First, we used a storyboard to identify important locations and events in the scenario. For each scene, a group of expert fire fighters identified the required sounds. Together with the experts we determined the type of sounds (i.e. ambient sounds, event sounds or communication sounds) and the expected effect (i.e. whether the sound would induce a sense of location, provide information on an event, increase the convincingness of the scenario, and/or increase the perceived danger of the situation). Because the instructor performs a dialogue with the participant there was no need to include communication sounds in the soundtrack. The sounds we used were either selected from existing databases, or recorded at firefighting training centers. If necessary, the recorded sounds were separated from other background sounds and noise was filtered out. Each individual sound was subjected to pretests, to evaluate whether the sound itself (without its visual counterpart) was evocative of the intended environment or event, whether it conferred the intended affective qualities, and whether it was considered appropriate for the visual environment or event that it represented. Because Quest 3D does not include a physics engine, sound volume changes that should accompany actions like entering or leaving a house, or opening or closing doors, were hardcoded in the scenario. This workaround successfully simulated a physics sound engine. After implementing the soundtrack, its volume was corrected, and the overall simulation was assessed and fine-tuned with the help of expert users.

The resulting soundtrack consists of a continuous, low, background noise, which is appropriate for an urban environment. All relevant events are accompanied by sounds: the engine of the fire truck, pumps, fire, explosion, extinguishing of the fire (with water), sirens indicate the arrival of ambulances and police, and cries represent a victim in distress. In real emergencies, fire commanders wear helmets, and continuously receive messages from other emergency responders, which affect the way they perceive the other sounds. We decided to adhere to the customary training situation in which helmets are not worn, and communication with the instructor is not hindered by other, less relevant information sources.

**Measures**

To assess the arousal and engagement of the trainees, and the experienced convincingness of the training, the following measures were applied:

- **Self-reported scores on 9-point Arousal and Valence scales of the SAM (Self-Assessment Manikin; Bradley & Lang, 1994).** The Valence scale was applied to determine effects on the perceived pleasure of the experience.
- **Questions on engagement from the ITC-SOPI.** The ITC-SOPI is developed to measure users’ experiences of media, without reference to objective system parameters (Lessiter et al., 2001). It consists of four factors: engagement, sense of physical space, ecological validity, and negative effects. Only the questions on engagement were relevant and usable in this environment. We made a selection and slightly adapted some questions to the situation. To gather additional information on different aspects of the user experience related to engagement and convincingness, we added questions on Spatial Presence (the sense of being physically in the VE), and on involvement (the attention devoted to the VE and the involvement experienced) from the Igroup Presence Questionnaire (Schubert et al., 2006). Some questions were slightly adapted to the situation in this experiment.
Questions about the convincingness of the virtual representation of the scenario, the events and victim in the scenario, and the perceived danger of the situation. We define convincingness as a quality of the representation that makes a trainee perceive the scenario, the events and the victim as realistic. In the sound condition, questions were included about the perceived quality and appropriateness of the sounds.

The instructor scored the performance of trainees on a scorecard, which was based on the official score form of the Netherlands Institute for Safety (NIFV: www.nifv.nl), and which had been adjusted to this virtual training scenario.

The skin conductance of subjects was measured during the whole training session by a GSR device, as a measure of physiological arousal (Lin, Imamiya, Hu, & Omata, 2007; Mandryk & Inkpen, 2004). Unfortunately, it proved impossible to analyze the results afterwards, due to an error in the recording software.

Procedure
First, the participants filled in a general questionnaire, and scored the SAM Arousal and Valence scales. Then they were fitted with the GSR device and had a few minutes to practice navigating a simple virtual environment, using the gamepad. During this practice run, we made a base level recording of the skin conductance. The participants then performed the training scenario, which typically took about 10-12 minutes. Directly after they had finished the training, the GSR device was removed, and the second questionnaire was presented, including the SAM. Then the instructor finished his assessment of the trainee’s performance using his score cards. Finally, the instructor evaluated his assessment with the trainee. This evaluation was not part of this experiment.

4.7.4 Results of experiment 1

Arousal and valence, SAM-scores.
Table 4.26 shows the scores on the Arousal and Valence scales of the SAM, obtained directly before and after the training. For Arousal, a score of 3 to 4 means the participants expressed themselves to be rather calm (1 is very calm, 9 very excited, 5 represents the middle of the scale) both before and after the training. A score of 7 for Valence means the participant feels pleasant or happy (9 is the highest score on this scale). We performed a Mixed Two-Way ANOVA, with the SAM scores as within subject factors. There were no significant main effects of condition, Arousal or Valence difference, and there was no interaction effect. Thus, the training did not have a significant impact on the self-expressed emotional state of the trainee.

<table>
<thead>
<tr>
<th>SAM scores</th>
<th>No Sound (N=10)</th>
<th>Sound (N=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arousal</td>
<td>before 3.60 (1.58)</td>
<td>4.00 (1.34)</td>
</tr>
<tr>
<td></td>
<td>after 3.60 (1.71)</td>
<td>4.27 (1.65)</td>
</tr>
<tr>
<td>Valence</td>
<td>before 7.30 (1.57)</td>
<td>7.09 (1.58)</td>
</tr>
<tr>
<td></td>
<td>after 7.00 (1.63)</td>
<td>7.18 (1.66)</td>
</tr>
</tbody>
</table>

Engagement and related aspects of the users’ experience.
The construct “Engagement” was measured (on a seven point bipolar scale) using six questions. Cronbach’s alpha for this construct was .80. There was no significant difference (p>.05) in the mean Engagement scores between the sound (M=5.35, SD=.79) and the no-sound (M=5.33, SD=1.20) conditions.
The degree of involvement was assessed (on a seven point bipolar scale) by three questions based on the Igroup Presence questionnaire (Schubert et al., 2006). Cronbach’s alpha remained below 0.6 for this scale (possible reasons are discussed in the conclusion of this section). There was no significant difference (p > .05) in the mean involvement scores between the sound (M=5.03, SD=.91) and the no-sound (M=4.9, SD=1.02) conditions.

Spatial Presence was measured (on a seven point bipolar scale) with four questions adapted from the Igroup Presence Questionnaire. Cronbach’s alpha for this scale was .74. There was no significant difference (p > .05) in the mean Presence scores between the sound (M=5.18, SD=.73) and the no-sound (M=5.45, SD=1.16) conditions.

**Convincingness.**

Convincingness was measured using 5 questions on the representation of the victim, the environment, and the events in the scenario. The answers were scored on scales from 1 (not convincing at all) to 7 (extremely convincing). Cronbach’s alpha for this construct was .76. The mean scores for the convincingness of the elements in the scenario were higher in the condition without sounds (M= 5.47, SD=1.0) than in the condition with sounds (M=4.73, SD=.62), and this difference was weakly significant (p=.051, 2-tailed).

**Perceived danger**

The Perceived Danger for the victim and the overall seriousness of the situation were assessed through 4 questions (on a seven point bipolar scale). The total of the scores did not show a significant difference between the sound and no-sound conditions (no sound M=4.65, SD=1.19; sound M=4.84, SD=.54). The danger was thus perceived as existent, but not high, in both conditions.

**Sound quality**

In the sound condition, the trainees were fairly positive about the appropriateness of the sounds for the visual environment (M=4.82, SD=1.17 on a 7 point bipolar scale), not annoyed by the sounds (M=1.91, SD=1.14 on a 7 point bipolar scale), and they agreed slightly that the sounds provided information about the seriousness of the situation (M=4.55, SD=1.21 on a 7 point bipolar scale). When presented with two options, 90% of the participants considered the sounds good (vs. bad), 90% thought they were loud (vs. low), and all participants considered them helpful (vs. unhelpful) and realistic (vs. unrealistic). All participants had heard the sound of the fire, 82% had heard the sound of the explosion, and 55% had heard the sound of the pumps.

Only about half of the participants in the no-sound condition (6 out of 10) noticed that sounds were not provided in the scenario.

**Performance**

The performance of the trainees during the training was assessed by the instructor, using standard scorecards. There was no significant difference between the mean performance of trainees in the no-sound and the sound conditions.

### 4.7.5 Conclusions on experiment 1

Contrary to our hypothesis, we found no significant effect of sound on the arousal and engagement of the trainees. The participants in both conditions felt involved and moderately present in the VE. Although expert fire fighters judged the added sounds appropriate for the simulated events, and of good quality, the participants in the study judged the convincingness of the virtual scenario even somewhat lower. However, a thorough check revealed two potential flaws in the experimental procedure that may have influenced these results. First, from our own observations and from comments by the trainees and the instructor, we inferred that the communication between the trainee and the instructor partly masked the sounds: both the trainee and the instructor raised their voices to make themselves heard. Secondly, by coincidence, trainees in the sound condition had significantly more experience with 3D virtual environments and games than trainees in the no-sound condition (on a scale from 0-3 respectively M=1.45, SD=0.82 and
M=0.36, SD=0.67; p=.003). With an analysis of covariance with 3D virtual environment experience as covariate, to correct for this experience, we found clearly no significant difference in convincingness between both conditions (F(1,19)= 1.42, p=.25). Experienced gamers are probably accustomed to high quality soundtracks, which may have affected the present results. We therefore decided to repeat the experiment at a different location, with an enhanced soundtrack, and with participants carefully balanced with respect to game experience.

4.7.6 Experiment 2: enhanced sounds in a virtual training environment

From the results of the first experiment we concluded that some sound events probably required more emphasis to be noticed in the context of a virtual simulation. Since it is known in film industry that important sounds need to be “hyper real” (i.e. they should be played louder than normal) to stand off against other sounds and background noise (Hawkins, 2005), we adjusted the sound volume of two important (informative) sound events in the virtual training scenario from experiment 1. These were the sounds of the explosion and the screaming victim, and they were respectively increased by 400% and 200% compared to their volume in experiment 1. The volume of the other sounds was kept the same, to prevent an overall increase in background sound level during the training, which would have interfered with the communication between the instructor and the trainee.

Participants

The participants in the second experiment were 18 crew commanders (all males), aged between 29 and 51 years (M=42.9 years, SD=6.1). All participants had 0 to 22 years of firefighting experience (M=21.2 years, SD=7.8) and had been crew commanders for 0 to 22 years (M=6.3 years, SD=3.3). Because we suspected an effect of gaming experience on the results in experiment 1, we divided participants with little and extensive experience with gaming and 3D environments equally over the sound and no-sound conditions, resulting in 9 participants in both conditions.

Materials and procedure

This experiment was performed in a dimly lit room at the Alblasserdam fire station (Alblasserdam, The Netherlands). The acoustics at this location were better than at the location of experiment 1. As a result, the overall audibility of the soundtrack was better than in experiment 1.

In the first experiment, participants answered questions about the perceived quality of the sounds, and their appropriateness. In this experiment we included additional questions to assess the importance of sound as a source of information, relative to other information sources that are available to the trainees. We asked the trainees from which sources they derived the information to construct their mental model of the “cube”: from respectively the virtual environment that was shown on screen, the sounds, the crew, the dispatch center, or their previous knowledge and experience. The answers were given in percentages, totaling 100%.

We also asked the participants to what extent they used these sources in assessing the dangerous episodes in the scenario. By adding these questions we hoped to understand the role of sounds in the actions and decisions of the fire commanders. The other measures and materials, and the experimental design, were the same as in experiment 1.
4.7.7 Results of experiment 2

Arousal and Valence (SAM-scores)

The scores on the arousal and valence scales of the SAM were obtained directly before and after the training. We performed a Mixed Two-Way ANOVA, with the SAM scores as within subjects factor. It showed no significant main effects of condition, of the arousal or valence difference, nor an interaction effect (p > .05)(Table 4.27).

Table 4.27 Mean (and standard deviation) scores on the SAM Arousal and Valence scales, directly before and after the training.

<table>
<thead>
<tr>
<th>SAM scores</th>
<th>No Sound (N=9)</th>
<th>Sound (N=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>arousal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>before</td>
<td>3.33 (1.00)</td>
<td>3.67 (1.94)</td>
</tr>
<tr>
<td>after</td>
<td>3.22 (.83)</td>
<td>3.89 (1.62)</td>
</tr>
<tr>
<td>valence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>before</td>
<td>7.11 (1.45)</td>
<td>6.89 (1.76)</td>
</tr>
<tr>
<td>after</td>
<td>7.11 (1.45)</td>
<td>7.67 (.71)</td>
</tr>
</tbody>
</table>

Engagement and related aspects of the users’ experience

As in experiment 1, the reliability was satisfactory for engagement and spatial presence, but low for involvement.

There was no significant difference (p=.79) in the mean engagement scores between the sound (M=5.47, SD=1.10) and the no-sound (M=5.35, SD=.69) conditions.

There was also no significant difference (p=.56) in the mean involvement scores between the sound (M=5.33, SD=1.15) and the no-sound (M=5.04, SD=.94) conditions.

Finally, there was no significant difference (p=.56) in the mean Spatial Presence scores between the sound (M=4.69, SD=.86) and the no-sound (M=4.53, SD=1.09) conditions.

Convincingness

Again, the mean of the scores for the convincingness of the elements in the scenario were higher in the condition without sounds (M=6.18, SD=.76) than in the condition with sounds (M=5.51, SD=.77). However, this time the difference was not significant (p=.085, 2-tailed). Cronbach’s alpha was satisfactory: α = 0.63.

Sound quality

Again only half of the participants in the condition without sound noticed that sounds were not provided in the scenario (4 out of 9 trainees). In the sound condition, the trainees were positive on the appropriateness of the sounds for the visual environment (M=5.78, SD= 1.09), were somewhat annoyed by the sounds (M=2.78, SD=1.64) and they felt the sounds provided them with information on the seriousness of the situation (M=5.33, SD=1.12, all on a 7 point bipolar scale).
with two options, all trainees (N=8, 1 missing value) considered the sounds to be good (versus bad), loud (versus low) and realistic (versus unrealistic), and all (N=9) considered them to be helpful (versus unhelpful). All had noticed the sound of the fire, the sound of the explosion, and 66.7 % had heard the sound of the pumps.

**Perceived danger**

The Perceived Danger for the victim and the overall seriousness of the situation were assessed through 5 questions. Cronbach’s alpha for this scale was .81. The total mean of the scores did not show a significant difference (no-sound M=5.04, SD=0.65; sound M=5.2, SD=1.60). The danger was thus perceived as existent, but not very high, in both conditions. The Perceived Danger for the crew did not differ significantly between conditions (no sound M=3.56, SD=1.67; sound M=4.11, SD=2.26).

**Performance**

The performance during the training was assessed by the instructor, using standard score cards. Again, the mean performance scores in the condition without sound (M=6.96, SD=.51) and the condition with sound (M=7.11, SD=.69) were not significantly different (p=.62, 2-tailed).

**Use of information**

We asked the trainees from which sources they derived the information to construct their mental model of the “cube”: from respectively the virtual environment that was shown on screen, the sounds, the crew, the dispatch center, or their previous knowledge and experience. In both conditions, the trainees estimated that they derived less than half of the information from the virtual environment; and almost the same fraction from the crew members and dispatch center (i.e. the instructor; Figure 4.34 and Table 4.28).

Knowledge and experience are also important information sources. Sound only accounted for 7 % of the information sources. The variation in the results also shows that the commanders use different strategies.

We also asked to what extent the trainees used these sources to assess dangerous episodes in the scenario; now the sounds were said to be important. A paired samples t-test showed a significant difference between the role of sounds as information for the “cube” and for the assessment of dangerous episodes (t(8)= -2.15; p = .03 (1 tailed)). Sound was considered an important source for assessing danger, second after the virtual environment. Again, the virtual environment itself accounts for only a third of the information (Figure 4.35 and Table 4.29).

![Figure 4.34 Information sources used by trainees to complete their mental model of the “cube”, in the no-sound (a) and sound (b) conditions (N=9). The scores represent percentages, adding up to approx. 100% for each participant (N=9).](image-url)
Table 4.28 Information sources used by trainees to complete their mental model of the “cube”, in the no-sound and sound conditions (N=9).

<table>
<thead>
<tr>
<th></th>
<th>No Sound M (SD) (%)</th>
<th>Sound M (SD) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual information from the VE</td>
<td>42.67 (20.0)</td>
<td>31.11 (12.69)</td>
</tr>
<tr>
<td>Sounds</td>
<td>1.11 (3.33)</td>
<td>6.67 (4.33)</td>
</tr>
<tr>
<td>Crew</td>
<td>31.67 (23.18)</td>
<td>33.33 (17.14)</td>
</tr>
<tr>
<td>Dispatch center</td>
<td>6.22 (6.87)</td>
<td>7.44 (4.04)</td>
</tr>
<tr>
<td>Knowledge / experience</td>
<td>18.33 (16.77)</td>
<td>21.67 (12.25)</td>
</tr>
</tbody>
</table>

Note: The scores represent percentages, adding up to approx. 100% for each participant (N=9).

Figure 4.35 Information sources used by trainees to assess dangerous episodes in the scenario, in the no-sound (a) (N=8) and the sound (b) (N=9) conditions. The scores represent percentages, adding up to approx. 100% for each participant.

Table 4.29 Information sources used by trainees to assess dangerous episodes in the scenario, in the no-sound (N=8) and the sound (N=9) conditions.

<table>
<thead>
<tr>
<th></th>
<th>No Sound M (SD) (%)</th>
<th>Sound M (SD) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual information from the VE</td>
<td>36.25 (34.20)</td>
<td>38.89 (28.48)</td>
</tr>
<tr>
<td>Sounds</td>
<td>.00 (.00)</td>
<td>28.89 (29.35)</td>
</tr>
<tr>
<td>Crew</td>
<td>30.00 (25.07)</td>
<td>17.78 (16.42)</td>
</tr>
<tr>
<td>Dispatch center</td>
<td>6.88 (17.51)</td>
<td>1.67 (3.54)</td>
</tr>
<tr>
<td>Knowledge / experience</td>
<td>31.88 (31.84)</td>
<td>12.78 (13.49)</td>
</tr>
</tbody>
</table>

Note: The scores represent percentages, adding up to approx. 100% for each participant.

4.7.8 Conclusions on experiment 2

The results of the experiment 2 confirmed the findings of experiment 1: a soundtrack with realistic ambient and event related sounds did not increase the engagement and arousal of trainees in this particular virtual training. In contrast to the realistic soundtrack used in experiment 1, the enhanced soundtrack did not lower (but also did not enhance) the convincingness of the virtual environment.

Sound appears to be more relevant for the assessment of danger, which is a task with a strong affective component, than for constructing the “cube”, which is a procedural and cognitive task. This
result agrees with the theory that emotional reactions to risks can diverge from cognitive evaluations of the same risks (Loewenstein et al., 2001; Slovic et al., 2004).

4.7.9 Conclusions and Discussion

The addition of a realistic soundtrack with emphasized task relevant sounds to a virtual fire fighter training scenario did not increase the arousal and engagement of the users. It is unlikely that the quality of the sounds or their implementation were insufficient, since the soundtrack was carefully designed and tested with the help of expert fire fighters, and the trainees judged the sounds as appropriate and adequate. The participants in the sound conditions were aware of the sounds, and were positive about the quality of the soundtrack.

The commanders in experiment 2 alleged that sounds were an important source for assessing dangerous episodes in the scenario, which is a task with a strong affective component. For constructing the “cube”, which is a procedural and cognitive task, sounds were considered far less important. This result agrees with the theory that emotional reactions to risks can diverge from cognitive evaluations of the same risks (Loewenstein et al., 2001; Slovic et al., 2004). Sound may stimulate mental images and associations which induce a feeling of risk (Slovic et al., 2004), and thereby determine the overall assessment of danger. This finding is also supported by the theory of listening modes of Tuuri et al. (2007), which holds that users have different listening strategies with associated focus on different aspects of a sound. We conclude that sounds did have some effect on the affective response of the trainees.

Nonetheless, the sounds used in the experiments did not induce an emotional change in the users strong enough to be measured with the scales used here. From our own observations of the training program for fire fighters over a longer period of time, and from discussions with expert fire fighters, we conclude that the sounds in this study probably did not have an effect on arousal or engagement, because the trainees were not fully involved with the virtual world, which is a prerequisite for engagement. One source of distraction from the virtual environment is the task performed in the training. In addition to visual and auditory information, commanders heavily rely on information provided by fellow crew members and other human sources to construct their mental model. The trainees probably focused on the information verbally provided by the instructor, and therefore paid less attention to the events that were displayed and indicated by the sounds. For instance, when the operator and the virtual simulation provided contradictory information (which sometimes happened when the operator made a slight mistake in the scenario), the trainee always relied on the information provided by the crew (i.e. the instructor), and ignored the events that were displayed on the screen or indicated through the soundtrack. About half of the trainees in the soundless condition did not even notice the absence of sounds. They apparently completed their mental model of the events using other information sources. Support for this argument is supplied by the importance which trainees attributed to the crew and the dispatch center, and to their experience, as information sources for completing their mental model (experiment 2). In addition, the problems that some trainees experienced with navigation and the use of the controller may also have diminished their engagement with the virtual world (Lessiter et al., 2001).

We expected that sound would raise the convincingness of the virtual training and thus increase the arousal of the trainees. However, the convincingness was judged lower in the condition with sound in experiment 1. Even if this had not been the case, the effect of sounds on the arousal of the trainees in this scenario would be small and difficult to measure on the scales that were used, such as the SAM. The scenario presented was not perceived as complex or dangerous, and the trainees were mostly in control of the situation. Fire fighters are generally self-assured individuals who do not easily get nervous or show strong emotional response to danger. Also, they are used to frequent training sessions in various forms, using photos, mockups, or setups in training centers. Measuring small effects on their emotional state will therefore require very sensitive scales or metrics. The GSR-response might have revealed fluctuations in their emotional stated during the training, but unfortunately the measurements were unreliable due to technical difficulties.

129
Factors related to user characteristics and content of the training may also have contributed to the unexpected results on the assessment of the convincingness of the virtual environment. Firstly, as we noted before, sound effects in entertainment games and films often need to be exaggerated (made hyper-real) in order to be noticed or be perceived as “real” or convincing. A similar effect may occur in virtual trainings. Thus, trainees experiencing a virtual scenario may expect similar “hyper-real” sounds, or may not even be aware of sounds when they are not accentuated. This may explain why our soundtrack did not raise the convincingness of the virtual training. Secondly, in real incidents, communication over the radio, directed to the commander and to the other professionals involved, is an important part of the “soundscape”. In those conditions, several voices may be heard simultaneously and indistinctly, and the commander must filter out the information that is relevant for his task. In the training, the auditory environment is much simpler, since the instructor is the only one speaking, and offers the information clearly audible and sequentially. By introducing event related sounds on the one hand, but omitting communication sounds on the other hand, the trainees may have become aware of this abnormality.

We conclude that verisimilar, diegetic sounds do not affect the arousal and engagement of users of a virtual training when distracting factors (like the communication with an instructor, problems with navigation, or operator errors) prevent them from focusing their full attention on the virtual simulation. Hyper-real (exaggerated) visual and auditory effects may be required to overcome the distractions provided by the training context and to achieve “affectively intense training”. However, enhanced sound effects and non-realistic sounds cannot be deployed when accurate audio-visual representations of incidents are essential for the training scenario. Returning to the IEZA model introduced earlier, the question is whether sounds in the “Affect”-category may be introduced in a more subtle way; for instance, a radio or television in the virtual environment may play music or generate fighting sounds to create the desired ambience. Another option may be to integrate the communication with the instructor in the virtual soundscape.

In the development of virtual training scenarios, individual and explicit requirements should therefore be made for the learning goals and for the desired affective/emotional response of the trainees. These allow the designer to establish and augment relevant sounds (Hoorn, Konijn, & van der Veer, 2003) that address the different listening modes involved (Tuuri et al., 2007). Dependent on the relative importance of the affective/emotional response of the trainees, these requirements may lead to the identification of different types of training environments.
4.8 Experiment: The influence of cybersickness of the affective appraisal of a virtual environment

This section is based on:

4.8.1 Introduction

In the experiments described in the previous paragraphs we assessed effects of manipulations of the content of the virtual environment, or of the representation. A third category of factors that may influence the response of the viewer included in the framework of determinants are individual characteristics (Figure 3.3). Individual characteristics include pre-existing mental representations, expectations or media schemata, and mood at the time of viewing. We hypothesize that cybersickness influences the mood of the viewer, and is thereby one of the factors that can influence the affective appraisal of a virtual environment. According to Russell’s attribution hypothesis (Russell, 2003), the perception of the affective qualities of an environment can be influenced by a person’s affective state. If cybersickness influences the affective appraisal of the virtual environment, it may restrict the validity and use of visualisations of manmade and natural environments.

The term cybersickness was originally used for the phenomenon of people getting sick from actively playing computer games or passively watching them (McCaugley & Sharkey, 1992). It is a form of visually induced motion sickness, comparable to cinerama sickness and simulator sickness (Emmerik et al., 2011). We conducted an experiment to assess the effects of cybersickness on the affective appraisal of a virtual environment that was developed for this experiment.

4.8.2 Related work

Cybersickness is a visually induced motion sickness which is due to the observer’s perception of self-motion without actual motion (Bos et al., 2007; Kennedy, Lanham, Drexler & Essex, 1997). The main symptoms include nausea, oculomotor disturbances such as eye strain, disorientation, vertigo, postural instability, and even vomiting (LaViola, 2000; Nichols & Patel, 2002). These effects may linger after the interaction with the VE has concluded (Nichols & Patel, 2002; Stanney, Kennedy, Drexler, & Harm, 1999). For many years driving, flying and sailing simulators have been known to cause sickness, afflicting up to more than 70-80% of the users (Sharples, Cobb, Moody, & Wilson, 2008; Stanney et al., 1998), which negatively affects the effectiveness of these trainings. Recently,
cybersickness with desktop and console systems has become an important issue, also due to the increasing use of these systems for serious gaming applications (Merhi, Faugloire, Flanagan, & Stoffregen, 2007; Sharples et al., 2008).

Russell’s attribution hypothesis holds that the perception of the affective qualities of an environment can be influenced by a person’s affective state (Russell, 2003)(see also Chapter 2.4.2). Core affect is object free (free-floating), but through attribution can become directed at an object. If a person attributes affective qualities to an environment by the shift in mood he or she experiences, but the mood change is not due to environmental properties, then this could lead to a misattribution (Schwarz, 1990). For example, if a person’s mood is improved, for instance when given a free gift, his or her liking for the environment also increases (Russell & Snodgrass, 1987).

Physical discomfort, for instance caused by cybersickness, may influence mood (e.g. Baron & Bell, 1975, Lecci, & Wirth, 2001). Thus, VE users suffering from cybersickness may experience the simulated environment as less pleasant, and possibly also as more distressing, than users who are not afflicted. In this study we investigated the effects of cybersickness on the affective appraisal of an urban environment simulated on a desktop computer system.

Until now no studies have been performed that specifically link the physical wellbeing of a person to the perceived affective qualities of an environment. However, it is well known that affective appraisals are influenced by environmental stressors (Holahan, 1986). For instance, helicopter noise was found to negatively influence the perceived aesthetic qualities of the Grand Canyon (Mace, Bell, & Loomis, 1999), crowding in store environments to diminish the affective experience of shopping (Machleit, Kellaris, & Eroglu, 1994), malodor to negatively affect the appraisal of objects and places (Rotton, 1983; Wrzesniewski, McCauley, & Rozin, 1999), and sleep deprivation to negatively influence the affective appraisal of images (Tempesta et al., 2010).

Although these environmental stressors are all external factors, and cybersickness may be considered an internal stressor, it seems likely that a person’s internal affective state could influence the appraised affective qualities of an environment through misattribution (Russell, 2003; Schwarz & Clore, 1983). Therefore, we hypothesize that users suffering from cybersickness may misattribute their unpleasant feelings to the affective qualities of the VE, and may therefore appraise it as being more unpleasant.

4.8.3 Method

Participants

A total of 32 persons (15 males and 17 females) participated in the experiment in return for monetary compensation. Two persons (a 68 years old man and a 47 years old woman) had to forfeit prematurely because they surpassed the pre-set level of cybersickness that was adopted as an upper limit in this experiment. The age of the remaining 30 participants ranged from 20 to 68 years, with an average of 45.8 years, and a standard deviation of 15.2 years. However, as a potential confounding factor, the ages are not evenly distributed, with participants being either college students, or normally distributed around the age of 50 years.

All participants were healthy and had no significant susceptibility to motion sickness and related phenomena, as determined through self-report by means of a health questionnaire. Only 6 participants had some experience with first person shooter (FPS) games. Two of those had some experience with Half-Life 2, but none of them recognized any of the scenes shown in the course of the experiment.

The experimental protocol was approved by TNO Human Factors internal review board on experiments with human participants. The participants gave their written informed consent prior to testing.

Apparatus

The experiment was conducted using two different computers. One was a Dell XPS700 desktop pc with an Intel Core2 Duo 6600 chipset, running at 2.4 GHz and with 2GB of RAM. The video card was a GeForce 7950 GX2 in dual SLi mode, with 1024MB of RAM. The other computer was a Dell
XPS600, with an Intel Pentium 4 processor, running at 3.6 GHz and containing 1 GB of internal RAM. This computer’s video card was a GeForce 7800GTX in dual SLI mode, with 256 MB RAM. Both computers used Toshiba TDP-P6 beamer to back-project the generated virtual environment on a semi-transparent (“frosted”) screen 1.46 by 1.09 meters in dimensions, at a resolution of 1024 by 768 pixels, with a screen refresh rate of 75 Hz.

**Viewing conditions**

The experiment was performed in a dimly lit room, using two similar set-ups separated by black curtains. Participants were seated in a comfortable chair with their head against a raised headrest, at approximately 1.2 meters viewing distance from the projection screen, resulting in a geometrical field of view of 60 degrees. Their visual straight ahead was aligned with the centre of the screen. The few participants that got highly nauseous were allowed to sit slightly further away from the screen, as this probably would not significantly influence the mental representation of the virtual world.

**Stimuli**

Subjects viewed a simulated walk through a virtual environment. The virtual environment was generated with the graphics engine of the game Half-Life 2 (the “Source” engine) and comprises rural, mostly coastline, areas, an industrial vista and an urban area, depicting a city with mixed European architecture (Figure 4.36). Due to the use of a contemporary game engine, the graphics are up to modern day standards and shows a variety of detailed textures, and high quality lighting effects.
The navigation through the virtual space was passive, i.e. without user control, in order to assure that every person had the same impression of the environment. The simulated walk started at a beach and ended in a city, and lasted about 12 minutes. At the end of the first walkthrough, the presentation started over and ran for another five minutes before stopping at a designated location in the city. The participant was then asked to appraise the environment displayed at that moment.

The virtual walk took place at a simulated speed of 13.6 km/h. Our aim was to induce cybersickness without hindering the formation of an affective appraisal of the environment. To increase the imagery’s potential to provoke sickness, we used a large geometrical field-of-view (60°) in combination with an erratic swaying camera motion about an axis at foot level and with a peak amplitude of 8°. These parameter settings were likely to induce cybersickness within a relatively small amount of time (Vries, Bos, Emmerik, & Groen, 2007), before participants would get bored (a mood which might confound the affective appraisal).

Previous research indicated that motion sickness related symptoms are most prevalent when either a roll or pitch motion is performed (Lo & So, 2001). In this experiment we used a roll motion since this ensures less distortion of the scenery to be appraised (see Figure 4.37 for a pictorial representation of the roll axis). However, during a pilot experiment it became apparent that participants were able to accustom to the roll motion when a regular sinusoid was implemented. Therefore an unpredictable complex harmonic swaying motion was used, further explained in Toet et al. (2010).

![Figure 4.37 Pictorial representation of the roll axis.](image)

**Measurement scales**

Most studies on motion, simulator and cybersickness use the Simulator Sickness Questionnaire (SSQ) (So, Ho, & Lo, 2001), a subjective self-report scale first introduced by Kennedy and Lane (Kennedy, Lane, Berbaum, & Lilienthal, 1993). However, the SSQ, while giving an accurate depiction of cybersickness, is a rather long questionnaire. As the affective appraisal questionnaire used in this study is also comprehensive, we feared that cybersickness could subside over the course of the final appraisal task. Furthermore, this is an exploratory research study and the effect of cybersickness was deemed more important than the cybersickness itself. We therefore decided to use a short 0-10 index score scale, the Misery Scale (MISC), for the evaluation of cybersickness (Bos et al., 2005). The MISC scores range from no symptoms (0), undeterminable symptoms (1) and slight determinable symptoms (2) gradually upwards to vomiting (10). We adopted a MISC score of 8 as a break-off criterion, i.e. subjects who reported a MISC score of 8 or higher were allowed to stop.
The affective appraisal of the virtual environment was rated using a slightly modified version of the eight point semantic differential scale developed by Russell and Pratt (Russell & Pratt, 1980). This scale reflects the circumplex model of affect (Russell, 2003) and describes the affective qualities of a molar environment as a 2D space, with the independent dimensions unpleasant-pleasant and arousing-sleepy on the two main axes. Diagonally through these axes are the dimensions exciting-gloomy and distressing-relaxing, which can optionally be inferred from the two main axes. It does not contain adjectives describing salient physical attributes of the environment, such as the SMB-scale (Chapter 3.5). Participants rate a list of 40 affective adjectives which are then grouped into the two dimensions comprising the appraisal.

**Procedure**

First, participants read and signed an informed consent.

Second, they filled in a questionnaire that assessed (a) their physical and mental state, (b) their past susceptibility to motion sickness and related phenomena, and (c) their proficiency with virtual environments and computer games, first person shooters and Half-Life 2 in particular. This latter information is important, since participants who have previously played Half-Life 2 may recall set pieces from the original game that were used in the construction of the virtual environment for the present study, and may associate these pieces with environments they previously encountered in the Half-Life 2 game. Such associations could in turn influence their affective appraisals of the environments shown in the present study. Only two participants reported having played Half-Life 2, but when prompted afterwards both stated they had not recognized any locations in the course of the experiment.

Third, participants rated the affective qualities of a picture of typical Dutch Neo-Classical / Neo-Renaissance style canal houses, and a picture of a busy New York street, on Russell and Pratt’s semantic differential scale (Russell & Pratt, 1980). This was done to accustom the participants with the measurement scale, and to ascertain the participant’s affective appraisal of these types of cityscape.

After completing these questionnaires, the participants were taken to the darkened experiment room and seated in the setup. There they watched the tour of the virtual environment. During this presentation they were regularly prompted to note down their MISC scores. At the end of the tour the camera froze on the area depicted in Figure 4.36d. Then, the lights in the room were turned on dimly, and the participants appraised this particular part of the environment with Russell and Pratt’s semantic differential scale.

**Hypothesis**

Prior to the experiment nearly all participants found an exemplary Neo-Renaissance style architectural cityscape to be pleasant and, for the greater part, slightly sleepy (Figure 4.38). Given the fact that cybersickness causes physical discomfort, we hypothesized that cybersick participants would find this type of environment less pleasant than participants who did not suffer from cybersickness.
Figure 4.38 Scatterplot with 95% confidence ellipse detailing the affective appraisals of a picture of Neo-Renaissance buildings

4.8.4 Results

**MISC (cybersickness)**

30 Participants finished the experiment. Of these, 7 reported no effect at all on the MISC during the whole exposure, and 13 reported feeling fine at the end. The last MISC report was used as the criterion for the following statistical analysis, as it most accurately described the participant’s wellbeing during the appraisal task. 16 Participants reported different degrees of cybersickness, although the MISC scores they reported were unevenly distributed over the spectrum, with 19 participants scoring a 0 or a 1, and a nearly even distribution over the remaining MISC scores. The small number of participants per MISC score meant that an accurate decomposition of the effect of different degrees of cybersickness on the affective appraisal of virtual environments was problematic. This was further complicated by the fact that participants could stop after reporting an 8, while the onset of nausea, which would hypothetically lead to the highest change in affective appraisal, doesn’t start until level 6. Although a simple multivariate regression did return a significant result, the small amount of participants that finished the experiment with cybersickness and the apparent high variation per MISC score, gives the model a low fidelity. Therefore the participants experiencing cybersickness were pooled into a single group, which makes for a more robust comparison. The MISC score can be divided into three categories: participants experiencing no effect (0), participants experiencing undeterminable symptoms (1), and participants experiencing determinable symptoms in varying amounts of severity (2-10). As determinable cybersickness is onset from a score of 2 or higher, this semantic divide was chosen as the demarcation criterion (incidentally nobody scored a two, hence the de facto divide is 0-1 and 3 and higher). Consequently two groups were formed, one consisting of 19 participants who were considered physically well, and one consisting of 11 participants, who were considered cybersick.
**Appraisal of the environment**

The scale of Russell and Pratt (Russell & Pratt, 1980) describes the environment on four dimensions: pleasant-unpleasant, arousing-sleepy, distressing-relaxing and exciting-gloomy. The scores on these dimensions are computed by rating individual adjectives, for instance “tranquil” or “repulsive”, on their applicability to the environment. Russell and Pratt have further computed the intercorrelations of the dimensions, resulting in the bipolar space in Figure 2.3. A person’s attribution of affective quality to an environment can be assigned to a single point in that space. Figure 4.39 shows the affective appraisals of the target area of Figure 4.36d for each of the two groups of participants (cybersick and not-cybersick).

![Figure 4.39 Scatterplot detailing the affective appraisals of the target area of Figure 4.36d.](image)

**Figure 4.39 Scatterplot detailing the affective appraisals of the target area of Figure 4.36d.**

![Figure 4.40 The effect of cybersickness on the affective appraisal.](image)

**Figure 4.40 The effect of cybersickness on the affective appraisal.**
To compare the results for these two groups we performed a MANOVA, with the cybersickness grouping as independent variable and the four different affective dimensions as dependent variables. The results are shown in Figure 4.40, where the affective dimensions differ significantly as a result of the cybersickness: $N = 30, \text{Wilks}'\lambda = 0.52, F(2,25) = 5.723, p < 0.003$. In this plot, a high score on the arousing-sleepy scale corresponds to a high level of arousal, while a low score indicates a low arousal (or highly sleepy). Likewise, a high score on the pleasant-unpleasant dimension corresponds to a high level of attributed pleasantness; a high score on the exciting-gloomy dimension indicates the participant found the environment exciting; and scoring high on the distressing-relaxing scale means the environment is distressing. Figure 4.40 shows even more change than was initially hypothesized. Participants who are cybersick ostensibly not only find the environment less pleasant, but also more arousing, and possibly more distressing.

Examining these scores with a post-hoc Tukey test reveals that the significant result of the four dimensions combined is mostly due to a significant effect on the pleasant-unpleasant dimension ($p < 0.05$), while the effect on the arousing/sleepy scale barely reaches significance ($p = 0.051$). Differences in the exciting/gloomy and distressing-relaxing scale are not significant however, so the slope of the regression line representing distressing-relaxing should be regarded with caution.

### 4.8.5 Conclusions and Discussion

In our experiment, participants who experienced cybersickness valued the virtual environment as less pleasant and more arousing, and possibly also as more distressing than participants who did not experience any appreciable symptoms of cybersickness.

It is well known that cybersickness causes physical discomfort. In this study cybersickness was induced by camera movement, and not by the cognitive or affective qualities of the buildings represented in the environment. In the pre-test, all participants had a preference for the Neo-Classical / Neo- Renaissance style buildings to be appraised. Furthermore, all viewers watched the same simulated walk through the environment, eliminating the possibility of different experiences other than those brought about by cybersickness. Therefore, whenever a participant was asked to affectively appraise the environment, his or her own change in affective state influenced the appraisal.

How cybersickness actually changes the affective appraisal of the virtual environment is not yet evident. We suggest that the most plausible explanation for this is that the change in core affect the participant experienced while watching the virtual tour was misattributed to the environment itself (Payne, Hall, Cameron, & Bishara, 2010; Russell, 2003). This result implies that cybersickness may compromise the effectiveness of simulations used for training, education and serious gaming applications, especially in combination with evidence that cybersickness may negatively affect presence (Nichols & Patel, 2002). Finding and eliminating the causes of cybersickness is an important factor in trying to construct the right ambience of a virtual environment.
4.9 Experiment: effects of stress and darkness on the appraisal of a virtual environment

This section is based on:

4.9.1 Introduction

This study investigated whether users appraise the nighttime representation of a virtual environment as less pleasant, more arousing, and more anxiogenic, than its daytime version, particularly after experiencing acute stress.

Very little is known of the validity of virtual environments for representing nighttime conditions, whether used for assessing the affective qualities of environments, or as training environments. Desktop VEs are increasingly deployed in training personnel to operate in dangerous or critical situations, where stress, and emotions such as anxiety, may influence performance. VE’s should be able to simulate a variety of circumstances, also nighttime conditions, when lighting conditions are unfavourable.

Stress is defined as a process whereby environmental demands evoke an appraisal process in which perceived demands exceed capacity, resulting in undesirable physiological, psychological, behavioural, or social outcomes (Driskell, Salas, Johnston, & Wollert, 2008). State anxiety is an emotional reaction “characterized by subjective, consciously perceived feelings of tension and apprehension, and heightened autonomic nervous system activity” (Spielberger, Gorsuch, & Lushene, 1970).

The effectiveness of desktop VEs for such training purposes depends critically on their ability to correctly address the user’s emotional, cognitive and perceptual experience (Gratch & Marsella, 2003). Emotionally compelling VE’s are thought to provide more effective training (Gratch & Marsella, 2003, Wilfred et al., 2004) and to result in a higher degree of initial learning and subsequent retention of the lessons learned (Shilling et al., 2002).

The independent variables researched in this experiment concern the content of the virtual environment (day or night); and the emotional state of the user (in this case stress) as a response moderator. Response moderators are characteristics of the individual that may influence the response to the virtual environment, and are included in the framework in Chapter 3 (Figure 3.3). We assessed the affective appraisal of the environment, and the emotional response of users, especially their anxiety, using verbal scales and questionnaires. We hypothesized that a nighttime version desktop VE would be assessed as less pleasant and more arousing than the daytime VE, and that a high stress-level would augment these effects and elicit anxiety in the participants.

4.9.2 Related work

Darkness is a phylogenetically relevant aversive context for diurnal creatures like humans. It connotes feelings of personal vulnerability to actual physical danger, and may automatically precipitate emotional responses consonant with those thoughts. Whereas ambient light may be reassuring, ambient darkness elicits fear by concealing potential dangers (Blöbaum & Hunecke,
2005, Nasar & Jones, 1997). Evidence for the anxiogenic nature of darkness is provided by the fact that the human startle reflex increases in the dark (Grillon, Pellowski, Merikangas, & Davis, 1997; Grillon et al., 1999). The innate fear for darkness which most people have also extrapolates to ecologically valid immersive virtual immersive environments (Mühlberger, Wieser, & Pauli, 2008). However, not much is known about the effect of the level of ambient lighting on the affective appraisals of a virtual environment on a desktop VE. Bishop and Rohrman (2003) reported to have found distinct similarities between appraisals of computer generated simulations of a suburban model and the real environment, in daytime and nighttime conditions. In commercial desktop games low-key lighting is sometimes deployed to evoke suspense and dread (Niedenthal, 2005). It has also been shown recently that color of simulated illumination in desktop VEs determines the user’s affect and performance (Knez & Niedenthal, 2008).

Because the affective appraisal process also has a cognitive component, a person’s internal affective state may in turn influence the perception of the affective qualities of an environment through misattribution (Schwarz & Clore, 1983; Russell, 2003; Clore & Huntsinger, 2007; Schwarz, 1990). People are inclined to make cognitive appraisals of unrelated topics and objects reflecting their affective state (e.g. induced by watching movies, experiencing sunny weather, or experiencing stressful exams) (Clore & Huntsinger, 2007; Lerner & Keltner, 2000; Payne et al, 2010). Also, they tend to attribute residual arousal from prior situations to external cues in subsequent situations (Zillmann 1971; Mattes & Cantor, 1982). Thus, people may also misattribute their own feelings to the affective qualities of VEs. Evidence for this hypothesis is our recent finding that users suffering from cybersickness experience a simulated environment as less pleasant and more arousing than users who are not afflicted (Chapter 4.8, Spek, Bos, Emmerik, Toet, & Houtkamp, 2007; Spek & Houtkamp, 2008).

Darkness induced anxiety is enhanced by prior exposure to a social stressor (Grillon, Duncko, Covington, Kopperman, & Kling, 2007). Anxiety generates a hyper-vigilance for threatening stimuli (Eysenck, 1992) induces selective attentional bias for anxiogenic cues (Mathews & MacLeod, 1986, Öhman, Flykt, & Esteves, 2001), causes a focusing on cues that support the subjective feeling of threat (Easterbrook, 1959) and impairs auditory perception (Simoens, Istok, Hyttinen, Hirvonen, Näätänen, Tervaniemi, 2007), thus reducing the intake of cues and information upon which to make accurate appraisals and attributions.

4.9.3 Hypotheses

We designed an experiment using a desktop virtual environment to assess whether darkness in a virtual environment has similar effects on viewers as darkness in reality. We included stress as a possible response moderator. Based on the theory described above we hypothesized that

1. participants would appraise a nighttime version of the desktop VE as less pleasant, less sleepy, and as more unpleasant, and more arousing, than its equivalent daytime VE, and that

2. participants with a high stress-level would appraise a nighttime version of the desktop VE as less pleasant, less sleepy, and as more unpleasant, and more arousing, than the non-stressed participants, and that they would feel more anxious.

It is reasonable to assume that stress also influences the appraisal of the daytime VE; however the goal of this experiment was to study the effects of nighttime VE’s so no hypothesis is added for this condition.

To test these hypotheses we first exposed 52 young male participants either to a non-stressful control (reading) task, or to a task generating stress, the Trier Social Stress Test (Kirschbaum, Pirke, & Hellhammer, 1993). The TSST is a validated psychosocial stress task. To check the response to the TSST we measured free salivary cortisol and heart rate of all 52 participants.

After performing this initial task, they explored either a daytime or a nighttime version of a virtual environment, representing a small village with typical Italian architecture. The affective appraisal of the VE was measured using the Russell and Pratt list of adjectives (Russell & Pratt, 1980). The level of anxiety of the participants was assessed through a self-report scale.
In addition, we tested a possible anxiety induced narrowing of attention for nonthreatening environmental cues (due to a focusing on threatening stimuli) through a recognition test: participants were presented with screenshots depicting details from the VE, both from parts they had explored during the experiment and from areas they had not encountered. We expected that participants in the nighttime condition would recognize less neutral details than participants in the daytime condition, and even more so after prior exposure to acute social stress.

4.9.4 Materials and instruments

**Participants**

The experimental sample consisted of 52 male volunteers, whose age ranged from 18 to 32 years (M = 23.37, SD = 3.27). A population consisting of young males was chosen to rule out potential effects of menstrual hormonal cycle and age (Kudielka, Buske-Kirschbaum, Hellhammer, & Kirschbaum, 2004). None of the participants had any experience with the computer game Counter-Strike®.

Participants were randomly assigned to one of four different conditions: two experimental (stress) conditions (the TSST followed by exploration of either the daytime or the nighttime VE) and two control (relaxation) conditions (a relaxation period followed by exploration of either the daytime or the nighttime VE). Each group comprised 13 participants. The participants received a modest financial compensation for their participation.

**Experimental Environment**

The experiments were performed in two separate rooms. Both rooms were windowless, temperature-controlled, and sound-attenuated, and their doors were closed during the experiments. The first room (A) contained the computer system that was used to present the virtual environment, a CRT display and video recorder that were used to monitor and register the events in an adjacent room (B), and some additional chairs and a table. The second room (B) was equipped with a camera, and contained an office table with three chairs.

**General questionnaire**

In order to control for relevant personal characteristics we asked the participants six questions from the Big Five questionnaire (Goldberg, 1992), three of which measure the level of extraversion, and the other three measure the level of neuroticism. Since extraversion (Swickert, Rosentreter, Hittner & Mushrush, 2002) and neuroticism (Ode & Robinson, 2007) are known to correlate positively with stress, these factors were used as covariates in the analysis of the results. Other questions concerned the participants’ experience with games and virtual environments in general, and with Counter-Strike® in particular.

**Heart Rate Measurement**

Heart rate and saliva sampling were used to confirm the stress inducing effect of the TSST. Heart rate was continuously monitored at a sampling frequency of 1 Hz during the entire period of the experiment, using a chest strap Polar Wearlink 31 wireless transmitter and a wrist strap Polar RS400 Running Computer (www.polar.fi). After the experiment heart rates were averaged over one minute intervals.

**Saliva Sampling and Biochemical Analysis**

Cortisol, a neurohormone released when the stress system is aroused, was used to index stress responsivity. When humans perceive an event as stressful, a cascade of physiological events occurs along the hypothalamic-pituitary-adrenal (HPA)-axis, which result in cortisol release (Holsboer, 1999; Levine, Zagoory-Sharon, Feldman, Lewis, & Weller, 2007). Salivary measures of cortisol are a valid and reliable reflection of serum cortisol (Umeda et al., 1981; Levine et al., 2007).

Saliva samples for the assessment of free salivary cortisol were collected from all 52 participants, both immediately before and again 35 minutes after the start of the experiment (i.e. 15 min after the start of the TSST, when cortisol levels peak (Kirschbaum et al., 1993)). For further details, see Van der Welie et al.( 2009).
**Stress inducing task: the Trier Social Stress Test**

The Trier Social Stress Test is a well-established psychosocial and cognitive task that can be deployed to induce acute stress in laboratory conditions (Kirschbaum et al., 1993). Its protocol consists of a 10 minute preparation period (causing anticipatory stress), followed by a 10 minute test period (a 5 minutes public speaking task followed by a demanding 5 minutes oral arithmetic task: e.g. serially counting down from 1,022 in increments of 13), both performed standing before a critical jury (Kirschbaum et al., 1993, Williams & Hagerty, 2004). The TSST reliably provokes significant changes of cardiovascular parameters, subjective distress ratings, and an increase of cortisol levels (Kirschbaum et al., 1993) even in a virtual reality (Kelly, Matheson, Martinez, Merali, Anisman et al., 2007).

At the start of the TSST the investigator led the participant from room A to the second test room (room B), where three jury members were seated behind a table. They were introduced as experts in nonverbal behaviour and public speaking analysis, the participant was then instructed to prepare a 5 minutes speech in which he had to convince the jury that he was the perfect applicant for a vacant employment position. He was also informed that his presentation would be video- and audio-taped for later analysis. The participant was then taken back to room A, where the investigator left him alone to prepare his speech. After 10 minutes preparation time, the investigator returned to the participant in room A, and instructed him to go to the interview room B, and to leave his notes behind. There he performed the public presentation (5 min) and the mental arithmetic task (5 min), while standing before the critical jury. The jury members were instructed to withhold any normal encouraging smiles and nods while the participant gave his presentation, and to remain silent, with no leading questions when the participant paused. When the participant had finished the mental arithmetic task, the jury chairman told the participant to return to room A, where the exploration task was performed.

**Control Condition**

The protocol for the control groups was identical to the protocol for the experimental groups, except that the control groups did not perform the TSST, but browsed some magazines instead (for the purpose of relaxation), during a period of 20 minutes (the duration of the TSST).

**The Virtual Environment**

The VE used in this study was the Italy level of Counter-Strike® (Source Engine 7). This level represents a small village with a typical Italian architecture, with narrow streets, steep stairs, and with a market place in its centre surrounded by houses (Figure 4.41). Participants could walk through the streets of this village or enter and explore some of its houses. While exploring, the participants heard a simulation of the sound of their footsteps, as well as all sorts of background noise (music, a singing voice, a passing airplane, wind, rumour) through their headphones. We used both a daytime and a nighttime version of this VE. The nighttime environment is identical to the daytime environment, except for the low light levels. Due to the use of a contemporary Counter-Strike® game engine, the graphics are up to modern day standards. The Counter-Strike® game was modified and all regular game features (opponents and strategic game elements like the gun, information on the number of bullets left, and the map) were removed. Also, most keyboard and mouse functions were disabled. The keys “w,a,s,d” could be used, to move for-, back- and sideways through the environment. The only remaining function of the mouse was to rotate the viewing direction around the vertical axis (to look around).

The Counter-Strike® game ran on a Dell Dimension XPS600 computer. The virtual environments were displayed on a 19” Iiyama Prolite E481S display. Sound was presented through Sennheiser EH 150 headphones.
The VE Exploration Task

The VE exploration task was performed in room A, and lasted 10 minutes. The participant was instructed to imagine that he was a military scout situated in an unknown village, whose task it was to perform a thorough reconnaissance of this village. This task served to ensure that the participants would perform a detailed visual inspection of the VE. The participant was unaware that his affective appraisals and recollection of scenic details would be tested at a later stage in the experiment. Room lighting remained on when the participant explored the daytime VE, and was turned off when he explored the nighttime VE.

Affective appraisal of the environment on the Russell and Pratt scales

The affective appraisal of the virtual environment was measured using a translated and shortened version of the verbal scales developed by Russell and Pratt (Russell & Pratt, 1980). This shortened version consists of twenty bipolar adjectives that are descriptive of the affective quality of physical environments. Each of the four basic affective appraisal states (arousing, sleepy, pleasant, and unpleasant) is tested by five adjectives, using an 8-point Likert scale.

Recognition Test

The recognition test comprised a self-paced PowerPoint presentation of twenty screenshots taken from objects that are represented in the computer game Counter-Strike®. Ten screenshots represented objects that participants actually encountered when they explored the virtual environment (Counter-Strike® level Italy). The other ten screenshots represented objects residing in levels of Counter-Strike® which the participants had not explored. The screenshots were randomly
ordered within the presentation. The presentation was shown on the same monitor that was used for the exploration task, to avoid any representation (colour or contrast) differences. Participants were provided with a pencil and a score list. They were instructed to mark on the score list for each object whether they had seen it in the virtual environment or not. No feedback was provided to the participants as to the correctness of their answers. Similar recognition tests have been used in the literature to investigate the user experience of virtual environments (McCreary & Williges, 1998; Tortell, Luigi, Dozois, Bouchard, Morie, & Ilan, 2007).

**Anxiety self report scale: Translated State Anxiety Inventory**

A validated Dutch translation of the state self-report scale (Ploeg, 2000) from the Spielberger State-Trait Anxiety Inventory (STAI) (Spielberger,1985) was administered to assess how anxious participants felt twice: the first time before the TSST, the second time at the end of the experiment, after the task in the VE. This scale, which contains 20 items that can each be rated on a 1- to 4-point Likert scale, is specifically sensitive to changes in anxiety, and can therefore be used to evaluate the effect of interventions.

### 4.9.5 Procedure

Experimental sessions lasted one hour. After arrival at the laboratory, the participant was taken to room A, where he first read and signed an informed consent form. Then, he put on the chest strap with the Polar Wearlink 31 wireless transmitter and the wrist strap with the Polar RS400 Running Computer. Next, a saliva sample was obtained by letting the participant gently chew on the Salivette dental cotton for 1.5 minutes. After that he was asked to complete the Translated State Anxiety Inventory (Ploeg, 2000), the General questionnaire, and the Game Experience questionnaire. Then the participant received a general instruction about the content of the two tasks he was asked to perform. Participants in the stress condition first performed the TSST, which lasted for 20 minutes. Participants in the control condition spent the same period reading magazines. The second task was the same for both groups and required the participant to explore a virtual environment for a period of 10 minutes. Next all participants performed the VE exploration task in room A.

After finishing this task, the participants were asked to complete the Russel and Pratt verbal scales on affective qualities and the Yes-no recognition test. Next, a (second) saliva sample was obtained. This was done 35 minutes after the start of the experiment (= 20 minutes stress/relaxation +10 minutes VE exploration + 5 minutes filling out questionnaires), and 15 minutes after cessation of the stress/relaxation condition, when cortisol levels are known to peak (Kirschbaum et al., 1993). Finally, the participant was asked to complete the Translated State Anxiety Inventory for the second time. At the end of the experiment the participants were extensively debriefed, paying special attention to participants in the stress condition.

### 4.9.6 Results

The general questionnaire did not show significant differences between the conditions, nor correlations in the results of the questionnaires on affective appraisal and anxiety.

In this section we first evaluate the results of the stress tests (heart rate, free salivary cortisol) to assess the extent to which the TSST evoked a stress response. Then we present the results of the Translated State Anxiety Inventory, the affective appraisal and recognition tests, which were administered after the task in the virtual environment.

**Stress Induction**

**Heart rate**

To control for individual differences in physical fitness levels, we first assessed the mean heart rate for each individual during a period of 5 min before the start of the experiment. Heart rate was then normalised by subtracting this baseline (resting) value. Figure 4.43 shows the normalised mean heart rate over the course of the experiment, for all 4 experimental groups. The two stress-groups showed an elevated heart rate over the period of the TSST stimulation (i.e. between 10 and 30
minutes after the start of the experiment). Shortly after cessation of the stressor, heart rates dropped to baseline. The two control (relaxation) groups showed a minor decrease in heart rate during the same period. This result indicates that the TSST procedure effectively provoked an increased heart rate.

Figure 4.43 Normalised mean heart rate (beats per minute) as a function of elapsed time since the start of the experiment (time=0), for all 4 conditions. Heart rate was normalised by subtracting the mean heart rate determined for each individual during a period of 5 min before the start of the experiment (baseline). The TSST was performed between 10 and 30 minutes after the start of the experiment.

**Salivary Cortisol**

Figure 4.44 shows the mean free salivary cortisol level (nmol/l) for the first sample at the start of the experiment and the second sample, 5 minutes after finishing the VE exploration task. The timing of the second sample is 15 minutes after the end of the TSST or the control condition. The exact values and their standard deviations are represented in Table 4.30. At the start of the experiment there was no significant difference between the level of free salivary cortisol of the four groups. A repeated measures t-test for mean cortisol level showed that the mean salivary cortisol levels significantly increased in response to the TSST simulation for both stress groups (stress day: t(12) = -5.48, p = .00, and stress night: t(12) = -2.89, p = .01), and significantly decreased after the relaxation period for both control conditions (control day: t(12) = 2.75, p = .02, and control night: t(12) = 4.24, p = .00). Thus, it appears that the TSST (stress condition) effectively provoked an increase of cortisol levels, whereas the relaxation task (control condition) effectively lowered cortisol levels.
Figure 4.44 The mean free salivary cortisol level (nmol/l) at the start of the experiment and 35 minutes after the start of the TSST or control condition, per group.

Table 4.30 The mean level of free salivary cortisol (nmol/l), in the first and second sample, and the mean difference between second and first sample, with their standard deviations, for each group.

<table>
<thead>
<tr>
<th></th>
<th>Stress</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Night</td>
<td>Day</td>
</tr>
<tr>
<td>First sample M (SD)</td>
<td>8.15 (3.19)</td>
<td>9.33 (8.94)</td>
</tr>
<tr>
<td>Second sample M (SD)</td>
<td>14.21 (7.92)</td>
<td>19.74 (11.62)</td>
</tr>
<tr>
<td>Difference M (SD)</td>
<td>6.06 (7.56)</td>
<td>10.41 (6.85)</td>
</tr>
</tbody>
</table>

**Affective appraisal of the environment on the Russell and Pratt scales**

Table 4.31 lists the means and standard deviations of the total score on each factor of the verbal scales (arousing sleepy, pleasant and unpleasant), for each condition.

Table 4.31 The mean and standard deviations of the total score on each scale of the Russell and Pratt list of adjectives, for each condition.

<table>
<thead>
<tr>
<th></th>
<th>Stress</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Night</td>
<td>Day</td>
</tr>
<tr>
<td>Arousing M (SD)</td>
<td>3.54 (.87)</td>
<td>3.40 (11.2)</td>
</tr>
<tr>
<td>Sleepy M (SD)</td>
<td>5.12 (1.16)</td>
<td>4.94 (1.13)</td>
</tr>
<tr>
<td>Pleasant M (SD)</td>
<td>3.85 (1.07)</td>
<td>4.45 (1.42)</td>
</tr>
<tr>
<td>Unpleasant M (SD)</td>
<td>4.72 (1.10)</td>
<td>3.77 (1.58)</td>
</tr>
</tbody>
</table>

Note: Stress (control) = group that performed the stress (relaxation) task. Day (night) = group that explored the daytime (nighttime) VE. Note that a score of 4.5 indicates a neutral assessment, on the 8 point bipolar scale.

To assess the reliability of the verbal scales (Russell & Pratt, 1980; Russell, Ward & Pratt, 1981), Cronbach’s alpha was calculated for each of its factors. All Cronbach’s alphas were larger than .70, which is sufficiently high to preserve all items per factor.
For each of the four factors a 2x2 ANOVA (stress level × time-of-day) was performed. There was a main time-of-day effect for the factor pleasant (F(1, 48) = 6.78, p = .01) but no significant stress or interaction effect. There was also a main time-of-day effect for the factor unpleasant (F(1, 48) = 6.59, p = .01 but no significant stress or interaction effect. The daytime environment was considered more pleasant, and less unpleasant, than the nighttime version.

No significant effects were found for the factors arousing and sleepy.

**Translated State Anxiety Inventory**

Table 4.32 shows the mean scores on the Translated State Anxiety Inventory, both at the start (begin score) and at the end (post score) of the experiment, and the difference between these two scores, for all 4 groups.

<table>
<thead>
<tr>
<th></th>
<th>Stress</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Night</td>
<td>Day</td>
</tr>
<tr>
<td>Begin score M (SD)</td>
<td>29.92 (6.22)</td>
<td>34.08 (6.54)</td>
</tr>
<tr>
<td>End score M (SD)</td>
<td>33.46 (6.97)</td>
<td>41.77 (9.89)</td>
</tr>
<tr>
<td>difference score M (SD)</td>
<td>3.54 (7.60)</td>
<td>7.69 (10.45)</td>
</tr>
</tbody>
</table>

All groups scored higher at the end of the experiment, which implies that their perceived anxiousness increased during the experiment. A 2x2 ANOVA (stress level × time-of-day) showed no main effects of stress or time-of-day. No significant interaction effect was found, however a trend was noticed: F(1,48) =3.29, p= 0.076. A follow-up t-test to compare the difference scores between the conditions revealed a significant difference only between the stress-day and control-day conditions, t(24)= 2.14, p=.043.

**Recognition**

Table 4.33 lists the mean number of correct responses for the Yes-no recognition test for each of the 4 experimental conditions. The number correct is the number of objects that participants have correctly recognized as actually present in the part of the VE they explored, plus the number of objects they correctly dismissed as not being there. The maximum score is 20. A 2x2 ANOVA (stress level x time-of-day) follow-up t-tests revealed no significant difference in the number correct responses between any of the four groups (p > 0.05).

<table>
<thead>
<tr>
<th></th>
<th>Stress</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Night</td>
<td>Day</td>
</tr>
<tr>
<td>number correct M (SD)</td>
<td>16.46 (1.39)</td>
<td>17.08 (1.38)</td>
</tr>
</tbody>
</table>

Summarizing, both the glucocorticoid (salivary cortisol level) and sympathetic (heart rate) components of stress response, and the scores on the Translated State Anxiety Inventory, all indicate that the TSST procedure effectively elicited signs of stress.

The nighttime environment was considered less pleasant and more unpleasant by the participants, but stress had no effect. The nighttime VE was not considered more arousing than the daytime environment. The Translated State Anxiety Inventory indicates that the night condition did not evoke anxiety in the participants who experienced stress. No recognition difference was found between the conditions.
4.9.7 Conclusion and discussion

In this study we investigated the effects of ambient lighting and prior stress on the emotional response to, and on the affective appraisal of, a desktop VE. Participants were first exposed to either the Trier Social Stress Test, or to a non-stressful control task. The daytime environment was considered more pleasant, and less unpleasant, than the nighttime version, by both groups. In contrast to our hypothesis, stress did not negatively influence the appraisal of the nighttime environment. There was also no significant effect of time-of-day on the factors arousing and sleepy of the Russell and Pratt verbal scale. This finding is in contrast with the results of Bishop and Rohrman (2003), who found that people appraised a nighttime VE as more arousing than a daytime VE. While the appraisal of an environment as pleasant or unpleasant may be strongly determined by the lightness level or time-of-day, the appraisal as arousing or sleepy (the static quality) is perhaps less dependent on the time of-day, since light and dark environments can both be static or dynamic. Contrary to our expectations, the Translated State Anxiety Inventory indicates that the night condition did not evoke anxiety in the participants who experienced stress.

The results of the recognition test showed no significant difference in the number of correctly recognized details between any of the 4 different conditions. This is in agreement with results reported by (Lueth, Meier, & Sandi, 2008): memory tests performed by participants after the TSST, indicated that the degree of stress and enhanced cortisol levels in the experiments were sufficient to “impair working memory, enhance spatial episodic memory, and to facilitate classical conditioning for aversive stimuli.” In contrast, they did not affect performance in verbal explicit memory tasks, or in implicit learning tasks that involved neutral or positive stimuli.” Differences in recognition scores may become only apparent for a more difficult recognition task, possibly in combination with more arousing events and emotional scenes (Lupien, Maheu, Fiocco, & Schramek, 2007). So, a reason for this lack of effect may be that the test details were too salient, both in the daytime and in the nighttime environment, so that the task may have been too easy in all conditions.

Summarizing, we find that the desktop VE was considered less pleasant and more unpleasant by the simulated nighttime lighting conditions, but not more arousing. Also acute prior stress did not evoke the expected appraisals, nor did it elicit higher anxiety in the participants during the task in the VE.

The absence of an effect of stress may be due to the combination of (a) an adaptation effect and (b) the lack of an engaging task. It has been observed that a prior affective experience biases the appraisal of the corresponding affective quality of an environment in a direction opposite to the value of the prior experience (Russell & Lanius, 1984). The TSST test was definitely an unpleasant and distressing experience for the participants in this study, which may have shifted their appraisal of the VE in the direction of pleasantness, and away from arousal, thus obscuring a possible darkness effect.

Another explanation of the fact that the stress and control groups showed no difference in their affective appraisal of the VE may be that distressed participants did not focus on the environment in an affective sense due to a narrowing of attention. Maybe they needed all their attentional resources to focus on the exploration task in the VE so that they were less susceptible to affective qualities of the environment. Engagement, and interest, in the mediated environment is an important factor for the feeling of presence, in a VE (Freeman et al., 2005), and some level of engagement is required to assess the affective qualities of an environment. More precise, according to Lazarus (1991), an emotion will only occur when the situation is important for the person at that moment.

Perhaps stronger differences in affective appraisal between the groups would have appeared if the stressor had been inherent in the VE itself (e.g. a virtual version of the TSST (Kelly et al., 2007)) or had been directly related to the task (Morris et al., 2004).

We conclude that merely lowering the ambient light level in a desktop VE is not sufficient to elicit anxiety, or make the affective qualities of the environment more arousing and less sleepy. In this respect desktop VE representations are different from immersive VE representations. High
stress levels, which may occur in training situations, do not per se make nighttime VE’s less pleasant and more arousing. This finding has implications for the use of desktop VE representations in visualisations, serious gaming and training applications.

**Acknowledgements**

We thank Erik van der Spek for adapting the Counter-Strike® level Italy (Source Engine 7) for the purpose of this experiment. We are also indebted to Roy Raymann and Ineke Klöpping-Ketelaars for their advice and help with the experimental design and the saliva sampling and analysis, to Etiënne Grassère and Susan Vrijkotte for their help with the heart rate registration equipment, and to Anja Langefeld for her support with the data analysis.
4.10 Summary of the results

The experiments described in Chapters 4.2 -4.9 encompass a broad range of virtual environments, which were developed to visualise urban environments and buildings, to use as training environments, or specifically for this research. In the experiments we manipulated different variables in the content and representation of these 3D models or virtual environments, and in response moderators, which are all included in the framework in Chapter 3. We made a selection of evidently important variables, based on our observations in several projects and on acknowledged issues in the development of 3D environments. They represent characteristics often commented on by users of visualisations and virtual environments. Examples are the clean and static appearance of 3D visualisations, and the absence of sounds in training environments. A third factor that determined the selection was the availability of appropriate 3D models and virtual environments, that were used in real training situations, and for which specific problems were articulated by trainers, users or developers. For instance, for the study described in Chapter 4.6, exam candidates in the virtual highway environment for traffic controllers emphasized that they did not recognize the experience of being on a real highway, which in their view influenced their decisions and actions. We assessed the affective appraisal of the environments, as well as aspects of the affective and emotional response of participants, such as engagement.

4.10.1 Experiments 4.2 – 4.5: Affective appraisal of environmental visualisations

Decay can be interesting in dull environments

In the experiment “Decay in an urban environment” (Chapter 4.2), the large number of negative elements such as nuisance elements and physical incivilities introduced in a virtual environment of an urban area only had a mildly negative effect on the viewer’s appraisal; the manipulated model only scored significantly lower on the SMB-factor Social Status than the original “clean” model. Although the objects were recognized as elements indicating urban decay, they also added (positive) interest to the rather dull and empty environment, typical of this style of visualisation. So, cues that have a negative effect on the assessment in reality may elicit a different appraisal in a virtual environment. In a limited second experiment we found evidence that weather conditions, especially blue sky and bright colours versus grey sky and dull colours, had a positive impact on the appraisal of the area.

Previous experience can dominate the affective appraisal

For the study of the effects of lighting on the affective appraisal of an indoor environment (Chapter 4.3), two versions of the original 3D model of the Carré Theater were created: one in which the lighting quality was diminished, a second with enhanced illumination. The results suggest that viewers are susceptible to differences in lighting conditions when visually comparing virtual environments, but far less without a reference. The scores on the SMB-scale showed few differences between the two versions, indicating that the affective appraisal of the virtual theater was not influenced much by the lighting quality. The narrow corridor was judged more enclosed in the enhanced version (factor Enclosedness of the SMB-scale), but more original in the degraded version (factor Originality). The models contain so many typical elements of a baroque style theater that we think they have invoked impressions of earlier visits to, or descriptions and images of, a similar theater. Consequently, they induced subjective responses that are based on these mental representations and not only on the actual represented environment.

Virtual can be more pleasant than real

The aim of the experiment described in Chapter 4.4 (“Affective qualities of an urban environment on a desktop computer”) was to examine the comprehensive effects of modeling, and representation on a desktop system, on the affective qualities of an urban area. The area represented in the 3D model, was considered more pleasant, more uniform, and more modern than
the same area in reality. The higher scores on the SMB-factors Pleasantness and Affection are probably determined by the general fresh and neat look of the area, and the colours, which were considered prettier and livelier in the model. The virtual environment was considered more uniform, which can be ascribed to the photo textures. These textures do not show differences in materials and relief present in the real world, and offer homogeneous colour hues. The factor Complexity, reflecting the perceived vivaciousness and colourfulness of an area, was not significantly different between the two groups (p > .05). Although the colours in the model were considered lively, the area is emptier and contains less movement and visual variation, and no sounds, which may have countered the effect.

**Dynamic features make the environment more arousing**

In Chapter 4.5 (“Visual dynamics and sounds in an outdoor environment”) sounds and movement were added to a visualisation of a Dutch polder area. We measured the appraisal of the environment using the Russell and Pratt scales, and the engagement of participants with a questionnaire and facial muscle tension (EMG). We concluded that sound and visual dynamic features can effectively be used to convey the impression of certain severe weather conditions in a 3D environment. The severity of the weather circumstances would in real life negatively influence the affective appraisal of the environment and increase the urgency of the situation. In this experiment, we found that sound increased the arousing qualities of the environment, so made the environment less “dull”; sound only made the environment less pleasant when not accompanied by visual dynamics. Visual dynamics increased the pleasant qualities of the environment. Together they made the environment livelier. Users thus appreciated diversity in the environment, even when the features in reality are considered unpleasant. Only the visual dynamics had a positive influence on the engagement of the participants. In contrast to what we expected, sounds did not contribute to the user engagement in this environment, at least in this experimental setup.

4.10.2 Experiments 4.6 and 4.7: Affective/emotional reactions to virtual training environments

**No direct correspondence between validity for cognitive tasks and affective/emotional responses**

The goal of the experiment using the exam environment for traffic inspectors was to assess the effect of enhanced visual and auditory depth and speed cues on the affective/emotional response of the user, especially the feeling of risk (Chapter 4.6). Our hypothesis was that the enhanced representation of speed and distances in the VE would increase the feeling of risk, by improving the convincingness of the simulation, and evoking past experiences. There were indications that participants in the enhanced version were able to make more accurate judgments about the distances and about egocentric and exocentric speeds of vehicles that were on the same part of the road as the participant. Moreover, participants in the enhanced version felt more certain of their estimates. However, the perceived speed of the vehicles near the user was not higher in the enhanced version, and objects were not estimated to be closer, which would create a more dangerous situation for road users. The average speed estimates for the video clips judged by the participants also did not differ significantly between conditions. Nonetheless, both the assessment of danger (for the participants in the enhanced scenario) and the feeling of risk were significantly higher. This experiment confirms our viewpoint that validity of a VE for cognitive tasks such as assessing speed and distances does not directly correspond to validity for affective tasks.

**Instructors and mental effort distract attention from arousing sounds**

For the study described in Chapter 4.7 we added realistic and task-relevant sounds to a training scenario for fire fighter commanders. We performed two field studies to assess how these sounds influence the arousal and engagement of the trainees. Although the soundtracks were developed in close collaboration with expert firefighters, they did not have the expected effect on the trainees’ emotional response. We concluded that the focus of attention of the participants was on the
communication with the instructor, and on creating a mental model of the fire and its surrounding areas, which distracted the attention of the participants. Hyper-real (exaggerated) visual and auditory effects may be required to overcome the distractions provided by the training context and to achieve “affectively intense training”.

4.10.3 Experiments 4.8 and 4.9: Effects of response moderators in environmental visualisations

**Cybersickness makes environments less pleasant and more arousing**

In Chapter 4.8 we reported on a study of the effects of cybersickness on viewers of a virtual environment depicting a city with mixed European architecture. We measured the appraisal of the environment with the Russell and Pratt verbal scales. In our experiment, participants who experienced cybersickness valued the virtual environment as less pleasant and more arousing than participants who did not experience any appreciable symptoms of cybersickness. The most plausible explanation for this is that the change in core affect the participant experienced while watching the virtual tour was misattributed to the environment itself (Russell, 2003).

**Nighttime conditions: less pleasant, not anxious**

Our final experiment (Chapter 4.9) investigated the effects of ambient lighting and prior stress on the emotional response to, and on the affective appraisal of, a desktop VE. The VE showed a small village with a typical Italian architecture. We again applied the Russell and Pratt verbal scales. The desktop VE was considered less pleasant and more unpleasant by the simulated nighttime lighting conditions, but not more arousing. Also acute prior stress did not evoke the expected appraisals, nor did it elicit higher anxiety in the participants during the task in the VE. The absence of an effect may be due to the combination of (a) an adaptation effect and (b) the lack of an engaging task. We conclude that merely lowering the ambient light level in a desktop VE is not sufficient to elicit anxiety, or make the affective qualities of the environment more arousing and less sleepy. In this respect desktop VE representations are different from immersive VE representations.

4.10.4 Outline of the experimental results

**Participants overlook, fill in and are easily distracted**

In the experiments, we noticed a number of recurring patterns. They will be discussed further in Chapter 5 but are summarized here.

First of all, features that are considered unpleasant in real environments (such as physical incivilities or severe weather conditions) do not have the expected effects when included in virtual desktop environments. Although they are recognized as disagreeable, they also seem to make the virtual environment, often considered rather dull and lifeless, more interesting and vivacious.

Secondly, features in the representation that may be seen as flaws, such as low quality of illumination, and the absence of sounds, do not seem to bother users, or only slightly. The features are only manifest when a version without these flaws is presented at the same time. Viewers naturally fill in or add information that is not presented in the 3D environment, and overlook these flaws when their attention is not directed towards them.

In the third place, the attention of the users plays an important role in the appraisal and experience of the virtual environment. In the experiments we frequently noticed that many other factors required the attention of the users, such as navigation, the tasks at time of use, and the real environment in which the participants were located. The impact of features of the virtual environment on the appraisal and affective/emotional response of the user seemed to be reduced by these distractions.

Finally, the affective appraisal of a virtual environment is an amalgam of the appraisal of the environment and its representation, and is influenced by the mental representation of the viewer at that moment, and his or her affective state. Participants in the experiments do not spontaneously
distinguish between these, and the measuring tools for assessing the appraisal of environments do not reveal the role of the different determinants either.

In Chapter 5, the consequences of these issues for the validity of virtual desktop environments are discussed.
5 Conclusions & Discussion

5.1 Introduction

This thesis focuses on the affective appraisal of and affective/emotional reaction to 3D visualisations and virtual training environments on desktop computers. We contend that affective/emotional responses to 3D models are often overlooked components in the design and use of virtual environments, which may compromise their validity for visualisation and training purposes. A valid simulated environment is defined by Wergles and Muhar (2009) as one that produces a cognitive, affective and behavioural response in the observer equivalent to the response produced by the real environment. The experiments described in this thesis reveal that factors in the 3D environment determine the affective response in a distinctive manner; and also the three types of response are distinct and must be addressed as such.

Central to this issue are the tasks for which a specific 3D environment is used, and requirements related to these tasks. They are generally described on a cognitive level (for instance learning goals, or functionality of an interface), but on closer examination often prove to include the affective appraisal of an environment, or an affective/emotional reaction to an environment in training. However, this affective component is not taken into account adequately in the development of most 3D environments. In this research project we studied for example 3D visualisations designed for informing citizens about alternative designs for an urban renewal project (Chapter 4.4). Viewers explore these visualisations to assess for instance the public facilities and accessibility of the area, but also (and often implicitly) the ambience of the envisioned area. For this appraisal, many features of the models and the display such as colours, indications of the function and use of the area, time-of-day, and season may be important. When these features are included offhandedly, without deliberate consideration, they may compromise the validity of the visualisation for this purpose. We also examined virtual training environments, which are expected to be valid substitutes for dangerous circumstances. If in virtual training environments darkness does not have the same effect on the affective/emotional response of the trainee as darkness in reality, it obviously limits the training scenarios, and the circumstances for which the trainee can be prepared (Chapter 4.9).

The affective appraisal of and affective/emotional reaction to virtual environments on desktop computers are complex topics. They involve knowledge from different research domains, and concern a large number of possible determinants, such as characteristics of the design, display, context of use, and the user.

The main research question for this thesis was:

*How do visual and auditory features in a virtual environment on a desktop environment influence the affective appraisal and reaction of the viewer?*

This comprehensive question was broken down into four sub questions:

1. How do users appraise virtual environments, and how is this different from the appraisal of real environments?
2. How do personal characteristics, momentary user states, and context of use moderate the effect of visual and auditory features on the affective appraisal of a virtual environment used for environmental planning or for training?
3. How can features of environments and the representation, and personal and contextual moderators, be organized into a framework of determinants of affective appraisal?
4. Which measuring instruments are sufficiently sensitive and accurate to assess affective/emotional responses to virtual desktop environments?

To answer these questions we used (1) studies in environmental psychology, human-computer interaction and media theory, (2) our observations of the use of 3D models and virtual
environments, and (3) the results of 8 experiments included in Chapter 4. After discussion of the sub questions we return to the issue of validity, and suggest recommendations to researchers as well as developers.

5.2 Answers to the four sub questions

Research Question 1

How do users appraise virtual environments, and how is this different from the appraisal of real environments?

Appraisal of virtual environments

In Chapter 2 (Figure 2.4) we introduced a schematic view of the appraisal of a 3D visualisation or desktop virtual environment. We now extend it by including affective appraisal as a specific component of the assessment by the viewer (Figure 5.1). The viewer perceives the virtual environment, and appraises its affective qualities. These differ from those of a real environment, because they are modified by the modeling process, the display, and the context of viewing. The momentary user state may influence the appraisal. In the process of appraisal, associations of similar places, and related subjective experiences, are triggered by cues in an environment. People have mental representations of places (also called schemas), that contain information about physical, social, organisational, purposive and affective qualities of these places (Imamoğlu, 2009); these may be very powerful (Chapter 2.5). Together with the perceived 3D model they constitute a composite mental representation. The viewer assesses this mental representation, which is updated continuously (Kim & Richardson, 2003; Kurby & Zacks, 2012). The appraisal may lead to a change in core affect (an affective reaction, for instance a shift in arousal), and to an emotional episode (an emotional reaction, such as anxiety).

Figure 5.1 A schematic representation of the affective appraisal and affective/emotional reaction of a viewer. The appraisal of the viewer is based on the combined information of the 3D environment and the pre-existing mental representation.

The mental representation

The importance of the pre-existing mental representation or schema was especially manifest in the experiment using the models of Carré Theater (Chapter 4.3). Although the quality of illumination was severely degraded in one condition, participants professed their immediate associations with
visits to theaters and with the festive atmosphere at such occasions. Their appraisal of this degraded version differed only slightly from the appraisal of the viewers of the version with enhanced illumination. Also other features in the representation that may be seen as flaws, such as low quality graphics (Chapter 4.4), and the absence of sounds (Chapter 4.7), do not seem to bother users, or only slightly. The features are only evident when a version without these flaws is presented at the same time. Viewers naturally fill in or add information that is not presented in the 3D environment, and overlook these flaws when their attention is not directed towards them.

In the appraisal of the urban area in Chapter 4.4, viewers who were familiar with real Venlo gave lower scores on the social status of the virtual environment, than viewers who were not familiar with Venlo. They apparently used their knowledge of the city in the appraisal of this aspect, whereas they were supposed to judge the virtual environment. In our experiment with a virtual training environment for leading firefighters we also observed that mental representations not only completed, but even overrode the input provided by the virtual environment (Chapter 4.7). In this case the mental model of the fire and its direct surroundings was put together by the participants using the spatial information offered by the virtual environment, and the verbal information provided by the instructor. During the training the participants focused so strongly on this mental model that the absence or presence of sounds did not have the expected difference in their response to the virtual training.

In environmental design and planning, 3D models are used to help non-expert viewers to understand and communicate about existing or future environments. They are meant to reduce misinterpretation and ambiguities on spatial layout and dimensions, which may easily occur when viewers must combine maps, elevations and artistic impressions into a complete three-dimensional mental representation. But it is generally overlooked that new ambiguities and biases are introduced in visualising the elements in the environment and its ambience, which especially influence the affective qualities of the environment: for instance, the depiction of atmospheric circumstances, colours, indications of the presence of people or the function of an area can have a strong impact on the affective appraisal of a user of an area (Chapter 4.4, 4.5).

The impact of these and similar “cues” cannot be predicted sufficiently using only evidence of the effect of these cues on the appraisal in real life. Environmental cues that were expected to have a negative effect on the perceived pleasantness of an area, such as indications of vandalism and urban decay, and severe weather conditions, in our experiments induced an effect in the expected direction, but not in intensity. Moreover, some of these cues also generated interest, and made the environment livelier (Chapter 4.2, 4.5). Therefore, an important conclusion is that cues cannot always be assumed to have the same effect in virtual environments as in real environments. Adding cues to a 3D environment may improve the perceived quality of the visualisation by creating a more photorealistic result, but does not per se lead to a more valid simulation.

**Convincingness of 3D environments**

The perceived realism of a 3D visualisation may create a misleading sense of confidence that the represented environment is an accurate representation of for instance an envisioned future. When humans view an environment in reality, they also draw from their memories and associations, and may in fact appraise their mental representations. But virtual environments are simplified and modified representations, from which most of the multimodal and varied information of a real environment is omitted; therefore viewers complete the model using their mental representation. This happens without effort, and viewers are often not aware of this process. Even with few visual cues, participants feel confident about their interpretation of elements of the model, such as season, weather, or materials used in the virtual environment (Chapter 4.2, 4.3), which influence their appraisal of the area.

From media history, it is known that whenever a medium provides a new sensory stimulation the audience at first is impressed with this new feature and experiences high perception of realism (Jusselsteijn, 2003). Several authors have determined that the confidence of viewers of the accuracy of their own perception is higher for technologically more advanced techniques, such as animations
and 3D models than for other, more conventional types of representations (Mahdjoubi & Wiltshire, 2001; Sirikasem & Shebilske, 1991; Skog & Söderlund, 1991).

The confidence in virtual representations can be effective in training environments, but may pose problems for the validity of 3D visualisations. It is well known that visualisations used for citizen participation can elicit strong reactions: viewers sometimes feel offended by the visualisations because the convincingness of the representation suggests that the shown plans are already agreed on. Viewers also react strongly to minor details in the visualisation that they consider as inaccurate. For this reason, there is an ongoing debate on techniques and styles that make the status of the plans and uncertainty of components obvious to the viewer (Appleton & Lovett, 2005). Until conventions are developed to convey uncertainty, it is recommended to offer viewers multiple versions of the visualisation showing possible variations of the elements that are not yet decided on (such as building materials) or that are dynamic (weather and lighting conditions). This allows viewers to construct a nuanced and better grounded appraisal of an environment.

When viewers appraise a virtual environment they do not distinguish between their appraisals of the represented environment, of the representation, and of their individual mental representation. The mental representation fills in more information than the viewer is aware of. At the same time, the way the information is presented does not alert the viewer to be cautious in his or her assessment.

Research Question 2

How do personal characteristics, momentary user states, and context of use moderate the effect of visual and auditory features on the affective appraisal of a virtual environment used for environmental planning or for training?

In the framework presented in Chapter 3 we included individual characteristics as a category of response moderators, such as pre-existing mental representations, expectations, mood, cybersickness, and previous experience. Other personal characteristics, such as cultural and technical background may also influence the affective/emotional response to a virtual environment (Gorini et al., 2009) but were outside the scope of this research. The influence of mental representations or schemas was explained above.

Personal characteristics

The affective state of a user may influence the appraisal of an environment, as we demonstrated in Chapter 4.8 and 4.9. Viewers who experienced cybersickness attributed their affective state to the virtual environment and valued the environment as less pleasant and more arousing than the control group. In contrast however, a state of acute stress did not negatively influence the affective state of the participants in our experiment. We attribute this to the effect of disengagement or lack of attention. Distractions and disengagement attenuate the impact of any cue. The attention of the stressed participants was probably distracted from the task in the virtual environment by the previous unpleasant experience.

Virtual environments are often expected to engage users and keep their attention. In the course of the experiments conducted, it became increasingly evident that personal relevance, involvement and attention are also response moderators, and have an important impact on the viewer’s response: an event or environment must be relevant for a viewer to elicit any emotion, and personal relevance or involvement also plays a role in the appraisal process. Without this, viewers may judge the represented environment on representational features, and value for instance the graphics quality. In contrast, when the model for instance represents their future living area they are more attentive to the content, so the environment itself, for positive and negative elements (Chapter 4.2).

Context of use and distracting elements

The attention of the user plays an important role in the appraisal and experience of the virtual environment. In the experiments we frequently noticed that many other factors required the
attention of the users, such as navigation, the tasks at time of use, and the real environment in which the participants were located. The impact of features of the virtual environment on the appraisal and affective/emotional response of the user seemed to be reduced by these distractions.

In practice, the importance of the context of use is strongly underestimated. By the context of use we mean the actual conditions under which the visualisation or virtual training environment is used in a normal day to day working situation. In the field experiments we noticed many distractions, for instance operators who are present in the training, problems when interacting with the system or navigation, and the environment in which the monitor is set. Noise and ambient illumination often distract the viewers’ attention away from the events on the screen.

In lab environments distracting factors may also occur. Because desktop monitors cover a small part of the human field-of-view, and only visual and sometimes auditory information are offered, the real environment may still distract the viewer and provide cues that are not in line with the virtual environment. The environment designed for the examination of road inspectors for instance was not considered convincing by the users, because they missed many cues that in reality add to the perceived danger of standing on a highway (Chapter 4.6). By enhancing the visual and auditory cues we slightly improved the perceived danger. It is doubtful that these relatively small additions also have sufficient effect in the exam context in reality, because this context is typically full of distractions such as described above.

In the course of this research we attended examinations and trainings using virtual environments, and sessions for citizen participation on urban planning in which visualisations were presented. In none of these instances, the location was adapted to create optimal circumstances for viewing or training. The ambient illumination was too strong, diminishing the perceived graphics quality of an environment and also the effects of the ambient conditions in the environment; the position of the trainee was often not on optimal viewing distance; people were allowed to walk in and out of the room, or even talked, which distracted trainees; or technical flaws interrupted the flow of the visualisation or training. Careful attention to the context of the presentation of a virtual environment can improve the overall effect on the viewers to a great extent. A second recommendation especially for training environments is to use priming techniques, for instance a task briefing or a preceding story, to prepare the viewer and create a frame of reference (Tortell et al., 2007; Park et al., 2010).

The absence of personal involvement, factors that diminish the perceived (graphics and audio) quality of the 3D environment, and factors that distract the attention of the user, attenuate the impact of cues and thereby the intensity of the affective/emotional response. Therefore, the importance of personal involvement and the context of use of a visualisation or a virtual training for their validity are underestimated.

Research Question 3

How can features of environments and the representation, and personal and contextual moderators, be organized into a framework of determinants of affective appraisal?
Chapter 3 introduced a framework for understanding and analyzing the affective appraisal of and reaction to virtual environments. It is based on the framework for understanding environment-user relationships in service organisations (Bitner, 1992), and extended with a category “representational modifiers and determinants”. It identifies the main categories of factors of a 3D model and its use that influence the affective appraisals of, or affective/emotional reaction to, a virtual environment.

_The first column_ “environmental dimensions and features” comprises three main types of elements of real environments, based on Bitner’s model (1992): 1. spatial layout and functionality, 2. signs, symbols and artefacts, and 3. ambient conditions. This division can also be applied easily to 3D models: static elements (buildings, roads, trees etc.) are the main building blocks of any virtual environment. Signs, symbols and artefacts are introduced in 3D models to reveal or stress how the area is used, and to give indications on for instance the social status of an area. Ambient conditions include temperature, lighting, noise, music, and scent.

Our experiments supported the importance of identifying these three main types of elements in environments, because they are affected by the representational modifiers (column 2) in different degrees. Regarding the category of signs, symbols and artefacts, we found that negative elements such as nuisance elements and physical incivilities in an urban area only had a mildly negative effect on the viewer’s appraisal and even generated interest. The representation of ambient conditions requires much effort from modelers and often goes beyond the representational capacities of desktop virtual environments. We found that ambient conditions represented in the 3D environments influenced the response of our participants but often weaker or with a different effect than expected. For instance, the illumination effects in the virtual theater described in Chapter 4.3, and the nighttime lighting in the model of a village described in Chapter 4.9, only mildly influenced the affective appraisals of the viewers, although the visual and/or auditory manipulations performed in the models to represent them were considerable. Visual dynamic elements made the polder environment (Chapter 4.5) more pleasant, although the elements themselves, indicating a thunderstorm, are generally considered as unpleasant.

_The second column_ of the framework shows the representational modifiers and determinants that influence the effects of environmental dimensions and features. It includes modifications that are related to the medium itself, such as the loss of multimodal information; choices in the modeling process, for example showing specific atmospheric conditions; and the display on a monitor that for instance confines the field of view. Variables from this category were included in all of the experiments discussed in Chapter 4, sometimes in conjunction with each other. Although the determinants are distinguished separately in the framework, in practice they often occur together and are difficult to separate completely.

All our experiments support the premise that manipulations of representational modifiers and determinants can influence the affective/emotional response of users. Modifiers and determinants...
caused by the medium itself and by the platform chosen were examined in several experiments described in Chapter 4. The absence of sounds and other sensory information was assumed to be important for the appraisal of the environments studied in Chapter 4.5 (the polder environment in severe weather), and Chapter 4.7 (the house on fire). Sounds accompanied by the appropriate visual information (such as the sound source) influenced the appraisal of the polder environment by making it more arousing. The sounds added to the training of leading firefighters showing the house on fire only had a minor effect, but we ascribe this to the strong effect of the mental representation which is explained earlier. To the highway environment used for the examination of traffic officers (Chapter 4.6) sounds and visual effects were added to enhance the depth and speed cues. They included cues for motion parallax, aerial perspective, and a larger simulated field-of-view. The manipulated version indeed elicited a higher feeling of risk.

In the modeling process many choices are made concerning the level of detail, illumination, atmospheric circumstances, etc. These were shown to influence the affective appraisals of environments. Bright colours and pleasant weather conditions positively influenced the appraisal of the urban environments as was shown in Chapters 4.2 and 4.4. The reduction of details generally resulted in a more coherent, clean and also rather dull environment, even when the details were related to negligence and vandalism, as was apparent in Chapter 4.2.

The third column in the framework concerns the “perceived holistic virtual environment”. It is assigned a separate column to indicate that it represents a virtual environment with a unique combination of environmental and representational features, viewed by a specific user, with specific goals and tasks, at a certain point in time.

Viewers generally do not distinguish between features of an environment and features of a representation: they perceive and assess one indivisible virtual environment. Therefore, together columns 1, 2 and 3 compose the Stimulus in Figure 5.2. The participants in our experiments invariably appraised the 3D environment as it was represented, and did not attempt to value the effect of features of the representation on the evaluation of the environment itself. For instance, they did not actively reflect on the possible influence of the low quality representation of vegetation on the ambience of the area, or on the confined field of view on their affective response to the environment. They only assessed features of the representation when prompted specifically with questions about the representation. We mention two examples here: the 3D model of the Carré Theater (Chapter 4.3) for instance displayed textures that were rather nondescript, but participants did not comment on them and easily accepted them as materials appropriate to the building. Participants in the experiment comparing the appraisals of a real urban area in Venlo with a 3D representation of that area did not notice or comment on chosen techniques (in this case photo textures stuck on flat facades) or reflect on the consequences of the representation on the area (Chapter 4.4).

This lack of discrimination holds a risk for the use of visualisations for communicating future designs of an area. Viewers may confuse the represented environment with an accurate view of an area and do not realize for instance the effect of the applied style on the perceived ambience of the virtual area. For training environments this is of course an advantage: trainees must react to the environment instinctively and not wonder for instance about the effects of the adopted style.

The fourth column concerns the “response moderators” on the affective appraisal and affective/emotional reaction to the virtual environments. In our experiments the influence of the mental representation, of attention and engagement, and of the context of use was more prominent than we assumed. Although the lighting quality of the two versions of the theater (Chapter 4.3) was very different, the appraisals of the versions were very similar. In our view the mental representation overrode the actual image on screen, showing the prominent influence of the mental representation on the users’ evaluation. In Chapter 4.7 we described an experiment in which sounds were added to a training environment for leading firefighters. The trainees reacted mainly to the mental representation of the incident that they constructed. The added sounds only contributed
slightly to their assessment of danger. In Chapter 4.8 we showed that cybersickness influences the affective appraisal of a virtual environment.

The response moderators in this column are personal and obviously difficult to address in the development of 3D environments for visualisations and training. User analysis before the start of the design, and guidelines for the context of use can be applied to control the effects to some extent.

The affective appraisal and affective/emotional reaction to the 3D environment, and assessment of the representation, are elements of the fifth column and constitute the main dependent variables in the experiments performed. They may result in a behavioural response (column six), such as approach or avoidance behaviour. In the exam environment showing a highway (Chapter 4.6), we asked participants to cross the highway when they felt it was safe. This was a very direct indication of the level of danger experienced in the environment. In the training for leading fire fighters (Chapter 4.7) the behaviour and performance of the trainees was assessed by an instructor. Although we addressed the behavioural response only in these two experiments, we infer that potentially the behavioural response is an effective manner of gathering information on the affective/emotional response of a user.

We conclude that our experiments corroborated the proposed framework. The arrangement of the determining factors in columns 1 and 2 proved useful, comprehensive, and supportive in designing the experiments. The effects found in the experiments can all be explained by the interaction of the representational modifiers and determinants (column 2) and the response moderators (column 4).

The framework can be used as a new tool for designing virtual environments. It supports identification of features of environments that are relevant for the objectives of the VE, but that until now have not been recognized as a meaningful or important category, such as ambient conditions. The list of representational and response moderators that may influence the effects of these features on the users’ response can be used as a checklist to discuss design choices before implementation.

The experiments included mostly determinants from the second column, that we have named choices in the modeling process. They highlight the importance of deliberate design choices instead of following conventions or trends in visualisations.

The framework can be refined by tailoring it for different types of 3D visualisations and virtual environments, which will reveal the factors that are more relevant for each of these applications.

After constructing and testing the framework, we noticed similarities with factors that have emerged in research as determinants for presence. Presence is defined as “a participant’s sense of being there in a mediated environment, arising from a perceptual illusion of non-mediation” (Freeman et al., 2005). For instance IJsselsteijn (2000) and Freeman et al. (2005) identify the following factors underlying presence: content factors (a broad category including the objects, actors, and events represented by the medium), media form factors, including the extent and fidelity of sensory information, and user characteristics. These categories correspond to the factors 1, 2 and 4 in our framework, indicating a similarity of the user’s response in the form of a feeling of presence, and an emotional response to a virtual environment. We will return to this issue in the final paragraph of this chapter.

We described the process of appraisal of VEs and distinguished a number of decisive subprocesses such as assessing the current core affect and mental representation (Figure 5.1). Furthermore, based on this process description we developed a framework (Figure 3.2) containing factors (features of environments and the representation, and personal and contextual moderators) that influence and modify the affective/emotional response to a VE. This framework appeared to be fruitful in examining the appraisal process and the resulting affective/emotional response under varying circumstances. More generally, we conclude that we constructed a comprehensive framework for measuring affective appraisal and affective/emotional reactions in VEs.
Research Question 4

Which measuring instruments are sufficiently sensitive and accurate to assess the affective/emotional responses to virtual desktop environments?

In the course of this study we used a number of measuring instruments, including verbal scales for affective appraisal, and graphic scales (for instance the SAM) as well as physiological measures for emotional reaction.

Affective appraisal of the virtual environment

The semantic scale of Russell and Pratt (1980) and the SMB-scale by Küller (1991) were developed for assessing affective qualities of environments but are not tailored, nor validated, for virtual environments. They do not distinguish viewers’ responses to the environments from responses to features of the representation, for instance low quality graphics, or the style, such as the lifeless ambience of an environment. This same problem was noted by Rohrmann, Palmer and Bishop (2000), who found that the quality evaluation of simulations was confounded with the perceived appeal of the presented environment, and that thus both aspects need to be measured and to be untangled in subsequent analyses. However, this problem has not been solved yet.

The internal consistency of the dimensions of the Russell and Pratt scale on arousal and valence was satisfactory in our experiments (Chapter 4.5, 4.8, 4.9) and reflected the affective appraisal of users to the perceived, visual, (holistic) virtual environments.

The Cronbach’s alpha (indicating the internal consistency of the factors) of the SMB-scale (used for the 3D models of the theater and both urban environments in Chapter 4.2, 4.3, 4.4) was low in some cases. Changes in language and preference over time may have had effect on some of the adjectives of the scale, and on the internal consistency of the factors. The SMB-scale nevertheless provided useful results, because the factors are more closely related to specific visual attributes of environments than those of for instance the Russell and Pratt scale (Chapter 3.3).

The scales developed for assessing affective qualities are especially suited for assessing the visual qualities of the environment; they therefore capture only a part of the experience of humans in their environment. We added open-ended questions in the experiments, which elicited many adjectives that were not included in either of the scales, but were very to-the-point for indicating the affective qualities of a (virtual environment), such as lifeless, inanimate, and sterile. A similar strategy is advocated by Wergles and Muhar (2009) and Benyon et al., 2006. Even using open-ended questions, participants tend to appraise only visual aspects of the environment, and ignore the other facets of the multimodal experience of an environment. Perhaps users are not used to assessing non-visual aspects of the environment, or consider them as separate from the environment, because they are often volatile and changing.

For assessment of the affective appraisal of a virtual environment it is recommended to use a multiple strategy:

- a semantic scale, for instance the scale of Russell and Pratt, or the SMB-scale by Küller for the built environment; however these scales require adaptation to match modern (Dutch) vocabulary, and validation. Recently Vogels (2008a) developed a new semantic differential scale in Dutch, using two underlying factors “cosiness” and “liveliness”, which is validated for indoor environments. We applied it in a recent experiment (unpublished) with satisfying results.
- questions or scales to assess perceived ambient conditions, if relevant for the goal or the environment, for instance the weather conditions in the virtual environment (Chapter 4.5), time-of-day, season, etc (Chapter 4.2).
- open questions allowing participants to express their appraisal of the environment in their own words. In our experiments the answers included many expressions that were specific for the ambience and style of many virtual environments (such as inanimate, lifeless, etc. in Chapter 4.2 and 4.4).
questions on the perceived quality of the representation, the perceived ease of use of
navigation, and distracting factors. The results may show a correlation with the appraisal of
the environment, as we found in a recent study (Houtkamp, Colijn, & Bouwman, 2009;
Lammeren, Houtkamp, Colijn, Hilferink, & Bouwman, 2010).

Finally, a novel approach (advocated by Wergles & Muhar, 2009) is to elicit
• participant generated reactions to the most notable elements in the environment revealing
which elements in the environment or model determine the observers’ evaluation.
Qualitative questionnaires can be used to elicit responses, which are content analyzed for
similarities and differences in their cognitive, affective and evaluative aspects (see Chapter
4.4).

The proposed multiple strategy is not yet a method to distinguish the assessments of the
environment and the representation, a problem that we identified in for instance Chapter 4.2, where
unpleasant elements in the environment were not considered as unpleasant for the virtual
environment. By asking specifically about the perceived quality of the representation, at least a little
of the assessment of the representation can be teased apart from the appraisal of the environment.

Affective and emotional reactions to the virtual environment

Nonverbal scales such as the SAM (Self-Assessment Manikin) proved to be quick and intuitive
for participants in the experiments, but the SAM-pictures representing arousal sometimes led to
confusion. A second problem was that the SAM was probably not sensitive enough for revealing
subtle response differences. In many of our experiments we examined relatively small differences in
the characteristics of the environment, which probably also resulted in small response differences.

In three experiments we used physiological measures to assess participants’ emotional
response. Positive aspects of these measures were that they could be performed continuously
during the experiments, were not dependent on language, and did not interfere with message
processing (Ravaja, 2004). However, we found that participants were sometimes distracted by the
application of the instruments and felt restricted in their natural movements. The measurement of
facial muscle reactions (Chapter 4.5) provided results that were difficult to interpret and was
possibly influenced by internal reflections of participants. The same problems occurred with the use
of galvanic skin response (used as a measure for arousal, Chapter 4.7). The measurement of heart
rate and cortisol measurement (Chapter 4.9) provided clear results, however in this experiment the
difference in stress level was large, and therefore relatively easy to determine. Although
physiological measures are widely used and provide interesting results, we also know of many failed
attempts which generally do not lead to publications. The notion that physiological measures are
objective and therefore more valuable than subjective, verbal scales is in our opinion too simple.

Tools to assess the affective/emotional reaction to a virtual environment depend of course on
the goal of the research, but in general the SAM (Self-Assessment Manikin) (Bradley & Lang, 1994),
the engagement section of the ITC-SOPI questionnaire (Lessiter et al., 2001) and the state self-report
scale (Ploeg, 2000, derived from the Spielberger State-Trait Anxiety Inventory (STAI), Spielberger,
1985) provided useful results. The questionnaires on perceived risk and perceived danger were
specifically developed for these experiments (Chapter 4.6 and 4.7).

For the assessment of affective/emotional reactions again a multiple strategy is advised:
- a validated scale as a measure for the element of the affective or emotional reaction under
  examination, such as the SAM, or STAI questionnaires.
- open questions allowing participants to express their response to the environment in their
  own words.
- tasks in the virtual environment can be designed to reveal an emotional state. In the
  experiment on the experience on risk on a highway for instance, participants were asked to
cross the highway when they considered it safe for themselves (Chapter 4.6). According to
the approach-avoidance theory, positively evaluated stimuli are inherently associated with
an approach orientation, whereas negatively evaluated stimuli are inherently associated with an avoidance orientation (Elliot, 2006). This can be used to evaluate participants’ navigational behaviour in a virtual environment.

- participant generated reactions to the most notable elements in the environment revealing which elements in the environment or model determine the observers’ evaluation. Qualitative questionnaires can be used to elicit responses, which are content analyzed for similarities and differences in their cognitive, affective and evaluative aspects (see Chapter 4.4).
- questions on the perceived quality of the representation, the perceived ease of use of navigation, and distracting factors. These may be derived from existing questionnaires but they must be tailored to the environment and representation.

So far no measuring instruments have been designed for assessing the affective/emotional response to virtual environments. At this moment we advocate a combination of available quantitative scales and measuring instruments that are used in real environments, with qualitative techniques. Quantitative techniques allow objective analysis of the experimental data. Qualitative techniques provide rich descriptive data that can reveal which elements determine the viewers’ response, and can elicit their judgments on the representation (its quality, style, or distracting elements).

5.3 Answer to the main research question and conclusions

We return to the main research question for this thesis:
How do visual and auditory features in a virtual environment on a desktop environment influence the affective appraisal and affective/emotional reaction of the user?

The answer to this question depends on four basic factors:
(1) characteristics of the user and (momentary) user state, such as experience with navigation and games, knowledge of an environment, and mood;
(2) characteristics of the represented environment, for instance natural or man-made, stereotypical or unknown, and static or dynamic;
(3) the purpose or goal of the 3D environment, used for visualisation or training;
(4) the response evaluated: the affective appraisal of an environment, or the affective/emotional reaction.

The four factors and their relationships, have to be considered simultaneously and interactively in every question on the effect of features in a 3D environment on the user.

1. Characteristics of the user and (momentary) user state

Users’ mood, physical well being (Chapter 4.8), experience and other characteristics may influence the affective/emotional response. The interaction with the other factors in the process can be quite extensive, so we will only provide some examples. The interaction with characteristics of the environment is apparent for instance when previous knowledge of an environment influences the affective response to a 3D environment (Chapter 4.4). Representations of environments that are well-known to users, for instance a modern business center, a quiet suburb, an old (Dutch) city center, or a baroque theater (Chapter 4.3) will easily trigger the place schemas of users that are accompanied by emotions. These personal schemas help to fill in visual details, even ambient sounds and information from other senses, and the ambience (affective qualities) of such places. To assess virtual environments that are less familiar, viewers require more information and details, or they fill in the details from other schemas. A second example is the personal preference of viewers: a preference for quiet natural or exciting man-made environments may result in very different affective responses to the same environment.

The interaction of the characteristics of the user with the goal of a 3D environment is apparent when users have different expectations of training environments than of visualisations. Training
environments usually represent generic types of buildings or environments, such as a hospital or a highway; in visualisations generally a specific, known location is represented. Users have different expectations of the accuracy and verisimilarity of these two types of environments.

2. Characteristics of the represented environment

Both in research and practice, there are many questions on the merits and the validity of environmental visualisations. Answers are sought in technical issues such as the level of detail, photorealism, and on display qualities. We propose that the environments that are represented themselves contain part of the answer, because some types of environments can be modeled more easily and accurately for the required purposes, than others.

In the framework of Chapter 3.3 environmental dimensions and features of real environments are divided into three categories: spatial layout and functionality, signs, symbols and artefacts, and ambient conditions (Bitner, 1992). Firstly, spatial layout refers to the ways in which elements in the environment are arranged, their size and shape, and their spatial relationships. Functionality refers to the ability of the items to facilitate performance and the accomplishment of goals. Signs, symbols and artefacts consist of actual signs and labels, and of objects that give implicit cues to users about the meaning of the place and norms and expectations for behaviour in the place, for instance quality of materials used in construction, or artwork. Ambient conditions include background characteristics of the environment such as temperature, lighting, noise, music, and scent (Bitner, 1992).

Real and virtual environments include elements from all three categories, but not to the same degree (Figure 5.3). Some environments are mainly characterized by elements of the first category, which are physical and static. Examples are urban areas including mainly buildings and roads, or large scale landscape views that also show hills, woods, and other large landscape elements. Static elements can be represented well by modern 3D engines. The static elements generally determine a large part of the affective appraisal of these environments in reality as well as in 3D, although the appraisal of the virtual environment is also influenced by the representational modifiers and the response moderators (Chapter 4.4). Studies that show a strong correspondence in appraisal between real environments, virtual environments, and other surrogates, for instance color slides, have often used environments in which the static elements are dominant, and in which ambient conditions are deliberately kept neutral (Stamps, 2010).

![Figure 5.3 Environmental features (on the left) and their relationship with different types of environments (static, symbolic and dynamic). Elements from the category spatial](image)

165
layout/functionality are predominant in static environments, displaying mainly buildings, roads and other static structures. Symbolic environments are mainly characterized by elements indicating the use by and presence of people. In dynamic environments ambient conditions or other dynamic elements are predominant in the overall ambience of the environment.

Environmental features belonging to the second category (signs, symbols and artefacts), are for instance traffic, behaviour and clothing of people, (street) furniture, indications of vandalism, or negligence. Some of these cues, like naturally behaving humans, are difficult to represent convincingly, or require a high level of detail. Other cues are not predominantly visual. Although on desktop systems sounds can be added to the environment, for instance to indicate human behaviour in an environment, sound design is labour intensive and requires expert knowledge to match the sounds to the visual information. For some environments, olfactory and tactile cues may even be stronger than the visual information. Elements of this category may have an important effect on the appraisal of an environment, and are relevant for many tasks in a virtual environment. Examples of environments that are characterized by features of this type are railway stations, or derelict urban areas (Chapter 4.2). The validity of 3D models of environments like these for assessment of affective qualities depends on whether the relevant cues are included in the model. Some environments are stereotypical, and contain cues that are easily recognized by viewers as for instance a Dutch city centre, a theater, an urban park, etc. Representations of those stereotypical environments easily evoke the ambience associated with these environments in reality.

The representation of ambient conditions (the third category) is even more challenging in virtual desktop environments. The elements in this category are dynamic, and often multimodal by nature, such as atmospheric circumstances, ambient sounds, time-of-day and season. Other dynamic features of landscape elements such as waterscapes can be included in this category as well. Ambient conditions in reality have a strong influence on the affective appraisal of and affective/emotional reaction to an environment. A shopping mall may be a pleasant place in sunny weather in spring, but threatening at three o’clock at night. A suburb may look peaceful in bright conditions, but desolate in rain and fog. For visualisation in public participation, favourable conditions are chosen, allowing good visual assessment of the area; they thus show only one sample, one snap-shot of the area. A valid visualisation should represent the same area in different conditions (for instance Laing, Davies, &Scott, 2003).

For some goals, ambient conditions are very important. Examples already mentioned in this research are visualisations for the assessment of social safety, and training environments for levee patrollers (Chapter 4.5). Graphic software and game engines are capable of simulating ambient conditions, but the validity of these simulations has not been studied often (e.g. Rohrmann & Bishop, 2002). We found an effect in the expected dimension of appraisal of cues for weather and nighttime on the appraisal of users, but the effect was not very strong.

This classification can help developers in understanding the determinant characteristics of an environment and predict whether the development will be time-consuming and costly, and whether it will be a valid representation and a satisfactory solution for the objectives. In visualisation research this approach improves the discussion on validity of visualisations and helps to structure the results of experimental research.

3. The goal of the 3D environment

An important conclusion is that there are no solutions that guarantee the validity of virtual environments for all purposes/applications. The exact goal of the virtual environment must be defined, users and their tasks specified (which may include assessing the affective qualities), as well as the context of use. This leads to exact requirements for that specific case. The environment to be visualised must be analysed in relation to the users’ tasks so it is known beforehand which elements in the environment are relevant. Developers are generally software experts, but are not
environmental psychologists. We introduce a new strategy here based on this analysis: the elements in an environment can be identified as static, symbolic and dynamic, so the environment can be assigned to one of the three main categories. This assists the development process and testing by focusing the effort on the relevant features.

Throughout the development process of a virtual environment, requirements for affective qualities must be distinguished from those related to or supporting cognitive tasks, and tested separately with future users.

3.a. Environmental visualisations

If the goal of an environmental visualisation is to convey the ambience of an area as it will be experienced by visitors, relevant cues must be included. Depending on the environment these include “nuisance objects” (Chapter 4.2, 4.4), and dynamic elements and sounds (Chapter 4.5). The virtual environment should be displayed in different lighting and weather conditions (Chapter 4.2, 4.5). Brightness of colours must not be exaggerated (Chapter 4.2, 4.4). 3D visualisations often offer a bird’s eye or aerial view, but the point-of-view should be on eye level, so the environment is experienced as it would be in reality. As an alternative, a third person point-of-view can be offered, which often permits more accurate estimations of dimensions in the environment (Chapter 4.6).

Although applying these techniques can improve the visualisation, they require testing and fine tuning to obtain the required effects. There are indications in our experiments and other studies (for instance Benyon et al., 2006; Huang, 2009) that virtual environments elicit appraisals that show the expected direction on a dimension or factor of the applied measure, but with a smaller intensity than presupposed. Furthermore, cues that have a negative effect on the assessment in reality may in a virtual environment not have this effect, and even elicit positive responses such as surprise or interest. The impact of negative cues may be smaller for several reasons: limitations of the display make the cue smaller (Reeves & Nass, 1996); limitations of the display make the cue unimodal and therefore lose important information; and when the content is not personally relevant for the viewer, the effective reaction will be less intensive.

3.b Training environments

In the determination of user requirements for training environments, the learning goals of training should be leading. The discussion of learning goals is outside the scope of this research, and only two examples of learning goals are used here to explain the relation of requirements to different training environments: procedural, rule-based training versus “affective” training.

In the training of procedures, the training application requires the trainees to focus on cognitive aspects of task execution. The training environment supports the events to which the trainee must respond. It must not distract the trainee and therefore its ambience can be neutral. An example of this type of training is described in Chapter 4.7. The training for firefighting commanders in that experiment showed that the effort of the trainees was directed on building a mental model of the situation and applying correct procedures for solving the incident. The original training environment was designed for procedural training and showed a neutral décor for a fire. The addition of arousing sounds had no measured effect on the affective qualities of the environment in this study, no impact on the performance, and only little on other responses of the trainee.

A very different category of training environments is used to prepare professionals to perform effectively in a stressful environment, for instance for practicing decision-making skills in highly emotional or dangerous situations. This so-called stress exposure training (Tichon, 2007) has been used successfully across many work domains, e.g. the military, aviation, and health care (Carvalho, Freire & Nardi, 2010). Recently, Tichon et al. (2010) introduced the Affective Skill Theory in the context of stress exposure training. In this theory affect or emotion is not only seen as a determinant of performance, but also conceptualized as a skill that can be trained in preparation of real life.

The events, the visual and the sound effects in the virtual environment may be designed in such a fashion that trainees experience the situation as very dangerous and that they induce an affective/emotional reaction, high on arousal. The effect of cues on a trainee who is experiencing
stress is however a complex process, in which many factors of the trainee, environments and context influence each other (Morris, Hancock, & Shirkey, 2004). In the experiment described in Chapter 4.9 for instance, the stressed participants did not respond to a negative cue (darkness) as expected, probably because their attention was distracted and the cue was not relevant to their emotional state.

Most training environments are situated between the two extremes of procedural and stress-exposure training, and developers and trainees do not have accurately and explicitly defined expectations on the affective response that will be evoked during the training. As with the development of virtual environments for visualisation, the wanted affective/emotional response should be specified in user requirements before development of the training environment.

For training applications, task analysis of the actions to be performed must reveal how the environment influences the actions of the trainees, but also how the environment supports the trainees’ situation assessment and decision making, also known as the situation awareness. The adaptations in the environment for traffic officers (Chapter 4.6) did not significantly influence the trainees’ objective assessments of speed and distances, but did result in an increase in the feeling of risk or danger, and thus in a different subjective assessment of the situation. In the training environment for fire fighters (Chapter 4.7), sounds did not influence the task performance of the trainees, but they affected the assessment of the danger of the situation. These findings are important for the validity of virtual environments used for training and require more attention in research.

4. The response of the user

In Chapter 2 we made the distinction between the affective appraisal of an environment, and an affective/emotional reaction. In our experiments we have found this is an important distinction. An appraisal is a judgment, and does not necessarily lead to an affective or emotional change. In our experiments the manipulations in the 3D environments generally influenced participants’ appraisals, but the affective/emotional reaction of the participants was not easily affected and showed only small changes, as in, for example, Chapter 4.6 and 4.7. Whether the appraisal also leads to an affective/emotional change also depends on the other factors, such as engagement, and the goal and tasks. It is difficult to elicit strong affective/emotional reactions with relatively simple visual and auditory cues in desktop environments of the kind we employed.

In this research project we have shown that the affective/emotional response to a virtual environment displayed on a desktop computer or projection screen is a complex process and is influenced by features of the environment itself, of the representation, by the context of use and the viewer’s characteristics and state. Whether this response is relevant for the validity of the virtual environment depends on the purpose and goal of the application. We have provided examples of the importance of the affective/emotional response as a requirement for virtual environments, which is generally overlooked or neglected. Recognizing the affective/emotional response as an important requirement and including it in the development process, will improve the validity of many visualisations and virtual training environments.

5.4 Recommendations for development and use of desktop virtual environments

Requirements for 3D visualisations and virtual training environments

In the requirements phase of development the objectives of the virtual environment and the tasks of the users in the VE are specified.

- Developers must determine if the objectives concern specific affective/emotional responses of the user, for instance conveying a sense of danger; and then specify both the required response
(in this case a certain level of fear and arousal) and the determinants for the response within the environment.
User tasks may include affectively appraising the area, for instance when users are expected to assess the represented area for its perceived public safety.
- Requirements should include the determining features in the environment for this specific objective; specification of these requirements demands domain knowledge (in this example of public safety) and have to be performed in close cooperation with domain experts.
- The environment to be visualised must be analysed in relation to the objectives and users’ tasks, so to decide which elements in the environment are expected to be important for the user’s response.
- The main elements in an environment should be identified as static, symbolic and dynamic, as explained in the preceding paragraph (5.3), and assigned to one of these three main categories. This assists the development process and testing by focusing the efforts on relevant features.
- The representation of ambient conditions (atmospheric circumstances, ambient sounds, time-of-day and season, waterscapes etc.) in virtual desktop environments is challenging. Because ambient conditions in reality have a strong influence on the affective appraisal of and affective/emotional reaction to an environment, their relevance for the application objectives and user tasks must be considered as a requirement.

**Development**
Through the development process of a virtual environment,
- requirements for affective qualities must be distinguished from those related to or supporting cognitive tasks, and tested separately with future users.
Visualisations by 3D models used for citizen participation can elicit strong reactions: viewers sometimes feel offended by the visualisations because the convincingness of the representation suggests that the shown plans are already agreed on. Viewers also react strongly to minor details in the visualisation that they consider as inaccurate.
- It is recommended to offer viewers multiple versions of the visualisation showing possible variations of elements that are not yet decided on (such as building materials) or that are dynamic (weather and lighting conditions). This allows viewers to construct a nuanced and better grounded appraisal of an environment.

**Use**
3D visualisations are used in communication to eliminate ambiguous elements in a plan, and to reveal information that is absent or incorrect. But representations are always abstractions of reality, and only show a selection of elements of the environment; therefore many gaps are still filled in by viewers individually. At the same time, the way the information is presented does not alert the viewer to be cautious in his or her assessment.
- Therefore, when presenting 3D visualisations for communication or participation organisations should inform the viewers of the limitations of the VE, show multiple renditions as suggested above, and offer other sources of information such as drawings.
Engagement with the environment is an important factor for experiencing emotions toward a virtual environment so the attention of viewers must remain with the VE. Especially when the VE is used for training
- any distraction must be avoided. Distractions include usability problems in the application such as navigational difficulties and inconsistencies between visual information and soundtracks.
- Careful arrangement of the room in which the virtual environment is shown can improve the attention and the overall effect on viewers to a great extent.
- Ambient illumination can diminish the perceived graphics quality of a VE and diminish the visibility of the VE, so should be kept low.
• The viewer or trainee must be positioned on optimal viewing distance from the monitor or projection screen.
• In training settings, the presence of individuals, including the instructor, may divert the attention of the trainee; they should be positioned out of sight.

A final recommendation especially for training environments is to use priming techniques, for instance a task briefing or a preceding story, to prepare the viewer and create a frame of reference (Tortell et al., 2007, Park et al., 2010).

5.5 Limitations of this research and recommendations for future research

The virtual environment

An obvious limitation in this research project (apart from the limitations mentioned in the framework) is the omission of people, or in general living beings, in the studied virtual environments. The presence of humans is an important determinant of the ambience of a place, because of their appearance and behaviour. In virtual environments used for visualisation of real environments such as used in this thesis, humans are often omitted or play a small role because even with advanced game engines modeling is time consuming, and the results are often not satisfactory. In virtual training environments, avatars are included and sometimes even distract the trainee by their clumsy appearance and behaviour. The trainees adapt quickly to the style, especially when the instructors in the training perform the role play well. However, with the further development of game engines the role of human presence in 3D visualisations and training environments will become increasingly important. Research on its influence on the appraisal of a virtual environment requires input from other branches of psychology as well, and needs a dedicated research project.

Another important topic for further research is the effect of nighttime conditions, which is highly relevant for visualisations of urban environments and for expanding the scenarios offered by training environments.

The framework of determinants

The factors we defined as determinants for the affective response to virtual environments show similarities to factors that have emerged in research as determinants for presence. The relation between presence and emotions is a relatively new branch of presence research. Presence is sometimes regarded as a mediating variable between the virtual environment and other affective, cognitive and physiological responses of the viewer. Lee (2004) mentions for instance the effect of presence on arousal, mood, memory, and task performance. Some authors (for instance Västfjäll, 2003) consider the possibility that emotion is determined by presence, following Frijda’s theory that the appraisal of something as being “real” (also if it is taken to be real when they in fact are not), is a necessary condition and determinant of emotion. Riva et al. (2007) found a bidirectional interaction between presence and emotions: the feeling of presence was greater in virtual environments that elicited anxiety and relaxation than in a neutral environment; on the other side, the emotional state was influenced by the level of presence. Baños et al. (2008) and Västfjäll (2003) also found correlations between higher levels of presence and emotions.

Engagement may play a pivotal role in this relation. Engagement (or a similar concept such as involvement) is considered as one of the presence dimensions by many authors (Witmer and Singer, 1998; Lessiter et al., 2001; Lombard & Weinstein, 2009). However, the exact relation between engagement and presence is not precisely defined: depending on the view on presence, engagement is considered as a determinant, a correlate or a factor within presence (Lessiter et al., 2001). The discussion has become more complex since different authors have now redefined engagement (for example Jennett et al., 2008; O’Brien & Toms, 2008). At this point we adhere to a simpler concept indicating a prolonged, effortless attention and involvement.

The close relation between engagement, presence and emotions is not only evident from the abovementioned research, but also from the similarity in factors that determine them. We do not offer here proof on the nature of the relation between presence and engagement and their relation
with emotions; we suggest that engagement is required for viewers to experience presence, and to experience an affective or emotional reaction; vice versa the experience of presence and of emotions may intensify the engagement with the virtual environment. This topic also deserves further attention in research.

**Research techniques and tools**

The research field of affective appraisal of virtual environments is a largely unexplored field. Because of the exploratory character of this research project, the experiments focused on many variables, and we applied different measuring instruments. This broad approach was necessary to gain a comprehensive understanding of affective/emotional responses to virtual desktop environments, and of the relative importance of determinants. As a consequence, in this research project we did not pursue experiments on the same determinants to further investigate or consolidate the results we found.

The results are also not a complete assessment of the proposed framework, and cannot fully explain all relations between the determinants and responses. The number of participants in the individual experiments was sometimes small, especially in the field experiments when professionals participated. However, the total number of participants in the experiments was almost 200, which provides more weight to the observations and conclusions.

We provided suggestions for improved techniques for assessment, but we have not developed new measuring tools. Especially for visualisation purposes, it is important that new verbal measurement scales are developed that use up-to-date vocabulary and distinguish the assessment of the environment from that of the representation.

We have studied applications that are designed for use in real world situations. Some experiments took place in a lab, some were field experiments. Especially in the field experiments, the contextual factors moderated the determinants or responses and influenced the results. Although it made the experiments complex it enriched our knowledge of the processes that lead to affective appraisals or affective/emotional responses. The challenge now is to further develop methods that can handle and evaluate the interaction of a variety of dynamic variables that are common “in the wild “.


Frydenberg, J. N., Lessiter, J., Pugh, K., & Keogh, E. (2005). When Presence and emotion are related, and when they are not. 8th Annual International Workshop on Presence, September (pp. 21–23).


Journal of Environmental Psychology, 3, 43-55.


Grout, T., Taylor and Francis.


Hillsdale, NJ: Educational Technology Publications.


Lawrence Erlbaum Associates, Mahwah, NJ.


Albuquerque, NM: Human Factors and Ergonomics Society.


Epileptic the Proceedings Journal of Biometeorology


Southern Psychologist, 2,6-16.


*Presence*, 8(6), 657–670.


7 Summary

Introduction

Virtual environments (VEs) on desktop computers are used in many domains, including science, heritage, engineering, design, and entertainment. In this dissertation two fields of application, environmental visualisation and training, are brought to the foreground. Interactive navigable 3D visualisations of built and natural environments have become commonplace in design and planning of urban environments and landscapes, and are regarded as potent prototyping and communication tools. They support planning decisions and allow greater involvement of the public by providing representations of urban areas or landscapes that can be easily recognized and understood. In training applications VEs displayed on desktop monitors or on projection screens are used to represent situations and scenarios that cannot be created in the real world for reasons of safety, cost, and time. They can represent generic environments (a hospital, a prison), or specific existing locations.

A valid simulated environment should induce a cognitive, affective and behavioural response in the observer equivalent to the response to the real environment. In many visualisations and training applications, the affective response of users of the VE influences their task performance. Thus, a validated affective response is an essential requirement for an effective application of a VE. The effectiveness of desktop VEs for training purposes often depends on their ability to correctly address the user’s affective experience, next to their cognitive and perceptual experience. Emotionally compelling VE’s are thought to provide more effective training and to result in a higher degree of initial learning and subsequent retention of the lessons learned, when the situation in the real world demands it.

The graphics quality of VEs has increased rapidly over the last years. In the development of 3D interactive environments, a common underlying assumption is that the highest possible level of photorealism ensures a valid representation. The apparent realism of 3D models also convinces users that these are valid representatives of the real environment, to be used for any purpose. However, the VEs are generally designed for spatial tasks or other cognitive tasks, not for assessment of the affective qualities of an environment. They do not contain the required information for that purpose, which can be visual, but also of other sensory modalities (audio, tactile and olfactory). Accurate and detailed visualisations do not necessarily create the same ‘ambience’ of a real environment. Vice versa, features that influence the perceived ambience of a VE do not automatically improve the accuracy of a visualisation. Therefore, the affective response to a VE must be considered and assessed separately during the entire development process, from requirements analysis to use and evaluation. Knowledge to support this process and of the importance of affective responses for the validity of VEs for visualisation and training is at this moment insufficient: research focusing on the affective responses to VEs on desktop computers is very scarce. Chapters 1, 2 and 3 discuss the research domain, theories on affective appraisals of and affective/emotional reactions to real and virtual environments, and develop a framework of determinants for empirical research.
**Research questions and research method**

The main research question in this dissertation is: How do visual and auditory features in a virtual environment on a desktop computer influence the affective appraisal and affective/emotional reaction of the user? This comprehensive question was broken down into 4 sub questions:

1. How do users appraise virtual environments, and how is this different from the appraisal of real environments?
2. How do personal characteristics, momentary user states, and context of use moderate the effect of visual and auditory features on the affective appraisal of a virtual environment used for environmental planning or for training?
3. How can features of environments and the representation, and personal and contextual moderators, be organised into a framework of determinants of affective appraisal?
4. Which measuring instruments are sufficiently sensitive and accurate to assess affective/emotional responses to virtual desktop environments?

In the framework environmental dimensions and features are divided into three categories: those relating to spatial layout and functionality, to the meaning and function of elements of the environment, and to ambient conditions. In a VE, these environmental features are modified by features of the representation, which are inherent to 3D models on a desktop computer. Representational modifiers and determinants reduce and modify information in the represented environment. They are classified into determinants and modifiers related to the medium itself, to the choices in the modeling process, and to the selected platform. User characteristics and momentary user states may also influence the response and are included in the framework. The users’ affective/emotional responses are further divided into affective appraisals, and affective/emotional reactions. The framework contains a large number of potential dependent and independent variables. For our eight empirical - partly quasi-experimental, partly true-experimental - studies we made a selection of evidently important variables, based on observations in several projects and on acknowledged issues in the development of 3D environments. They are described in Chapter 4.

**Experimental results (Chapter 4)**

In the experiment “Decay in an urban environment” (Chapter 4.2), the large number of negative elements such as nuisance elements and physical incivilities introduced in a VE of an urban area only had a mildly negative effect on the user’s appraisal. Although the objects were recognized as elements indicating urban decay, they also added (positive) interest to the rather dull and empty environment, typical of this style of visualisation. So, cues that have a negative effect on the assessment in reality may elicit a different appraisal in a virtual environment. A blue sky and bright colours had a positive impact on the appraisal of the area.

In the study of the effects of lighting on the affective appraisal of the Carré Theater (Chapter 4.3), the results suggest that users are susceptible to differences in lighting conditions when visually comparing VEs, but far less without a reference: the enhanced illumination had an effect on only a few factors of the appraisal of the user. Memories and impressions of this or similar baroque theaters probably induced affective responses that are based on mental representations and not only on the actual represented environment.

The aim of the experiment described in Chapter 4.4 (“Affective qualities of an urban environment on a desktop computer”) was to examine the comprehensive effects of modeling, and representation on a desktop system, on the affective qualities of an urban area. The results indicate
that the style used for this type of modeling, creates an environment that is appraised as more pleasant, more uniform, and more modern than the real environment.

In Chapter 4.5 ("Visual dynamics and sounds in an outdoor environment") sounds and movement were added to a visualisation of a Dutch polder area, most of them depicting severe weather circumstances. These would in real life negatively influence the affective appraisal of the environment and increase the urgency of the situation. In this experiment, we found that sound increased the arousing qualities of the environment, so made the environment less ‘dull’, however sound only made the environment less pleasant when not accompanied by visual dynamics. Visual dynamics increased the pleasant qualities of the environment. Together sound and visual dynamics made the environment livelier. In contrast to what we expected, sounds did not contribute to the user engagement in this environment, at least in this experimental setup.

The goal of the experiment using the exam environment for traffic inspectors was to assess the effect of enhanced visual and auditory depth and speed cues on the affective response of the user, especially the feeling of risk (Chapter 4.6, "Experiencing danger on a virtual highway"). Although the perceived speed of the vehicles was not higher in the enhanced version, and objects were not estimated to be closer, both the assessment of danger and the feeling of risk were significantly higher. This experiment confirms our viewpoint that validity of a virtual environment for cognitive tasks such as assessing speed and distances, does not automatically lead to validity for tasks that involve the affective response of users, and vice versa.

For the study described in Chapter 4.7 ("The effect of sounds on engagement and arousal in a virtual training") we added realistic and task-relevant sounds to a training scenario for fire fighter commanders. The soundtracks did not increase the trainees’ engagement and arousal as we expected. The focus of attention of participants was on the communication with the instructor, and on creating a mental model of the fire and its surrounding areas, which distracted the attention of participants. Whereas the sounds were a minor information source for the procedural tasks in the scenario, participants did consider them important for the assessment of danger, which has a strong affective component.

In Chapter 4.8 ("The influence of cybersickness of the affective appraisal of a virtual environment") we reported on a study of the effects of cybersickness on users of a VE. Participants who experienced cybersickness valued the VE as less pleasant and more arousing than participants who did not experience any appreciable symptoms of cybersickness.

Our final experiment (Chapter 4.9, "Effects of stress and darkness on the appraisal of a virtual environment") investigated the effects of ambient lighting and prior stress on the emotional response to, and on the affective appraisal of, a desktop VE. The dark environment was considered as less pleasant than the daytime environment, but not more arousing. Acute prior stress did not influence the users’ appraisals, nor did it elicit higher anxiety. We concluded that merely lowering the ambient light level in a desktop VE is not sufficient to elicit anxiety, or make the environment more arousing. In this respect desktop VE representations are different from immersive VE representations.

**Conclusions from the experiments**

Although the experiments addressed different variables, we noticed a number of recurring patterns. First of all, features that are considered unpleasant in real environments (such as physical incivilities or severe weather conditions) do not have the expected effects when included in desktop VEs. Although they are recognized as disagreeable, they also seem to make the VE, often considered rather dull and lifeless, more interesting and vivacious. Desktop VEs of the type we studied reduce
the effect of common negative (unpleasant) elements in the environment which limits their applicability as environmental simulations.

Secondly, features in the representation that may be seen as flaws, such as low quality of illumination, and the absence of sounds, do not seem to bother users, or only slightly. Users naturally fill in or add information that is not presented in the 3D environment, and overlook these flaws when their attention is not directed towards them. The perceived realism of a 3D visualisation may create a misleading sense of confidence. Further findings from the experiments are included in the answers to the research questions.

**Answers to the research questions**

Chapter 5 discusses the main conclusions of the research.

1. When users appraise a VE they do not distinguish between their appraisals of the represented environment, of the representation, and of their pre-existing individual mental representations. Mental representations of places contain information about physical, social, organisational, purposive and affective qualities of these places. Together with the perceived 3D model they constitute a composite mental representation. The user assesses this mental representation, which is updated continuously. The mental representation fills in more information than the user is aware of. At the same time, the way the information is presented in a VE does not alert users to be cautious in their assessments. Measuring tools for assessing the appraisal of environments do not reveal the role of the different determinants either.

2. The user’s affective state, induced by other factors than the VE, may influence the appraisal of the VE. The absence of personal involvement, factors that diminish the perceived (graphics and audio) quality of the 3D environment, and factors that distract the attention of the user, attenuate the impact of cues and thereby the intensity of the affective response. The importance of personal involvement and the context of use of a visualisation or a virtual training for their validity are therefore underestimated.

3. The proposed framework includes the important modifiers and determinants of the affective response to a VE and can be refined for specific applications or domains. Moreover, the three categories of features of real environments, relating to spatial layout and functionality, to the meaning and function of elements of the environment, and ambient conditions, can be used to guide the modeling process. They help to identify the determinant characteristics of an environment, to predict whether the development of a VE will be time-consuming and costly, and whether it will be a valid representation and a satisfactory solution for the objectives. Real and virtual environments include elements from all three categories, but not to the same degree. Environments characterized mainly by elements of the first category, which are physical and static, can be represented adequately with modern 3D software. The validity of 3D environments that are characterized by features belonging to the second category depends on whether relevant cues are included in the model. Examples are traffic, behaviour of people, indications of use and negligence in an environment. It is difficult to represent these elements convincingly, and they require a high level of detail. The representation of ambient conditions which are inherently dynamic, and often multimodal by nature, is even more challenging. For some goals, ambient conditions such as atmospheric conditions are very important. Graphic software and game engines are capable of simulating ambient conditions, but the validity of these simulations has not been studied adequately; this research project makes a contribution to this issue and provides possible designs for further studies.
4. So far no measuring instruments have been designed for assessing the affective response to VEs. At this moment we advocate a combination of available scales and measuring instruments that are used in real environments, with qualitative techniques. Quantitative techniques allow objective analysis of the experimental data. Qualitative techniques provide rich descriptive data that can reveal which elements determine the viewers’ response, and can elicit their judgments on the representation (its quality, style, or distracting elements).

We described the process of appraisal of VEs and distinguished a number of decisive subprocesses. Furthermore, based on this process description we developed a framework containing factors (such as features of the environment, representational modifiers and response moderators) that influence and modify the appraisal process in VEs. This framework appeared to be fruitful in examining the appraisal process and the resulting appraisal under varying circumstances. More generally, we conclude that we constructed a comprehensive framework for measuring affective/emotional responses in VEs.

We complete our research with guidelines for the development and use of desktop VEs for visualisations and training applications. Main recommendations are:

- Requirements for affective qualities must be distinguished from those related to or supporting cognitive tasks, and tested separately with future users.
- The main elements in a (virtual) environment can be identified as static, symbolic or dynamic, and be assigned to one of these three main categories to assist the development process. Ambient conditions are important determinants of affective qualities and require special attention.
- When presenting 3D visualisations for communication, participation and decision making, organisations should inform the users of the limitations of the VE, show multiple renditions, and offer other sources of information such as drawings.
- Especially in virtual training applications, distraction must be avoided. Distractions include factors in the application itself as well as the context of use.