MASTER

Researching the basis of ART
comparing the relative effect of auditory and visual natural stimuli

Lochmans, R.H.

Award date:
2022

Eindhoven, March 23, 2022

Researching the basis of ART: comparing the relative effect of auditory and visual natural stimuli

by R.H. (Roelof) Lochmans

identity number 0601495

in partial fulfilment of the requirements for the degree of

Master of Science

in Human Technology Interaction

Supervisors:

Dr. K. (Kynthia) Chamilothori

Prof. Dr. W. A. (Wijnand) IJsselsteijn

Abstract

Restoration from stress is becoming increasingly important, as the consequences of stress put a large socio-economical burden on the public health system. Nature in various forms can help restore from stress and so reduce the load for individuals and society. These beneficial effects have been found for mediated environments, which are particularly promising for stress restoration in conditions without access to real nature, but little is known about the role of modality and of nature scene content on stress restoration. To address this knowledge gap, the present study examines the influence of stimulus modality and of the presence of water on subjective stress reduction and fascination. Furthermore, the concept of soft fascination — among others — is thought to be a main driver for restorative nature experiences. Soft fascination lacks a fixed definition and means to measure this concept in the Attention Restoration Theory and in environmental psychology. The current study addresses both issues and includes a first attempt in composing a novel soft fascination questionnaire.

Using an online participant database, 227 participants were recruited in an online experiment. A three (auditory stimulation only, visual stimulation only, audiovisual stimulation) by three (low water content, medium water content, high water content) factorial between-subjects design was employed. Differences between the experimental condition and a control condition (a progress bar) were tested within subjects. Participants were cognitively stressed in both the control and experimental condition using a Sustained Attention to Response Task (SART). Subjective stress levels were collected before and after each stress induction and after each stress relief. Reported fascination and soft fascination were assessed for each experimental condition.

Contrary to the hypotheses, neither modality nor water content were found to significantly influence mental restoration or fascination. A near-significant effect was observed for water content with a small effect size. When comparing each condition with the corresponding control condition, the medium water level condition showed a small but significant effect on stress relief. For fascination and soft fascination this effect was not present. Examining further each of the nine combinations of modality and water content, a near-significant increase in stress relief with a small effect was found for the audiovisual - water medium condition. Lastly, the soft fascination questionnaire showed promise for future development with a moderate internal consistency.
The limited identified effects might be explained by high levels of noise due to the online nature of the experiment, larger than expected individual differences, a non-normal distribution warranting non-parametric testing and an order effect. Future research is encouraged to use a lab setting and a counterbalanced experimental design to prevent noise and remove confounds, use covariate analyses to identify factors influencing results and limiting the number of experimental conditions.

Keywords: mental restoration, acute stress, environmental psychology, ART, directed attention fatigue, mediated nature, fascination, soft fascination, modal presentation, water, green space, blue space

Table of Contents

1 Introduction

Motivation

Stress is prevalent in modern society and has far reaching consequences for peoples' health and the healthcare system. Natural stimuli have a calming effect on the affective state and nervous system of people when they are stressed (Ulrich, 1981). When calmed with nature – ideally a softly fascinating stimulus – people tend to restore faster and deeper than when they receive disturbing stimuli or no stimulation with nature (Beute & de Kort, 2018). The restorative effect of natural stimuli is even present when real nature is unavailable, as evidence shows that mediated nature can help relieve stress faster as well (Menardo et al., 2019). In environmental psychology two leading theories, Stress Reduction Theory (SRT) and Attention Restoration Theory (ART), explain the restorative effect from an affective and cognitive viewpoint respectively. The mechanisms that are responsible for the restorative effect of natural stimuli are still subject to debate. Attempts were made to describe what makes a natural environment more restorative. Both the modality the stimulus is conveyed in and the content of the scene seem to play a role.

In the literature manipulations of modality were undertaken showing effects on objective as well as subjective restoration (Annerstedt et al., 2013; Wooller et al., 2018). For different levels of water in a stimulus an effect on restorativeness was also observed (White et al., 2010). Examining the differences between auditory, visual and audiovisual presentation can show which modality or combination of modalities evokes the strongest restorative effect. Audiovisual stimuli can produce a more immersive experience than unimodal presentation and could show an additive restorative effect, possibly due to a more encompassing experience that stimulates more than one modality (Annerstedt et al., 2013; de Kort et al., 2006; Wooller et al., 2018). The presence of water is of special value for mental restoration, because it has a large impact on physical survival and psychological health (White et al., 2010). Research that investigates the effect of different levels of water presence or accessibility in a natural scene shows varying results (see section 1.3.3), so it is still unknown what the effect of water on mental restoration is. Testing the effect of different amounts of water in a stimulus can determine if there are differences in restorative effect, and in which direction.

A comparison of effect sizes between different modal stimuli is currently lacking in the literature. Also, the robustness of the restorative effect of nature is not sufficiently tested. Most former research did not employ designs to specifically measure the differences between stimulus properties. Effect sizes for manipulations and their differences were rarely reported in literature (Sullivan & Feinn, 2012), or they were not comparable because of the non-matching setup of within or between-subjects research (Menardo et al., 2019). Moreover, the circumstances under which the manipulations took place were not always well-reported, making comparisons difficult.

Mixed findings were reported in the body of research on the stress relieving effects of nature, especially for the differences between the restorative effects of stimulus modality, as well as different stimulus content (in the form of percentage of water content) in mediated natural environments.

The current research examines the effect of modality and amount of water on mental restoration while extensively describing the de-stressing stimuli (i.e. naturalistic scenes) used to invoke mental restoration, as well as give effect sizes with the findings. The latter makes the current research more easily comparable to past and future research.

To initiate the theoretical background on stress and the effect of stress on people, prevalence and de-stressing are discussed below. Subsequently, the theory on mental restoration is discussed, after which a synthesis of the theories is posited. Lastly, nature and elements stemming from nature that show restorative properties are described, as well as restoration in this mediated environment.

1.1 Stress and stress recovery in scientific research

Stress

Stress is omnipresent and a part of everyday life as well as work situations. One in four office workers in Europe reports experiencing work-related stress for all or most of their working time (Publications Office of the European Union, 2014). In the field of environmental psychology stress in an organism can also be seen as an evolutionary advantage allowing an extra boost of action in an emergency (Ulrich et al., 1991). Starcke & Brand (2012) conclude in their research that stress affects decision making, but stress can be seen as an advantage or disadvantage in terms of outcome depending on the specific task or situation (Starcke & Brand, 2012). A meta-analysis by Byron et al. (2010) supports this and shows that low levels of stress can temporarily increase creativity where large amounts of the same stressor will negatively influence creativity. Moreover, in humans rational decision making is impaired because “the brain regions that underlie intact decision making are regions that are sensitive to stress-induced changes” (Starcke & Brand, 2012). Stress thus has an influence on humans, and prolonged forms of stress can have adverse effects.

Effects of stress on people

The physiological response to stress can vary from person to person (i.e. due to anxiety, impulsivity or extroversion) and is researched very well (Hockey, 1983). On a physical level, stress is shown to impact heart rate (Ekman et al., 1983; Lin et al., 2011) as well as high frequency components of heart rate variability (HF HRV), a measure coupled with parasympathetic activation (Gudi et al., 2019; McCraty et al., 1995). Blood pressure is reported to increase during stressful events (Wooller et al., 2018) and muscle tension increases when a person is experiencing stress (Alvarsson et al., 2010). The increased blood flow is coupled to an increased respiratory rate (Marci et al., 2007). Skin conductance increases, as a person is sweating more (Lin et al., 2011). The hormonal system gets activated as well, leading to higher cortisol levels in saliva and in the blood stream (Annerstedt et al., 2013). Finally, activity in specific brain regions alter as a consequence of stress (Gould van Praag et al., 2017; Martínez-Soto et al., 2013).

On a psychological level effects on social behavior and cognitive capacities can be observed. Stress negatively impacts the ability to make rational decisions (Starcke & Brand, 2012), possibly due to loading of the cognitive system. On the affective side, depression (a result of prolonged stress) diminishes the ability to experience positive emotions and diminishes accomplishments in the work setting. Creativity is also affected, which was historically tested with Duncker's candle task (Isen et al., 1987). Affective as well as cognitive stress can thus influence work performance (Glaser et al., 1999).

Stress can also encompass the subjective experience of being stressed, although stress can also be described as mentally fatigued. Mental fatigue can manifest itself in negative emotions, reduced tolerance for frustration, impatience, reduced performance and increased errors. A lesser ability to take in information, irritability and a larger chance for people taking risks are also reported (Berto, 2014).

Types and prevalence of stress

Stress can be divided into acute and chronic types and be perceived as affective or cognitive stress. Acute stress is short lived and usually occurs due to a sudden external event, where chronic stress can also have an internal component, is built up over time and is longer lasting (McGonagle & Kessler, 1990). Prolonged exposure to stress works against the individual and can lead to sleep problems (Danielsson et al., 2013; Salvagioni et al., 2017). In severe cases of prolonged and unrelieved stress, a person can experience burn-out complaints (Salvagioni et al., 2017). Stress is correlated to depression as a consequence of too much unresolved stress (Wolkowitz et al., 2010) and even a higher risk for cardiovascular diseases (Larkin & Chantler, 2020). High costs are associated with depression and depression related complaints (Luppa et al., 2007). For Australia, depression is estimated to cost AUD 12.6 bn per year (Shanahan et al., 2016). In the Netherlands costs per year are estimated at 3.1 bn euro per year (TNO, 2020). Depression and burn-out related complaints flood waiting lists for psychotherapy (Wooster, 2008) while waiting times for treatment as high as 18 months are reported (van der Molen et al., 2020). Absenteeism from work due to mental or physical illness burdens the social security system, with workplace stress costing USD 300 bn per year in the USA alone (Foy et al., 2019). Work related stress is a growing problem in the western world. Of all work-related illness, 75% is related to stress and burn out type of complaints (van der Molen et al., 2020).

Affective stress concerns emotions, feelings or moods, for example when a person is stressing for a presentation, stress due to trauma or when someone in close social proximity is troubled (Watson & Pennebaker, 1989). Cognitive stress can occur when a person needs to focus or direct attention which in turn induces directed attention fatigue; a central component of Attention Restoration Theory (ART) as described by (Kaplan & Kaplan, 1989). So, when a person is under cognitive stress, directed attention will fatigue over time (Kaplan, 1995). Both described forms of stress have a valence and extension (i.e. how severe a good or bad form of stress is experienced). Stress is usually accompanied by negative feelings and cognitive disadvantages which fortunately can dissipate over time or else be mitigated. ART theory will be further discussed in section 1.2.2.

Stress thus, is a risk-factor capable of inducing large scale problems with health and consequently, problems in the healthcare system. Little stress can give a short performance boost. Too much stress, especially over long periods of time, can lead to serious illness. An area where this is particularly disadvantageous is the workplace. Data on stress related illness shows that the need for stress reduction is high and should therefore be a priority.

Stress recovery

After a stressor is no longer affecting a person, stress levels diminish naturally and stress reactions in the body will return to normal. So by removing a stressful stimulus, stress can be relieved over time naturally. Faster and deeper restoration from stress is possible with the help of nature (Nilsson & Berglund, 2006; Ojala et al., 2019). Hartig et al. (2014) discuss that “natural areas and features can reduce exposure to challenging environmental conditions by increasing distance to stressors and/or decreasing their perceptual salience”. The adaptive resources of a person can restore by distancing oneself from everyday demands with nature, resulting in reductions in fatigue, sadness, self-reported anger and anxiety (Hartig et al., 2014). In literature, relief from stress is described in many different forms. Affective, cognitive, psychological and physiological relief from stress, also called restoration, is identified (Martínez-Soto et al., 2013). Nature is not only perceived as mentally restorative, it is also proven to induce psychophysiological restoration (Annerstedt et al., 2013; Gould van Praag et al., 2017; Ma & Shu, 2018; Valtchanov et al., 2010; Wooller et al., 2018). Research by Shanahan et al. (2016) shows that "people who made long visits to green spaces had lower rates of depression and high blood pressure, and those who visited more frequently had greater social cohesion" (Shanahan et al., 2016). During the course of a week, 30-minute-long visits to outdoor green spaces already reduced the prevalence of depression and high blood pressure with 7 and 9 percent respectively (Shanahan et al., 2016).
Furthermore, nature not only has a restorative effect, but can also deliver a buffering or 'instorative' effect for future stress. In some cases this effect is referred to as immunization from stress (Beute & de Kort, 2014; Gidlow et al., 2016; Leather et al., 1998; Parsons et al., 1998; Wooller et al., 2018). After restoration, people are more productive and can handle stress better. Better handling of pain during restoration is also reported (Diette et al., 2003), as well as better recovery from surgery requiring less analgesics and with lower incidence rates of complications (Ulrich, 1984).

The stress relieving effect of nature can thus not be ignored. Stress can be alleviated using suitable stimuli stemming from nature. Mental restoration from stress has been subject of research for decades, yet still consensus on why nature is restorative is not reached in the field of environmental psychology. Twedt et al. (2019) support the notion that nature has restorative qualities, branched over more than one mechanism: "It is likely that multiple mechanisms underlie the restorative process, and that the impact of a single mechanism will vary by situation, person, and environment". There is still discussion because the exact mechanism behind stress relief is still unknown, although two prevailing theories explain parts of the restorative effect nature has on people.

1.2 Theories on the role of nature as a source of restoration

The two key theories explaining the restorative effect of nature and what elements should be present for nature to be restorative include Stress Reduction Theory (SRT) and Attention Restoration Theory (ART). These theories are not mutually exclusive and are an attempt to explain why nature is restorative, from an affective and a cognitive standpoint respectively. These are currently the leading theories.

1.2.1 (Psycho-evolutionary) Stress Reduction Theory

The Stress Reduction Theory (SRT) by Ulrich (1983) is evolution based and focuses on positive affect which elicits an immediate affective response leading to restoration by stimulation with an environment containing natural elements. Ulrich (1983) also calls this immediate affective response an immediate pre-cognitive response to nature. According to Ulrich (1983), this innate response originates in psycho-evolutionary adaptation in the organism. Ulrich (1983) assumes that attractive nature is mostly perceived visually and induces pleasant affect, holds interest, reduces stressful thoughts and so induces psychophysiological restoration. According to SRT, exposure to a pleasant and unthreatening natural scene reduces negative affect and physiological arousal, increasing positive affect. Attention and arousal are down-regulated (i.e. reducing or suppressing the response to a stressing stimulus) by the natural stimulus, where approach-avoidance behavior is managed in favor of less stress (Ulrich, 1983). Hartig et al. (2014) and Joye et al. (2013) add that in a person experiencing acute stress, nature evokes a rapid positive affect, blocking negative thoughts and feelings while reducing psychophysiological activation, giving a break from stress and arousal.

Ulrich (1983) introduced the existence of preferenda after an adaptation of the work of (Zajonc (1980): “mediated restorative natural elements elicit an immediate and unconscious affective response” (Zajonc, 1980). So, recovery from stress starts when the scene is displayed, before cognitive judgment is possible. From an evolutionary perspective, natural elements like water, food and shelter are important for survival and are thus preferred.

According to Ulrich (1984), the restorative effects of nature fit with SRT theory. Support for the restorative effect of nature was found in the recovery of patients undergoing surgery and after-treatment care. Patients with a natural view from a window recovered quicker and required less pain medication (Ulrich, 1984). Using students, Ulrich et al. (1991) measured the effect of mediated nature and urban environments after induced stress on restoration more rigorously with self-report and psychophysiological measures: muscle tension, skin conductance and pulse transit time to measure stress and heart period to measure intake/attention. Although not all psychophysiological measures showed significant differences in this within subjects design, the effects go in the anticipated direction and show evidence of enhanced psychophysiological recovery. In a more recent study, Diette et al. (2003) used self-assessment after a flexible bronchoscopy procedure and found that the patient group receiving nature images and sounds reported significantly less pain, while no difference in anxiety was measured with respect to the control group.

Ulrich et al. (1991) comments that “both the stressor film and the nature settings elicited high levels of involuntary or automatic attention, which contradicts the notion that restorative influences of nature stem from involuntary attention or fascination”. The latter is implied because involuntary attention due to threatening nature is not restorative. More criticisms on SRT are stated in literature, like Joye & van den Berg (2011), who comment that a rapid response for facilitating food acquisition is not needed. Also, unthreatening stimuli like clouds — a potential source of water — do not offer food or shelter yet are still restorative.

Partial support is found in a study by (Gould van Praag et al., 2017), who used fMRI based research. Compared to the artificial condition, in the naturalistic condition a shift in brain area activation (from anterior to posterior midline; parts of the Default Mode Network coupled to inward-directed focus and outward-directed focus of attentional processing respectively) was found (Gould van Praag et al., 2017). Alongside the changed brain area activation, peak high frequency heart rate variability (indicating an increase in parasympathetic activity) was observed. This is in line with the Stress Recovery Theory of nature exposure. Activation of the Default Mode Network (DMN) is also coupled to mind wandering, which is discussed in section 1.2.3 under research by Williams et al. (2018).

In conclusion, SRT theorizes the effects of stress restoration on stress through an immediate affective response. Empirical evidence shows that natural properties lead to restoration where man-made properties lack that quality, although the underlying mechanism behind the restorative effect is not fully explained with SRT. A cognitive approach to mental restoration is discussed in Attention Restoration Theory.

1.2.2 Attention Restoration Theory

Kaplan & Kaplan's (1989) Attention Restoration Theory (ART) focuses on cognition and the depletion or replenishment of a resource called directed attention. Directed attention was defined by James (1892) as “the ability to control distractions through the use of inhibitory mechanisms”. Joye et al. (2013) expands that definition and states that “Directed attention is a capacity controlled by the central executive, recruited during tasks that require focus and concentration (e.g., proofreading), under voluntary control, demands effort, and requires the inhibition of competing activities and tasks”. Since focus — or conversely the inhibition of competing activities — costs effort, directed attention can fatigue resulting in directed attention fatigue. Mental restoration can then be utilized to overcome the detrimental effects of directed attention fatigue. Replenishment of directed attention happens faster and deeper when a stimulus is presented to a person with directed attention fatigue that adheres to the four concepts proposed by Kaplan & Kaplan (1989).

ART encompasses the concepts of being away, extent (consisting of scope and coherence), compatibility and fascination. According to Kaplan (1995), natural stimuli are particularly suitable for mental restoration, and thus replenishing directed attention fatigue. For a restorative stimulus this translates to the following characteristics:

- provide a sense of being away from normal routine, either physically or psychologically, causing a person to feel freed from everyday cares and obligations;
- offer extent, such that the observer can sufficiently explore the environment (Kaplan, 1995; Kaplan & Kaplan, 1989; Menardo et al., 2019).
- give the participant the perception that the space is safe and thus compatible with one’s goals of seeking restoration and reflection - the correspondence between the expectations of a person and the observed qualities of the environment;
- be fascinating in the sense that the stimulus effortlessly captures involuntary attention of an individual;

Extent is further subdivided into scope — the way an environment affords opportunities for exploration — and coherence — whether an environment is perceived as structured or not. Kaplan's (1995) concept of extent discusses that an environment must be rich and coherent enough to provide optimal stimulation.

The original components of restorative environments from Kaplan (1995) still offer a relevant framework, accompanied by Kaplan's (1995) definition of fascination: "Nature is certainly well-endowed with fascinating objects, as well as offering many processes that people find engrossing. Many of the fascinations afforded by the natural setting qualify as ‘soft’ fascinations: clouds, sunsets, snow patterns, the motion of the leaves in the breeze — these readily hold the attention, but in an undramatic fashion. Attending to these patterns is effortless, and they leave ample opportunity for thinking about other things" (Kaplan, 1995). Thinking about other things is also called reflection; a state in which internally oriented conscious thought about important life issues is used to reflect in a non-goal-oriented manner, and thus process impactful life events. Williams et al. (2018) add to the definition of fascination that "interesting and aesthetically pleasurable environments effortlessly and gently attract attention". Note that fascination is a necessary component, but not the only component needed to make for a restorative experience. Fascination is further subdivided into hard and soft fascination by Kaplan (1995), where Joye et al. (2013) discern an affective valence, attentional bias and effortless attention dimension within the concept of fascination. More on the dimensions that make up fascination can be found in Figure 2 of section 1.2.3.

A large body of research validates ART as restoration theory. Berman et al. (2008) validate ART by objectively showing better performance on a backwards digit-span task and the executive portion of the Attention Network Task. An fMRI study with a between subjects design by Martínez-Soto et al. (2013) tested differences in brain activation between presenting high restorative potential (HRP) and low restorative potential (LRP) environments. The results showed that the "areas more active in HRP than in LRP (a) and LRP than in HRP (b)" indicate a greater distribution of activations for LRP environments than for HRP environments. This implies that brain areas that are related to involuntary attention were more activated when a participant was shown HRP environments. In contrast, brain regions related to directed attention were more active when a participant was viewing LRP environments (Martínez-Soto et al., 2013). Martínez-Soto et al. (2013) verifies that the hypothesized neural pathways are activated when a restorative and a non-restorative environment is viewed. This shows that the tested HRP environments effectively have lower total activation in the brain than LRP environments, which adds to the theoretical validity of ART. The exact brain regions that become activated and their functions in both LRP and HRP conditions fall out of the scope of this paper.

Even though fMRI studies where post stress brain states are compared to post-restorative brain states strengthen the position of ART as the leading theory, constructive criticisms on ART are also present. Joye et al. (2013) pose critical questions on the theoretical evidence underlying ART. The main objection is that ART fails to objectively test its main theoretical predictions. Also, little support is given for restoration being an ancient evolved adaptive response. Joye & Dewitte (2018) argue that the concept of soft fascination is vaguely described, as a fixed definition is not available. Finally, most measured dimensions are measured with self-reported measures which could introduce biases. Joye & Dewitte (2018) continued research on the theoretical basis of ART and conclude that the elements of a natural scene that restore the mental capacity of a person, or which mechanism is responsible for the restorative effect of nature is yet unknown.

No consensus on the definition of (soft) fascination is reached, therefore Kaplan's (1995) definition stated above is not complete or acknowledged. Compelling evidence for the foundations of mental restoration through soft fascination is yet to be produced. More on the concepts of fascination and soft fascination can be found in section 1.2.3. Apart from soft fascination, being away is also a component mentioned by Kaplan (1995) which can be a literal but also a mental construct. The sense of being away from the normal environment can then be a meditative experience, a transportation to a different reality. Mental restoration is multifactorial, so multiple factors influence restorative potential, which indicates that not one singular mechanism is responsible for the restorative effect.

Differences between people are present and the restorative effect that is perceived by people. The following factors influencing restorative potential are also to be considered: "brain areas related with attentional process (bottom up and top down), multimodal sensory integration, episodic memory, topographic orientation, and brain areas related with emotion processing" (Martínez-Soto et al., 2013). When doing fundamental research on mental restoration all these factors should be considered.

When analyzing the characteristics of a natural scene it is yet unclear which environmental psychological theory or which element of the theory explains the restorative effect. Within ART, the components of being away, extent, compatibility and fascination are present. Progressively more research tries to expand on and adapt the original components of ART to alleviate objections to the theory and find compelling evidence for the validity of ART. Much discussion on fascination and its impact on mental restoration is present in the literature. In the next section the current state of the concept of fascination is examined.

1.2.3 Fascination and Soft fascination

The four cornerstones of ART are being away, extent, compatibility and fascination, where fascination is a central component of a restorative experience (Kaplan & Kaplan, 1989; Kaplan, 1995). Fascination is believed to be important to the restorative effect because it attracts people and prevents boredom without having to use directed attention. Joye et al. (2013) consider the fascination component to be the most important source and main driver for improved restoration in people. Therefore, fascination plays a key role in this thesis.

Fascination as component of an environment alone does not guarantee that directed attention can rest and replenish. Even though fascination is an important component for recovering directed attention, the other components of the ART model (i.e. extent, being away and compatibility) are also needed as basis for recovering directed attention (Kaplan, 1995). One of the reasons that restoration is not guaranteed when a person is fascinated is because fascination does not discriminate between soft and hard fascination (Joye & Dewitte, 2018). Thus, to distinguish what is restorative, a division between hard and soft fascination is to be made. As Kaplan & Kaplan (1989) states; “Some fascination is so powerful that one cannot at the same time think of anything else. Soft fascination, by contrast, permits a more reflective mode” (Kaplan & Kaplan, 1989). Cacioppo et al. (2007) described soft fascination as “involuntary attention to inherently interesting and positively toned environmental features and activities”. Note that the wording "positively toned" is also seen in the positive affect dimension of SRT. Also, the immediate and unconscious processing of a natural stimulus in SRT can be seen as involuntary, like in the fascination component of ART. Hard fascination is experienced when an environment offers fascination and extent but falls short as an environment delivering a restorative experience (Kaplan, 1995).

Joye et al. (2013) discusses three dimensions of fascination and stresses the importance of the fascination dimension for mental restoration: "the three dimensions of fascination are thought to be the main drivers of restorative nature experiences " (Joye et al., 2013). Soft fascination requires more attention in research since the definition of soft fascination is not formalized yet. To further investigate the concept and implications of soft fascination, research by Joye et al. (2013), Williams et al. (2018) and Basu et al. (2019) is reviewed below.

Affective valence, attentional bias and effortless attention — Joye et al. (2013)

Joye et al. (2013) expanded on the concept of fascination and suggested that it consists of three dimensions: affective valence, attentional bias, and effortless attention. As discussed below, Joye et al. (2013) criticized most prior literature on fascination and soft fascination, claiming that both topics are discussed extensively in the literature but fixed definitions are not yet available. Moreover, most literature does not contain objective measures to validate the theoretical concepts of fascination, soft fascination and the three dimensions Joye et al. (2013) discuss. Joye et al. (2013) discussed that Kaplan's (1995) "claim that fascination with nature involves these three dimensions is to a large extent based on intuition or derived from introspection-based measurement methods, such as self-reports". Joye et al. (2013) examined the validity of fascination within ART theory and the affective, attentional and effort dimensions with experiments using objective measures. A visual overview of how the components apply in ART can be found in Figure 2.

The affective valence dimension is assumed to attenuate unpleasantness that can follow when thinking about serious (life) issues through positive affect. When a natural scene preceded a neutral Chinese pictograph, the mean affective evaluation was more positive then with a preceding urban scene indicating an affective bias of natural scenes (Joye et al., 2013). The attentional bias dimension is indicated to be a (stimulating) distraction using an unthreatening natural environment which can turn one's attention away from Directed Attention Fatigue (DAF) inducing sources. Participants showed more accuracy and speed at identifying the location of a dot in the nature condition, indicating an attentional bias towards natural environments. Effortlessness in fascination with nature assures that directed attention is no longer being expended and can gradually recuperate. In the urban condition, participants were faster recognizing sequences. In the natural condition, accuracy on a cognitive task was significantly better in the cognitively most difficult task, although sequences were more rapidly recognized in the urban environment. The effort dimension is discussed as valid, with the premise of a sufficiently difficult cognitive task. Joye et al. (2013) validated Kaplan's (1995) theorized components of fascination, although more research to draw more sound conclusions is needed.

The experimental data from the Joye et al. (2013) study demonstrates that people have an attentional bias to natural environments (which is in alignment with the biophilia hypothesis); natural settings are perceived as more aesthetically pleasurable experiences and are more effortless to attend to than urban settings. Joye et al. (2013) conclude their research with the notion that of the affective, attentional and effort dimension, the effort dimension is considered most in ART literature because restorative experiences are believed to be based on the relative effortlessness with which natural environments are mentally processed. It is interesting to note that according to Joye et al. (2013) the affective dimension (describing the attenuation of feelings of unpleasantness) is embedded in the cognitive based ART theory, where pleasant or positive affect is a key concept in SRT.

Mind wandering — Williams et al. (2018)

According to Williams et al. (2018), when restoration is induced through an unthreatening nature experience either soft fascination (with attention dimensions of *effort, intensity, and affective valence*) or mind-wandering can be induced. Kaplan & Kaplan (1989) already make notion of an invitation for the mind to wander next to the concept of soft fascination. In their theoretical paper, Williams et al. (2018) expand on the concept of mind wandering with a definition: "Mind wandering, sometimes referred to as daydreaming, is internally focused thought that is unrelated to the task at hand" (Williams et al., 2018). Williams et al. (2018) give the following definition of soft fascination: "Soft fascination is supported and sustained in environments that additionally provide a sense of being away from everyday concerns and mental routines and devoid of stimuli that might trigger demanding thoughts, that have substantial scope and coherence (together feeding a sense of extent), and that are compatible with intended activities of the individual".

Mind wandering is considered to add to the restorative effect but is not the same as soft fascination; the mechanisms complement each other but are separate concepts that are temporally distinct and believed to work on neurally distinct pathways. As opposed to externally oriented soft fascination, mind wandering induces withdrawal from the immediate environment and the task to an internal focus, which is non goal oriented. Soft fascination, on the other hand, induces mental restoration and creativity benefits "primarily through restoration of a directed attention capability via an externally oriented, effortless form of attention called soft fascination" (Williams et al., 2018).

Williams et al. (2018) seek to incorporate ART and mind wandering into a cognitive dual process model on attention restoration for incubation during a nature experience; "Considering how soft fascination and mind wandering might co-occur during nature experience does not challenge ART". A consistent enhanced creativity as a result of exposure to natural stimuli is reported, as well as more divergent thinking, idea generation and evaluation. After restoration with natural stimuli a person is more creative, which increases relaxation and well-being (Williams et al., 2018). The theory describes the mechanism as gentle shifts between externally oriented soft fascination (unthreatening nature experience — associative unconscious bottom-up processing) and internally oriented mind wandering (reflecting — goal directed conscious top-down processing). Mind wandering is more likely to occur during a low demand, goal-oriented task (Williams et al., 2018).

Considering that the switching between the two mechanisms can provide optimal stimulation resulting in optimal restoration can thus produce an 'additive' restorative effect. Theoretically, the switching could happen involuntary, on an unconscious level. If focus would be put on the switching that would be conscious and goal directed attention, making a switch back to soft fascination less likely.

Williams et al. (2018) do not test the validity of mind wandering in an experiment, which is also criticized by Joye & Dewitte (2018). A theoretical relation between soft fascination and mind-wandering needs further investigation. No comprehensive scientific experiment has systematically examined coupling the concept of mind wandering to cognitive mechanisms (or neural pathways) in the brain yet. Williams et al. (2018) propose research to test the mutually reinforcing pathways by measuring attentional focus, brain network activation, cognitive effects and the temporal progression of these processes across the stages of creativity (i.e. divergent thinking, free association, flexibility and new associations of ideas). The concept of mind wandering as described by Williams et al. (2018) shows remarkable similarities to the concept of reflection (see section 1.2.2) as described by Kaplan (1995). Joye & Dewitte (2018) add that mind wandering most probably causes people to reflect in soft fascinating environments. Kaplan (1995) proposes that inhibition (i.e. suppression of competing activity) is essential for reflection and that soft fascination provides an opportunity for reflection. A link that is proposed by Williams et al. (2018) is that nature boosts creativity through mind wandering, although they fail to explain convincingly that increased creativity explains more of the restorative effect of nature. Future research might be able to scrutinize the nuances between the concepts. In Figure 2 the discussed concepts and their interactions are shown. In Williams et al. (2018), a suggestion for subjectively measuring the switch between soft fascination and mind wandering is given as "the ebb and flow of attention". Research using fMRI techniques by Gould van Praag et al. (2017) couples activation of the Default Mode Network to the state of mind wandering and found that the neural and physiological data support the differences in experience between naturalistic and artificial stimuli.

In resume, (intentional) reflection is goal oriented, mind wandering is non-goal oriented although both are internally oriented. According to Williams et al. (2018), soft fascination is externally oriented while soft fascination and mind wandering are temporally exclusive although both are assumed to invoke a restorative effect.

Mental bandwidth — Basu et al. (2019)

Basu et al. (2019) emphasize that the role of fascination in mental restoration is an interaction between a person's current state (i.e. the continuously variable personal characteristic of mental bandwidth) and the restorative qualities of the environment. According to Basu et al. (2019), attentional effort alone is not sufficient to distinguish between soft and hard fascination, but the type of fascination induced also depends on how much mental bandwidth is available. Mental bandwidth is therefore placed apart from fascination in Figure 2. Mental bandwidth is then defined as "the amount of space available in one’s head for processing", which describes the current state of the person in the restoration transaction (Basu et al., 2019). It is implied that mental bandwidth is a characteristic of a person that can change over time. Basu et al. (2019) state that "Soft fascination requires being effortlessly engaged while simultaneously having some free mental bandwidth for reflection" (i.e. internal processing). In this sense, Basu et al. (2019) does not support nor contest the switching from soft fascination to the state of mind wandering that Williams et al. (2018) describes. Mental bandwidth can be seen as a prerequisite for the type of fascination that is induced with a certain stimulus or environment. Basu et al. (2019) gives an example in that walking in nature (low attentional effort and high available mental bandwidth) is usually perceived as softly fascinating, whereas watching television (low attentional effort and little available mental bandwidth) is mostly a source of hard fascination. The mental bandwidth by attentional effort matrix by Basu et al. (2019), which also presents an example activity for each state, is shown in Figure 1.

	High Attentional Effort (Directed Attention)	**Low Attentional Effort** (Effortless Attention)
High Mental Bandwidth (Available space for reflection)	Soft Focus; Uses Directed Attention, but allows resolution of internal noise *Activity:* Hanging out at home	Soft Fascination; Permits resolution of internal noise *Activity:* Walking in nature
Low Mental Bandwidth (No space for reflection)	Hard Focus; Uses Directed Attention and impedes resolution of internal noise *Activity:* Using a smartphone	Hard Fascination; Impedes resolution of internal noise *Activity:* Watching television

Figure 1: Activities varying in a 2 × 2 space of attentional effort and mental bandwidth

Basu et al. (2019) also state that "Under Effortless Attention, low and high bandwidth equate to the ART notions of hard and soft fascination, respectively". Considering this definition, natural stimuli can then also be perceived as hard fascination when little mental bandwidth is available, which is contradictory with the attentional bias theorem discussed by Joye et al. (2013). Basu et al. (2019) add that “Although many settings and activities that we prefer are in fact less demanding of attentional effort, they may not be equally restorative due to variable demands placed on mental bandwidth” (Basu et al., 2019).

To account for individual differences, a more nuanced view on the interaction between soft fascination and mental bandwidth is proposed. When high mental bandwidth is available, a person can handle or require more stimulation to restore, and when low mental bandwidth is available, a person can handle or require less stimulation to restore; i.e. boredom and overstimulation are to be prevented. To cater for differences in mental bandwidth here should be enough soft attention grabbing elements in a stimulus. Moreover, when low mental bandwidth is available, a person should be able to effortlessly filter out an excess of (details of) soft attention grabbing stimuli, while still perceiving restorative nature. In that case, there is most probably less room for reflection. When high mental bandwidth is available, a person takes in more (details of) stimuli and can also involuntarily switch to the mode of reflection (giving rise to mind wandering or day dreaming), while perceiving restorative nature. The former — more nuanced — view corresponds to Kaplan's (1995) scope and extent criteria, in which there is enough to see and experience to take up space in one's head (Kaplan, 1995). Please note that this view also coincides with optimal stimulation handled in section 1.2.4. Another critical note is that the mental bandwidth theorem is not linked with positive affect, although no negative affect should be invoked. Also, the role of reflection is considered second order in the restoration process by Basu et al. (2019), where the most important restorative effect is reached by a diminished load on directed attention since reflection only reduces the internal noise.

Concluding the definitions and descriptions of fascination and soft fascination, and reviewing the recent literature on both, a synthesis of theories and framework is proposed describing the factors that are argued to influence fascination and soft fascination.

1.2.4 Evidence and synthesis of theories

Research on mental restoration advanced significantly in the past decade, partly formalizing and verifying proposed theories. SRT and ART are both evolution based and are the most popular theories. Both SRT and ART theories involve the concept of complexity, giving rise to the importance of sufficient fascinating details for stimulation. In very broad lines, SRT describes a process of feeling first, and processing after that. ART considers processing first and feeling afterwards. The primary affective dimension as described in SRT is needed to complement ART in the sense that soft fascination describes the affective response which is believed to relieve directed attention fatigue. As Menardo et al. (2019) describe it, SRT and ART are not mutually exclusive.

To recapitulate the theory on fascination, first the definitions need to be clear. Based on Kaplan's (1995) paper, the following definition for fascination will be used:

> Fascination is a mental state that influences the restorative potential of experiences and can enhance recovery from directed attention fatigue.

The concept of fascination is subdivided into hard and soft fascination. An experience that requires directed attention (i.e. goal oriented) is considered hard fascination and is not inherently mentally restorative. An example of hard fascination is watching a soccer match.

Based on above mentioned theories and reviewed literature (Basu et al., 2019; Joye et al., 2013; Kaplan, 1995; Williams et al., 2018) an encompassing definition for soft fascination is formulated:

> An externally oriented experience that is inherently fascinating whilst not being attention grabbing is considered soft fascination. Inherently fascinating stimuli involuntarily hold the attention, but in an undramatic fashion; attending to these stimuli is effortless. This makes the stimulus inherently mentally restorative through means of thinking about other things or an opportunity for reflection.

An example of a soft fascination inducing activity is walking in a natural setting, observing the motion of the leaves in the wind. In Figure 2, a summary of the reviewed papers on fascination and soft fascination can be found, describing which components are present and the dependencies between them.

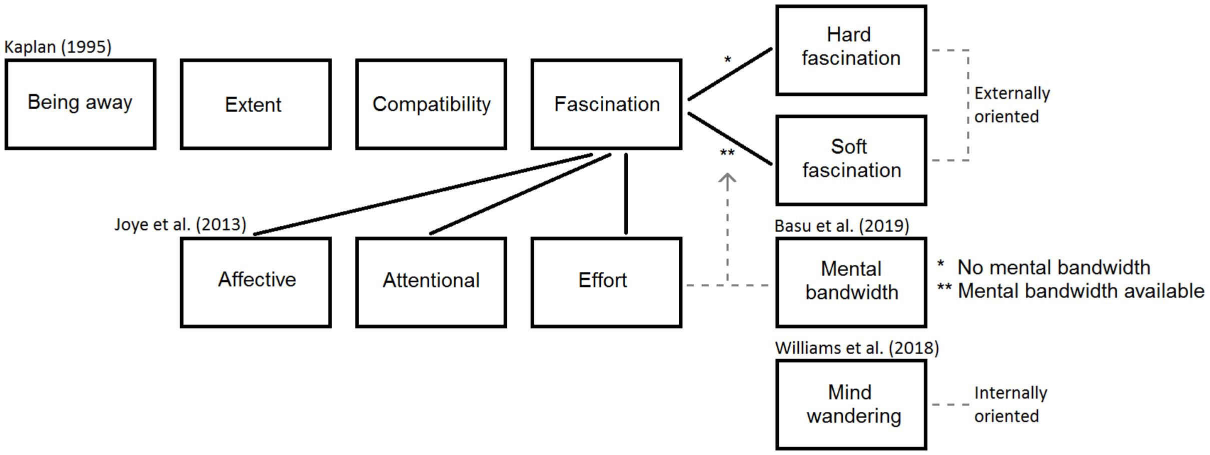

Figure 2: Fascination framework, combining components by Kaplan (1995), Joye et al. (2013), Williams et al. (2018) and Basu et al. (2019)

In Figure 2, the top row describes the components Kaplan (1995) proposed when introducing ART. The second row includes the affective, attentional and effort components by Joye et al. (2013); these components give an expansion on Kaplan (1995) fascination component covering soft fascination only since the induced affect (by nature) is considered to be positive.

Joye et al. (2013) objectively verified the affective, attentional and effort components of fascination in three experiments, where Kaplan (1995) only made assumptions about the characteristics of restorative stimuli. Objectively testing these elements formalizes and adds to the validity of ART, while complementing ART with an affective component as found in SRT. The components Joye et al. (2013) discusses (see Figure 2) also help distinguish between hard and soft fascination inducing stimuli. Another dimension influencing whether a stimulus is experienced as hard or soft fascination is mental bandwidth, as discussed by Basu et al. (2019). The component of effortless attention as discussed by Joye et al. (2013) is dependent on the mental bandwidth that is available, as argued by Basu et al. (2019). Therefore, the components of Joye et al. (2013) and Basu et al. (2019) are on the same height in Figure 2. The definitions seem to accommodate for the notion that individual differences and characteristics of people that perceive the restorative stimuli play a role as well. This can further the field of environmental psychology, as Menardo et al. (2019) suggest that environmental and individual variables that influence the perception of restorative potential should be identified.

As final part of the fascination framework, mind wandering is a theoretical construct that is very close to soft fascination (Williams et al., 2018). According to Williams et al. (2018) the difference was found in internally and externally oriented focus respectively. An unconscious and involuntary switch can be made between the state of soft fascination and mind wandering, presumed to induce an increased restorative effect. Williams et al. (2018) address the difference between goal-oriented and non goal-oriented tasks but do not discuss the difference between internal and external focus, coupling with reflection (Kaplan, 1995) and soft fascination respectively.

Concluding, discussion on the definitions of the components that make for a restorative experience is still ongoing, especially for fascination and the underlying effort dimension as discussed by Joye et al. (2013). A consolidated definition for soft fascination is not available yet. More research is needed to investigate the properties of soft fascination with respect to mental restoration. Explorative research into soft fascination is therefore essential. The theory regarding fascination and soft fascination led to the fascination framework as displayed in Figure 2. The fascination framework is used to answer the research questions in section 1.5. First, the elements in natural stimuli that are found in literature to induce restoration are discussed in section 1.3 below.

1.3 Nature as a source of restoration from stress

Substantial evidence shows that faster and deeper restoration from stress can occur when a person is stimulated with a natural environment (Nilsson & Berglund, 2006; Beute & de Kort, 2018; Ojala et al., 2019). In early restoration research a bias towards natural settings called the biophilia response was introduced by Wilson (1984). This response was assumed for nature in general. More recent research shows relatively consistent patterns for stimuli containing bird sounds, different species of greenery and streaming water (Ratcliffe et al., 2013; Shu & Ma, 2019; Valtchanov et al., 2010). When considering nature as a source of restoration not all natural environments are considered equally restorative; rough terrain for example shows a lower restorative potential than wide open spaces (Parsons et al., 1998). Which natural elements of nature are responsible for the restorative effect is still topic of more research. Menardo et al. (2019) describe characteristics like vegetation, water, caves and qualities such as trees that give a depth of a scene or smooth ground texture of natural environments which can produce affective and cognitive benefits from stimuli stemming from nature. Thus, it is important to examine how the restorative qualities of nature vary across different types of natural environments (Bratman et al., 2015). To be able to measure the differences between modalities (auditory, visual and audiovisual, see section 2.2) and water content, an overview is needed of what makes a stimulus restorative. This is also described on a high level in the four cornerstones of ART (Being away, extent, compatibility and fascination).

1.3.1 Urban and rural environments

Nature is found in several domains, each domain with different qualities. Urban nature and rural nature share characteristics but are also notably different. In urban areas it is more likely to encounter man-made features and other people. In the literature, High Restorative Potential (HRP) and Low Restorative Potential (LRP) elements of surroundings are discussed (Martínez-Soto et al., 2013; Shu & Ma, 2019; Twedt et al., 2019). Usually, urban areas contain more LRP elements. Most people populate urban areas, with an estimate of 55% in the year 2021 (Ritchie & Roser, 2018), and that percentage is rising. Although a large portion of people live in urban areas, natural environments show more restorative potential. Ulrich states that “everyday unthreatening natural environments, compared with most urban settings, should tend to foster greater stress recovery” (Ulrich et al., 1991). When comparing rural nature with urban settings, a natural setting provides deeper and better cognitive restoration, which shows in better performance on a cognitive task in the nature condition with respect to an urban condition (Berman et al., 2008; Bratman et al., 2015). Moreover, rural nature is rated as more restorative than urban nature, like a city park (Twedt et al., 2019). Finally, Menardo et al. (2019) conducted a meta-analysis evaluating 22 studies and found that natural environments are perceived as more restorative than urban environments.

As stated before, urban environments usually contain smaller amounts of, and less diverse, green space than rural areas, usually making them less suitable for mental restoration. In contrast to a large body of mental restoration research, Wilkie & Stavridou (2013) comment that participants with a preference for urban environments show equal perceived restorative potential for urban and natural environments. In the literature, the term “green space” is used synonymously when discussing rural and urban nature. Green space encompasses plant life as well as animals and insects. With regards to mental restoration, green space is to be unthreatening and softly fascinating. Discussion on the term green space did not lead to a set scientific definition of this term (Taylor & Hochuli, 2017).

Ulrich (1983) uses the preferenda that Zajonc (1980) introduced to explain a preference for vegetation and water. The components of preferenda are categorized as (visual) complexity, structural properties, focality, depth, ground surface texture, threat or tension, deflected vistas and the absence of threats. More recent scientific studies give a more detailed view on what is considered restorative. Some of the higher restorativeness of rural nature might be explained by a higher biodiversity that is usually found in rural areas: "biodiversity per se also influences perceptions; people value green space significantly more with high than with low measured biodiversity" (Gunnarsson et al., 2017). Different species of greenery are also suggested to be beneficial for mental restoration (Ulrich et al., 1991; Valtchanov et al., 2010; Sun et al., 2018). The structural complexity of vegetation and visual characteristics of green space is argued to have an influence on stress reduction and mental and physical health (Shanahan et al., 2016). In a Chinese study using only subjective measures, higher biodiversity is rated more positively as well (Wang & Zhao, 2019).

1.3.2 Most effective dose of nature

The duration and frequency of stimulus exposure that lead to restoration should be considered when administering a stimulus. Studies on restoration with mediated nature report exposure time to restorative stimuli in the order of magnitude of several minutes. Examples of restoration times with exposure to a natural stimulus found in mental restoration studies from the literature ordered from low to high restoration time are: 30 seconds stimulus (Shu & Ma, 2019), 40 seconds visual stimulus (Lee et al., 2015), 1 minute stimulus (Sun et al., 2018), 1 minute stimulus (Krzywicka & Byrka, 2017), 85 seconds of video stimulus (Wang & Zhao, 2019), 3 minutes recovery period (Ulrich et al., 1991), 4 minutes and 12 seconds (Hunter et al., 2010), 5 minutes and 6 seconds (Martínez-Soto et al., 2013) and finally 5 minutes 25 seconds (Gould van Praag et al., 2017). The reviewed studies were effective in restoring acute mental stress within several minutes, showing that recovery from acute stress is brief and little time is needed to mentally restore.

In a study by Ojala et al. (2019) a period of 15 minutes sitting in real nature demonstrated to be very effective and showed significant differences on Restorative Outcome Scale (ROS) and Subjective Vitality Score (SVS) with respect to urban environments. An additional period of 30 minutes walking did not change the psychological and physiological measures (Ojala et al., 2019). The most effective dose of nature was researched by measuring at several time intervals. In a meta study by Barton & Pretty (2010) a restoration time of 5 minutes was found to be most effective after acute stress was induced. A restoration time of 10 to 60 minutes was still effective but showed diminishing returns (Barton & Pretty, 2010). This is consistent with the restoration time of around 5 minutes, although the effect of dose seems to depend on whether stressing was acute or not. For prolonged stress, a longer de-stress cycle is most probably needed to obtain the same effect. The duration of dose is further discussed in method section 2.3.3 and 2.3.6.

More research on nature dose investigates the intensity (i.e. the qualities or quantity of nature itself), the frequency and what types or characteristics of nature are most restorative. The qualities of restorative nature are described in section 1.3.1. Jiang et al. (2014) showed that tree density (i.e. the amount of trees in the visual field) influences stress reduction, but only up to a certain point. An inverted U-shaped curve for stress reduction was found with a peak at 35% visual field filled with trees (Jiang et al., 2014). Stress reduction in men (but not in women) was highest when tree coverage was 30%. Lower and higher percentages of tree coverage still led to stress reduction, albeit lower in strength (Jiang et al., 2014). In a study by White et al. (2010), preference and affect were significantly different when comparing green space (Green-Built) and a mix between urban and green space (Built-Green). These results led to a nature dose-response relationship between peoples' preferences and affective responses, and the amount of green space in a scene (White et al., 2010). Herzog (2003) researched the interaction of attractiveness of nature and the perceived restorative potential. The findings indicated that perceived restorative potential is predicted by Kaplan's (1995) components of compatibility and being away. Joye & Bolderdijk (2015) investigated extraordinary nature (with a higher degree of fascination or even hard fascination) compared to mundane nature (soft fascination) and found beneficial effects from extraordinary nature regarding the degree of beauty, awe, and positive mood change. A beauty bias might thus be in effect, apart from how much attention is grabbed by the natural stimulus.

The accessibility or nearness of nature is also a factor influencing the most effective dose of nature. The restorative effect reportedly can be induced with real nature as well as with mediated nature (for example video, sounds or images of nature). Examples of restorative mediated environments are found in literature (Diette et al., 2003; Hartig et al., 1997; Ulrich, 1979). Sun et al. (2018) investigated the differences in restorative effect between unobstructed and mediated nature. A picture projection, nature through a glass pane and an unobstructed view on a Japanese garden were compared for restorative potential using eye movement tracking (Sun et al., 2018). In this study, the way visual information was presented influenced the evaluation of a space. Viewing a Japanese garden in an unmediated way was perceived as most restorative, followed by viewing the same scene through a glass pane, which was found to be second best (Sun et al., 2018). A mediated view through photo projection was perceived as least restorative (Sun et al., 2018). Individual differences can influence the need for nature as restorative environment. The connection between an individual and nature can be measured with the Connectedness to Nature Scale (CNS) for example (Mayer & Frantz, 2004). In an individual showing a low connectedness to nature, green urban environments could be perceived as more restorative then rural nature. The same could hold for mediated environments being perceived as more compatible with the need of the individual. The quality of a simulation when exposing people to mediated environments was researched by de Kort et al. (2006), empirically showing that immersion is influential on objective measures of restoration. In this study, presence was found to correlate with self-reported affect. When restoration with nature was experienced during a walk or run, real nature was preferred over mediated nature (Bowler et al., 2010; Wooller et al., 2018). However, exercise induces arousal which can influence relaxation (Annerstedt et al., 2013). Therefore, nature without physical activity is still preferred for mental restoration research (Kjellgren & Buhrkall, 2010; Sun et al., 2018).

1.3.3 Presence of water

Water is described as interest evoking and higher in stimulation while increasing scenic quality and providing better restorative effects, possibly through biological appeal (Ulrich, 1983; Ulrich et al., 1991). Water is thus considered part of the preferenda. Water sounds have a good restorative effect on self-reported fatigue and annoyance mitigation, as well as result in an enhanced sense of involvement and immersion (Ma & Shu, 2018; Liu et al., 2019). Water sounds also have a positive effect on recovery and even a cognitive measure like better short-term memory (Shu & Ma, 2019). Water could have been overlooked in earlier restoration research and is referred to in recent literature as "blue space" (White et al., 2010). The distinction between "blue space" and "green space" (i.e. rural or urban nature, containing plant life and possibly animals and insects, with a non-threatening nature) is a more recent development in restoration research with inconclusive results, mostly due to low statistical power (Grellier et al., 2017). The presence of water could have unintentionally biased results of earlier research on this topic (White et al., 2010). Considering the primal need for survival, water is described as a necessity (Ulrich et al., 1991). Park et al. (2020) conducted a detailed research on water and conclude that water adds to restorativeness in the rural condition. Adding water to green scenes leads to a higher preference, greater positive affect and higher perceived restorativeness than scenes without water content (White et al., 2010).

Accessible water is mentioned as "first important predictor for restoration potential of a visual landscape, which implies that touching water is much better than just viewing it" (J. Zhao et al., 2018). So an important promoter for restorative potential of a nature environment is easy accessibility to waterscape. Accessible water can also be seen as more attractive, and therefore leads to higher preference ratings (Wilkie & Stavridou, 2013). Accessible water is flowing — or even more preferable, potable — water that is located somewhere it can be collected for drinking. Clouds for example can contain large quantities of water, but the water in them is not readily available. Another example of inaccessible water is murky water from a sewer (Wang & Zhao, 2019). A study by Aarts et al. (2001) uncovered that participants with induced feelings of thirst showed increased attention to, and recall of, drinking related items during an incidental recall task; i.e. thirsty participants noticed a bottle on a table significantly more often than non-thirsty participants.

Urban space, countryside and woodland habitats were compared in a meta study by Barton & Pretty (2010), who found that significant differences were present between the 3 surroundings. In green spaces, the presence of water had an even larger effect on self-esteem and mood (d = 0.29 and d = 0.19 increase respectively) (Barton & Pretty, 2010). Gidlow et al. (2016) investigated differences in salivary cortisol levels, HRV and subjective measures against urban, green and blue environments but did not find significant differences in the objective measures. The subjective measures showed that total mood disturbance (TMD), cognitive function and restorative experience were significantly influenced by measurement time (Gidlow et al., 2016). Photos of scenes containing little water led to significantly higher subjective preference (Mdiff = 1.22, $p < .001$) and affect (Mdiff = .75, $p < .001$) than scenes containing only green. Increasing the amount of water in the scene showed the same effect on preference (Mdiff = .63, $p < .001$) and affect (Mdiff = .56, $p < .001$). Making the entire scene aquatic resulted in significantly lower ratings of preference (Mdiff = .28, $p < .007$) and affect (Mdiff = .29, $p < .006$). In a second study by White et al. (2010), the same effects were found for subjective "Willingness to pay" and "Restorativeness". Note that no stressor was used in this study (White et al., 2010).

Zhao et al. (2018) give indicators on what types of water sounds are restorative and found that "introducing flowing water sound to the landscape with less still water is a better choice to improve the restorative capacity" (J. Zhao et al., 2018). In their classroom lab study, combinations of pictures and sounds were rated for restorative quality with a questionnaire, followed by analysis with regression and correlation. A follow up study by Wang & Zhao (2019) showed that running water sounds were found to improve the aesthetic quality of an environment. To examine which elements in a sound are restorative, Watts et al. (2009) investigated the spectra of sounds generated by water under different circumstances with a small number of participants. Significant differences were found in terms of subjective impact on tranquility. Variable water sounds with relatively more high frequency content were found to be perceived as more tranquil (Watts et al., 2009).

Summary

SRT and ART give an explanation on why nature results in a restorative effect but are subject to debate. Restorative elements of nature are mapped and researched in mostly empirical research using subjective measures. For mental restoration, natural rural environments are generally preferred over green urban areas although both environment types show restorative potential. Non-green space and non-blue space urban environments with many man-made features show the lowest restorative potential. Green space helps the restorative effect and an increase in biodiversity is perceived as more restorative up until a certain maximum, then diminishing returns can be observed (Jiang et al., 2014). The most effective dose of nature can be expressed in terms of duration, frequency, biodiversity and proximity (i.e. unmediated and mediated). Nature should be easily accessible (i.e. less mediation is perceived as better), so it holds attention in an undramatic fashion (Sun et al., 2018). Auditory and visual elements like water help the restorative experience in a softly fascinating fashion (White et al., 2010). The presence of water is also called blue space. Accessible water shows higher restorative potential than non-accessible water (J. Zhao et al., 2018). Like in the tree density results of the Jiang et al. (2014) study, an inverted U-shape was found for the restorativeness of water over the amount of water (White et al., 2010). When experiencing acute stress, a restoration time with 5 minutes exposure time to a natural stimulus proved optimal (Barton & Pretty, 2010), where a restoration time of 1 minute could already show significant results (Sun et al., 2018). Increasing the exposure time beyond five minutes showed diminishing returns. The frequency with which nature is optimally visited is variable between people. Few studies looked at the effect of varying the perception modality and the effect on mental restoration. Clearly, nature has a restorative effect on people, but the underlying mechanism why this effect of deeper and faster restoration is fostered by natural stimuli is still largely unknown. Mediated environments are a good solution for stress restoration when real nature is not available. Research covering modal variations of exposure to natural stimuli show significant differences for restorative potential. Water content in nature improves the restorative experience. Mental restoration with mediated nature is further discussed in section 1.4. Open questions about the underlying mechanism of mental restoration are stated in section 1.5.

1.4 Restoring from stress with mediated nature

Real nature that is in close proximity is not always available. Then, mediated nature can still help to create a restorative experience that alleviates stress without the presence of actual nature. Menardo et al. (2019) comment that "reproduced (virtual) environments are valid representations of real environments, confirming the implicit assumption that exposure to simulated environments produces the same effects as exposure to real environments". Mediated nature is also called simulated nature and can be any display or auralization of natural elements (i.e. video, photographs, sounds), where even smell and rumble have been used when exposing a participant to a mediated nature environment (Annerstedt et al., 2013; Sun et al., 2018; Valtchanov et al., 2010; Wooller et al., 2016). The impact of stimulation through mediated nature instead of real nature is discussed below. An advantage of lab studies using mediated nature is that these studies can effectively be conducted while changing any of the variables and measuring the influence of each variable. Moreover, interactions or additive effects can more easily be measured. One of the researched subjects investigating mediated nature is the influence of modality on restorative potential. Among others, auditory, visual and audiovisual natural stimuli can be effective, the differences in effectiveness are discussed below.

Stimulus modality

The role of stimulus modality on stress restoration is receiving increasing attention in the field of environmental psychology. Olfactory stimulation was researched in mediated environments but only showed a marginal effect in research by Wooller et al. (2016) and Sona et al. (2019), while the effect remained unreported in research by Valtchanov et al. (2010). Most of the early work on mental restoration focused on visual stimulation only, since the visual sense was believed to be the most important sense for mental restoration (Ulrich, 1981). Alvarsson et al. (2010) compared auditory stimuli and found that nature sounds induced a faster recovery from stress. Audiovisual stimuli have been compared in a variety of experiments under different conditions (Annerstedt et al., 2013; Diette et al., 2003; Hunter et al., 2010; Jahncke et al., 2015; Lindquist et al., 2016; Moreno et al., 2018; Park et al., 2020; Sona et al., 2019; Ulrich et al., 1991; Wooller et al., 2018; Zhao et al., 2018). Findings illustrate that overall stimuli stemming from nature are perceived as restorative, although mixed results were found for the restorative impact of auditory and visual stimuli.

Research indicates that the delivery of sound has a larger effect on restoration than visual stimulation (Ma & Shu, 2018). Research by Wooller et al. (2018) shows that post-stressor (Trier Social Stress Test) total mood disturbance and perceived stress were significantly lowered in exercise coupled conditions with nature sounds, nature video and an audiovisual nature condition, but not in an exercise only or resting condition. The experimental conditions were compared as a between-subjects factor, where the difference between control and experimental was analyzed as a within-subjects factor. The restorative effect of the sound only condition was larger than the effect of video only. Audiovisual stimulation showed an additive effect, although not significantly different from the audio or video only conditions. The restorative effects would remain 10 minutes post-intervention.

Research where visual stimuli were more restorative than sound was also found, as Lindquist et al. (2016) researched the interaction between visual terrain type (terrain only, terrain with built form and terrain with vegetation) and sound (no sound, anthropogenic, mechanical and natural). The highest preference ratings were elicited by the combination of vegetation and birdcall and vegetation and no sound. Park et al. (2020) used two visual conditions, a sound or no sound condition and a rural or urban condition. They demonstrated that for psychological recovery a significant difference was present in the rural setting with water sound between the visual only and audiovisual conditions. The objective measures of heart rate and facial muscle tension failed to show significant differences (Park et al., 2020). Zhao et al. (2018) reports an additive restorative potential when adding sounds of birdsong to visuals of a landscape containing a high coverage of plants and natural water. The study by Zhao et al. (2018) compared restorativeness quality of combinations of 20 pictures and five sounds. The interaction between sound and vision was also researched by Jahncke et al. (2015), who found that nature sounds and visuals are more restorative than office sounds and noise sounds. This was both true when showing a natural scene and an open office scene. An interaction is visible between sound and visuals, as well as an additive effect between sound and visual stimulation for perceived restorativeness. Annerstedt et al. (2013) researched stress restoration with a control condition, a virtual forest and a forest with the addition of an auditory stimulus. As stressor, the Trier Social Stress Test was used. The study by Annerstedt et al. (2013) showed that the audiovisual group recovered significantly, where the visual only group did not show a significant effect on objective measures as saliva cortisol and a variety of HRV measures. The control group showed an even lower restoration than the visual only condition. Finally, Deng et al. (2020) carried out a mixed-modal rating experiment and found that audiovisual stimuli are related to better health benefits than visual stimulation only.

With mixed modal studies the salience of congruent stimulus material of the offered modalities was also researched. Stimulus congruency was found to positively impact realism and restorativeness, among others. Carles et al. (1999) compared the effects of congruent sound and images and found that when the modalities match, the resulting stimulus is “rated higher than the mean of the component stimuli”, indicating a possible additive effect of stimulus modality on stimulus preference. Lindquist et al. (2016) investigated differences between visual and acoustic stimuli and found that “sounds and visuals that are congruent receive higher realism and preference ratings while the more incongruent the combination is, the lower the corresponding ratings” (Lindquist et al., 2016). Higher preference ratings can lead to higher affective ratings, which can give rise to a higher restorative power. Sona et al. (2019) endorses that finding and adds that congruent scent leads to greater pleasantness, higher fascination and a higher score on the measure of being away. Wang & Zhao (2019) found a higher aesthetic preference when visual and auditory stimuli were congruent. Kaplan (1995) incorporates some of the previous items in his compatibility dimension, where the stimulus should match what the person wants to do or achieve, although Kaplan (1995) only takes the stimulus properties into account and not the properties of the person.

Why mixed effects for auditory and visual stimuli are observed, even with congruent stimulus material, is still largely unclear. In an attempt to investigate which qualities make mediated nature restorative, studies in the literature have also examined experiential realism and naturalness. A higher level of diversity and complexity — integrating the auditory and visual modalities — have been found to induce a higher restorative potential (Deng et al., 2020). To test an increase in immersion with respect to experiments using pictures or video material and test the effects on restorativeness, Annerstedt et al. (2013) conducted experiments using a VR forest environment as stimulus material. Only the group that was subjected to the VR environment with addition of sound showed parasympathetic activation, leading to enhanced stress recovery and the conclusion that “The higher the immersion of a VR system, the better the restorative potential that can be expected from the mediated natural environment” (Annerstedt et al., 2013). Valtchanov et al. (2010) added smell and even a rumble platform to stimulus presentation with VR glasses and headphones to create a fully immersive VR experience, which resulted in increased positive affect and decreased stress when compared to a control condition. Unfortunately, many studies did not measure experiential realism and it is impossible to compare the level of perceived immersion over studies to draw conclusions about the influence of immersion on restorativeness.

Most empirical research on mental restoration makes use of subjective measures, where objective measures can confirm that mental restoration really takes place. However in most research, objective measures do not show significant differences when comparing different modal presentations. Moreover, the relative effect size of auditory and visual (or other) modalities are not always commented on, making comparisons difficult. Small differences between restorativeness of modalities are to be expected, not large effects. In research that compares urban environments with natural positive environments, the observed effects — or more specifically differences — are usually larger than when comparing only (affectively positive) natural environments. The same holds when comparing two stimuli or modalities in a stimulus stemming from nature.

1.5 Problem statement and hypotheses

Stress affects daily and work life for many people leading to psychological and psychophysiological responses, sometimes with detrimental effects. Under certain conditions nature is empirically proven to relieve stress faster and deeper, in an effective and cheap manner. Mediated natural environments showed to be a good alternative when real nature is not available. Although the two key theories in environmental psychology — SRT and ART — attempt to theoretically explain the restorative influence of natural stimuli on mental restoration, the relative effect sizes of modality and water content on mental restoration, fascination and soft fascination are not documented comprehensively. Also, a measurement device for soft fascination is not available yet.

In the literature regarding mediated environments, little attention has been paid to the relative impact of natural auditory and visual stimuli, which leaves a fundamental gap of knowledge on the mechanisms of technology assisted stress restoration. The role of stimulus modality has been researched before, although effect sizes of comparisons between modal conditions were rarely reported. Fortunately, increasing attention is directed towards the effect of stimulus modality on mental restoration. Auditory, visual and multimodal stimuli are scarcely compared in literature, and with varying results. By investigating the modal differences with well documented stimulus materials and by including effect sizes, increasing insight in what makes mediated nature fascinating and mentally restorative can be obtained. A knowledge gap needs to be addressed for the effect of modal and congruent multimodal presentation of natural stimuli on mental restoration, fascination and soft fascination. Small differences between restorativeness of modalities are to be expected, not large effects.

The positive role of the presence of bodies of water in a mediated restorative nature experience has been demonstrated in a large number of scientific papers, indicating that a restorative effect of accessible water in both auditory and visual modalities is to be expected (see chapter 1.3). An evolutionary survival perspective is mentioned as reason for the restorative effect of water, although no consensus in the literature is reached about why the presence of bodies of water adds to the restorative effect. A knowledge gap exists about what the effect is when modulating the amount of water presence. This knowledge gap will be addressed with the current study by exploratively investigating the effects of water content on mental restoration (stress relief), fascination and soft fascination.

Fascination and soft fascination as discussed in section 1.2.3 are key concepts in ART, and are considered the main drivers of mental restoration by Joye et al. (2013). Joye & Dewitte (2018) add that a lack of knowledge exists about what makes a stimulus fascinating. No consensus on the definitions of fascination and soft fascination has been reached in the literature and a way to measure soft fascination is not available yet. Joye et al. (2013) made an effort to experimentally validate the dimensions comprising the fascination dimension as discussed by Kaplan (1995): affective valence, attentional bias, and effortless attention. An overview of these components is found in the fascination framework as displayed in Figure 2. In an attempt to address the lack of an accepted definition and a lack of a measurement tool for soft fascination, the theoretical concepts of mental bandwidth and mind wandering as discussed by Basu et al. (2019) and Williams et al. (2018) were added to the fascination framework. Definitions for fascination and soft fascination were formulated. A novel soft fascination questionnaire is to be developed, measuring soft fascination using an online questionnaire.

Concluding the former with proper research questions, the research questions of the current thesis are twofold:

1) What is the relative effect size of the restorative response to auditory, visual and audio-visual mediated nature stimuli in stressed participants with regard to mental restoration, fascination and soft fascination?

2) What is the effect of water content on mental restoration, fascination and soft fascination?

To answer the research questions regarding the effect of modality and water content in a stimulus two hypotheses were posed:

Hypothesis 1a	"There is an effect of modality on perceived mental restoration and on fascination".
Hypothesis 1b	"Auditory stimuli induce higher mental restoration and fascination than Visual stimuli".
Hypothesis 1c	"Audiovisual stimuli induce higher mental restoration and fascination than both Auditory and Visual stimuli".
Hypothesis 2	"A higher (accessible) water content in the scene leads to higher mental restoration and fascination".

In addition, considering the aforementioned components of soft fascination (see Figure 2) as discussed in research by Joye et al. (2013), Williams et al. (2018) and Basu et al. (2019) it is of interest to take the first steps in constructing and testing a questionnaire measuring soft fascination. In this thesis, a questionnaire incorporating the dimensions of soft fascination, mind wandering and mental bandwidth is developed and analyzed as an additional exploratory research. The soft fascination questionnaire is discussed in section 2.5.3.

2 Method

2.1 Participants

Participants were recruited via the participant recruitment platform Academic Prolific. In order to be eligible to participate, participants are between 18 and 45 years of age (people in this age range are assumed to have shorter response times), speak fluent English, have normal or corrected-to-normal vision, are not colorblind and have good hearing (i.e. no hyperacusis, ringing or tinnitus). A test for hearing problems is in place, more on this in section 2.4.2. Other criteria included the use of a desktop or laptop computer with Windows 10 (not Linux, Android or Apple) and stereo headphones, the Chrome, Edge or Opera browser and no apps that change the appearance of colors on the screen (i.e. f.lux, Iris or SunsetScreen). These criteria, including participant age, are inquired where possible at the demographics questionnaire, see section 2.4.9. Participants were paid £2,10 (€ 3,29 at the time of running the research) for participating. A total of 207 participants were needed following an a priori power analysis, see section 2.7.1. The full set of demographics questions is displayed in appendix 8.19.

A total of 227 participants (110 male, 113 female, 4 other) enrolled in the experiment. Participants were between 18 and 44 years of age with a mean age of 23,96 (SD = 5.73). All participants were checked to have passed the hearing test and color blindness questionnaire as described in the method section. The country participants resided in was reported by the participant. For brevity, the data is grouped by continent: Africa (3), Asia (4), Australia (7), Europe (170), North America (36), South America (7). Data were collected between December 17th, 2020 and January 18th, 2021.

Of the 227 participants, 92 used a fixed PC setup and 135 used a laptop. Screen size was binned in categories, while giving participants room to give an 'other' response. One participant filled in "1920 x 1080", this was edited to "I don't know". Of the 227 participants, 91 used headphones of the ear bud type where 136 used over ear headphones. The distribution of screen size and the distribution of participants over experimental conditions including gender is shown in appendix 8.19 Table 35 and Table 36 respectively. A good distribution of number of participants over conditions was observed. A separation was made for gender over conditions, showing that also a uniform spread of gender over conditions was achieved.

Modal preference data was collected, but was not included in the statistical analysis of the current study. A preference for "Hearing sounds, for example music" was indicated by 183 participants, 12 indicated to prefer "Looking at pictures, for example photos". Finally, 32 participants answered to have "No preference between the two options".

A final open question was asked to assess issues during the experiment, distractions that occurred or any comments. Comments that influenced results are discussed in section 4.2.

2.2 Experimental design

An experiment was designed as a within subjects design with modality as a between subjects factor with 3 levels (Audio, Video and Audiovisual) and presence of water level as between subjects factors with 3 levels ($Water_{Low}$, $Water_{Medium}$ and $Water_{High}$). The study design is visualized in Figure 3.

Independent variables are the stimulus modality (with Auditory, Visual or Audiovisual conditions) and water level (amount of water present in the scene in both the auditory and visual modality – with $Water_{Low}$, $Water_{Medium}$ and $Water_{High}$ conditions). Dependent variables are the perceived stress level, fascination, soft fascination and scene quality. Stress level is measured before applying the stressor, directly after the stressor and directly after applying the stress relieving stimulus. Fascination, soft fascination and scene quality were only assessed after the experimental condition (i.e. not after the control de-stressor stimulus). See Figure 5 for a timeline of the experiment.

Assignment of participants to experimental conditions is semi-randomized. Participants are not made aware of the existence of other experimental condition groups. Experimental condition groups are available one at a time in Academic Prolific (i.e. Audio - $Water_{Low}$, then Audio - $Water_{Medium}$ and so on). Participants can only enter a group until it is full, usually within 30 minutes of opening a group. 25 participants per group are admitted, with a spread between men and women approaching 50%. After a group is completed, data is checked for non-compliant participants. Those will be removed and that particular experimental group re-opened. This process will be repeated until all groups contain usable data. Participants can technically enter in more than one study, although the prerequisites state this is not allowed. Participants that try to enter more than once will be removed based on their Prolific ID.

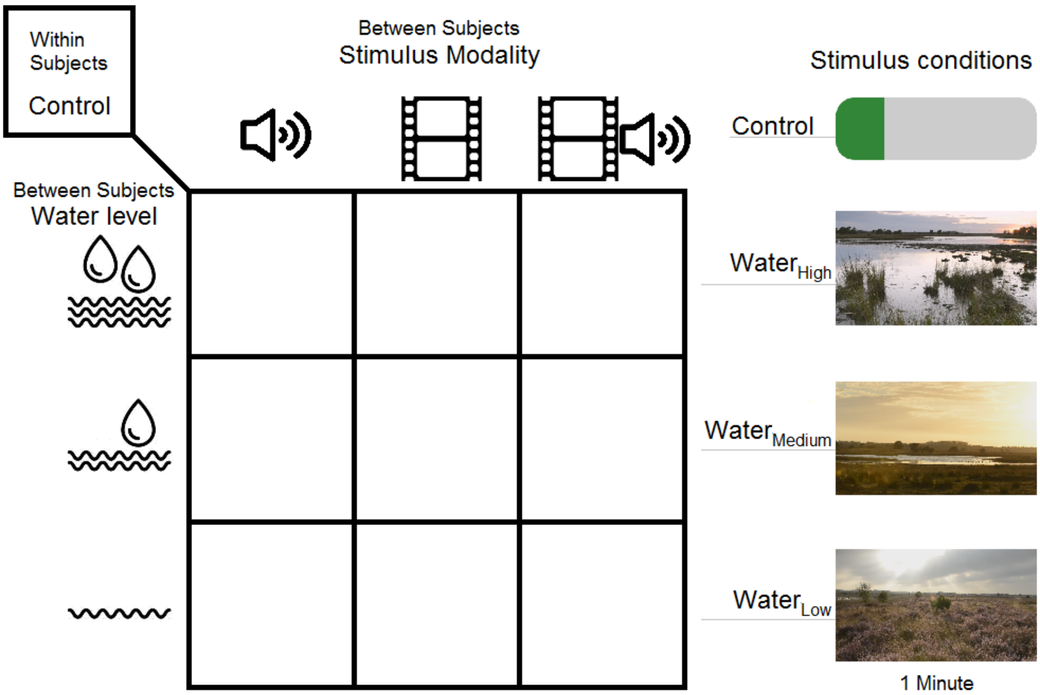

Figure 3: Experimental design — condition matrix

A different experimental design was envisioned at the start of this research. Due to COVID restrictions, another approach was necessary. The preceding setup of the study and the technical details of the design can be found in Appendix 8.1.

2.3 Materials

The experimental design as described above calls for congruent restorative audio and video material that can be separated into audio only, video only and audiovisual stimulus material. The restorative environment is thus of a mediated natural kind. To accommodate for content with 3 different levels of water ($Water_{Low}$, $Water_{Medium}$ and $Water_{High}$), 3 scenes are needed. In literature, guidelines for restorativeness criteria can be found in several papers. Kaplan (1995) and Joye et al. (2013) mention several fascination and/or restorativeness criteria for the stimuli. In section 1.3 the criteria are summarized at the end of the chapter.

Stimuli

Downloading or buying stimulus material was considered but was soon abandoned for practical reasons; many nature videos contain music or narration, image switches, non-natural elements (e.g. car sounds or buildings) or people and other hard fascination inducing elements. The video material was to be uploaded to YouTube for viewing convenience (Google LLC, 2021).

The following sections explain in more detail the main elements that lead to the production of the congruent auditory and visual stimuli. The 9 stimuli are uploaded as video to YouTube (Google LLC, 2021) and embedded in the experiment software. The details are placed in the appendixes and are referred to in the corresponding sections.

2.3.1 Recording

The stimuli used for this study were recorded in a protected nature preserve at the Strabrechtse heide, the Netherlands in the month of August, 2012. A Nikon D3100 DSLR camera was used for video recording in 1080p resolution at 24 fps. An array of microphones (one cardioid and one bi-directional) was used in a mid-side configuration with a Zoom H4n recorder set at 24 bit - 48 kHz. The H4n internal XY microphones were recorded alongside the mid-side microphones. Audio and video were thus recorded separately. Recordings were mostly done at sunrise and sunset when animals were in a relaxed state. The video and audio recording techniques are discussed in greater detail in appendix 8.2 and appendix 8.6 respectively.

2.3.2 Audio editing and loudness matching

Audio recordings needed to be free of car noises, human sounds, domestic animal sounds or any other distressing sounds. Wind noise was present on all recordings, so the most quiet passages were selected and then high pass filtered at 40 Hz to minimize distraction of the participants by wind flutters. All sound editing and montage was done with Wavelab 6 (Goutier, 2007). The sound needed to be congruent to the video, so events visually present in a scene like a flock of geese taking off needed to be audible as well. After the final montages were made and rendered, mid-side matrixing was done with the VST GoodHertz Mid-Side matrix processor plugin (Kerr et al., 2014).

A just noticeable difference of 0.25 dB in loudness is observed in the literature for audio material, following Pierce (1983). To prevent a loudness bias the loudness was matched within 0.1 LUFS (Loudness Units relative to Full Scale) as measured with Youlean Loudness Meter 2 (J. Nikolic, 2020). For a more detailed description of all audio editing, matching and the steps involved, see appendix 8.6 and 8.8.

2.3.3 Video editing, color & lightness matching

Video sections of 1 minute in length were selected from video files initially made and edited for the original setup of this study (see section 1.3.2 and appendix 8.1 for more information on video length and source material). Video editing was done with Sony Vegas 12 (Sony, 2012) software. A total of 6 video scenes were extracted from the material, leading to 4 preliminarily selected scenes used for pre-pilot testing with the aim to select 3 of them for the final experiment. The selection process is explained in section 2.3.5 and 2.3.6.

The human visual system is very sensitive to differences in light, computer displays have properties that need to be taken into account (Kimpe & Marchessoux, 2016). To prevent biases due to differences in lightness (Lakens et al., 2013), a lightness matching over the scene videos was done by measuring and averaging the lightness of the individual video frames. Video frames were turned into a series of images using FFmpeg (Bellard, 2020) and analyzed for Lightness (L) with ImageJ (Rasband, 2019) in the system independent and perceptually correct L*a*b* color space (Mokrzycki & Tatol, 2011). The mean value of all frames was then averaged, giving a lightness level per scene. Another dimension that is to be similar across scenes is the average color. Again, ImageJ (Rasband, 2019) was used to analyze the average value of the A and B component of the L*a*b* color space. An iterative process assured that the best match on lightness, and secondarily color, emerged while respecting the hue of elements like bodies of water in the scene. Lightness was matched as closely as possible; more details about lightness matching can be found in appendix 8.3. Full average color matching was more difficult to achieve since the hue of the water turned more and more un-natural. A trade-off was made and only a slight color correction was administered to the $\text{Water}_{\text{Medium}}$ condition. More details about video color matching can be found in appendix 8.5.

2.3.4 Video and audio rendering

After the audio and video was edited, video and audio needed to be rendered as one video per condition for upload to YouTube (Google LLC, 2021). Audio and Video assembly and rendering was done with Sony Vegas 12 software (Sony, 2012). After rendering, videos for four scenes with three different water levels were available for the current study. Files were rendered as MP4 files with content according to the conditions of the experimental setup. For audio only conditions, a white background was used to prevent image switches between the white background of the LimeSurvey software (Schmitz, 2020) used. Audio was faded in and out over a 2 second period. For video only conditions, no sound was present. A video fade in and out from white lasting 3 seconds was implemented. For the audiovisual conditions, both auditory and visual fade ins and fade outs were implemented.

2.3.5 Scene characteristics & comparisons

After the paradigm shift of changing from a lab study to an online study (see appendix 8.1 for more information), video clips one minute in length were needed for the nine experimental conditions (see section 1.3.2 for more information). The scenes were edited and chosen to reflect as much of the key components of ART as possible; being away, fascination, extent and compatibility. In a pre-pilot test (see section 8.13), the one-minute video clips were then tested for suitability for use in the experimental study using the criteria described in section 1.3. Water content in the stimuli was distributed over 3 levels (Low, Medium and High) in the scene material on all manipulations of modality. The scenes were comparable on most items, while $Water_{High}$ showed highest restorative potential. More details on the selection process and data on the pre-pilot test can be found in Appendix 8.14.

$Water_{Low}$

$Water_{Medium}$

$Water_{High}$

Figure 4: Screen shots of 1-minute video scenes

Apart from the color and brightness matching discussed in section 2.3.3, a good characterization of the used video material can help future research because a better comparison between studies can be made (see appendix 8.4). A characterization is made for water content in the auditory conditions and the amount of water in the visual parts of the scenes in Table 1 and Table 2 respectively. On all other aspects, the scenes are as similar as possible.

The duration of water sound samples was measured and added per scene. In Table 1 the duration and characteristics of the water sounds are displayed. Care was taken to also take the diminished spatiality of water sounds from a distance into account.

Scene	Duration of water sounds (seconds / percentage)	Characteristics
$Water_{Low}$	0 s / 0%	-
$Water_{Medium}$	4.6 s / 7.66%	Water sounds are monaural, level is 7.6 dBFS lower than in $Water_{High}$ and high frequencies are damped to mimic distance effects
$Water_{High}$	21.2 s / 35.33%	All sounds are binaural, water sound level is high

Table 1: Scene characteristics — Water sounds

To objectively characterize the visual amount of water present in the scenes, a measurement of the amount of pixels in the screenshots of the scenes representing water surface was made using Adobe Photoshop CS (Adobe, 2004) as well as the percentage of water content with respect to the total image resolution, shown in Table 2.

Scene	Water amount of Pixels	Water %
$Water_{Low}$	3677	0.28%

$Water_{Medium}$	101951	7.87%
$Water_{High}$	854954	65.97%

Table 2: Scene characteristics — Visual water

The water content roughly shows 3 orders of magnitude with the chosen scenes. This should give a good range with also enough space between the conditions to be sufficiently different. The video links to the stimulus material are found in Table 26.

2.3.6 Pre-pilot test

The aim of the pre-pilot was twofold. One objective was to make a selection between two scenes for the $Water_{High}$ condition. The second objective was to verify that the scenes were comparable on important restorative, attention grabbing, natural and bias inducing metrics for both modalities. Most importantly, the stimuli needed to be restorative and attention grabbing (i.e. soft fascination), but not too attention grabbing as that could induce hard fascination.

After completion of the preliminary stimuli described in section 2.3.4, four remaining videos with accompanying congruent audio were uploaded to YouTube (Google LLC, 2021) and made available privately. Only users with the link were able to watch the videos. Ratings were done by 5 persons (3 males, 2 females) from the social network of the researcher. A questionnaire with 17 ratings from 1-10 and 4 open questions was used for audio, 17 ratings from 1-10 and 5 open questions for video and 6 ratings from 1-10 and 2 open questions were used to compare the preliminary scenes. The dimensions of pretty, naturalness, interesting, spatiality, feeling of being there, relaxation and attention grabbing were inquired, mapping onto the components of Kaplan (1995). The full rubric with original questions is displayed in Appendix 8.13, in Table 29 and Table 30. The means and standard deviations were compared for all items of the questionnaire. The audio and video ratings were very similar on most items. Therefore, a more in depth analysis was made and the most important aspects of fascination and soft fascination were used to determine whether Scene A or Scene F (for screenshots, see Figure 14) was to be used in the pilot study as the $Water_{High}$ scene. The rating items were condensed into the following metrics: pretty, naturalness, interesting, spatiality, feeling of being there, relaxing / gives me peace and attention grabbing. These metrics and their ratings are displayed in section 8.14. A preference for scene F appeared from the addition of number of preferred metrics per scene.

None of the 5 participants that pre-listened, pre-viewed and rated the manipulations mentioned adverse effects regarding the fade in or fade out procedure. All 5 participants commented that the wind-noises were distracting them, leading to a choice to re-edit the audio to reduce wind noise. Those audio files were not re-pilot-tested due to time constraints.

2.3.7 Pilot study

A small sample pilot study (n=30) was used to confirm that the selected stressor task sufficiently mentally depleted — and thus stressed — participants in a measurable and repeatable manner. Also, the de-stressing effect of both control (through progress bar) and experimental conditions (through nature intervention) needed to be verified. If one or both manipulations appeared to be inadequate, a different manipulation was to be chosen and pilot tested. The same procedure as described in section 2.4 below was used for the pilot. However, participants were randomly assigned to groups and videos were randomly assigned to participants (see appendix 8.17). All experimental conditions were completed by between one and six participants per group. More information can be found in results section 3.1.

2.3.8 Experiment and questionnaire software

Since the experimental setup changed from a lab study to an online study, the software used to administer the experiment changed. After careful consideration, a choice was made not to ask participants to install software, but to run the whole experiment in a web browser, and preferably in a single tab. As this choice does have implications for the type of interactions that are available for research, a strength and weakness analysis was made between the available software packages. A functional integration of the stressor task (see section 2.4.5) was a key factor in the selection process. A careful analysis of different online platforms (PsychoPy, LimeSurvey, Millisecond Inquisit and lab.js) to conduct the experiment was made (see Appendix 8.11). Functionality, timing resolution (see Appendix 8.10) and stressor task integration were important metrics for selection. A final choice was made for LimeSurvey Version 3.22.15+200505 (Schmitz, 2020), for which the Eindhoven University of Technology (TU/e) Human Technology Interaction (HTI) department holds a licence. A webserver to run the experiment is available on the TU/e campus, running LimeSurvey (Schmitz, 2020), hosting the SART script and the images needed to run the experiment. Using this TU/e server also covered legally mandatory data protection needs like Algemene Verordening Gegevensbescherming (AVG) and General Data Protection Regulation (GDPR).

2.4 Procedure

In the following sections, all procedures that participants went through during the experiment are discussed in a chronological order. Participants were recruited with Academic Prolific and redirected to a link with the LimeSurvey (Schmitz, 2020) study. The whole experiment was run integrally in a single browser tab. A simple Captcha prevented activity of bots, this was implemented because of the online nature of the study. Participants were then informed about the procedure of the study. Time spent in each LimeSurvey (Schmitz, 2020) page was recorded to examine non-compliance after the experiment. In Figure 5 below, a timeline for the whole experiment is displayed.

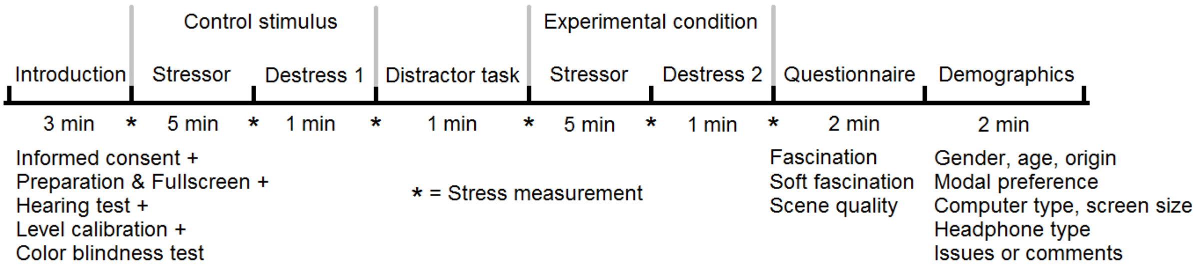

Figure 5: Experiment timeline

In the introduction, participants were explained what was expected of them in text and with visuals. An informed consent form to cover legal requirements was implemented before going to the experiment section.

2.4.1 Experiment introduction

Participants were to read the participant information. After completion, a "Next" button was pressed. Then, an informed consent form was displayed. By clicking the "Next" button, the participant was assumed to have taken notice of all legal information and expectations relevant for the experiment that were put on the participant. The participant information and informed consent form are found in Appendix 8.15 under headings with the same name.

After giving the informed consent, participants were softly drawn into the experiment by giving more instructions. It was expected that participants were sitting in a calm and quiet place behind a desk or table and turned off their phone. Furthermore, participants were to close all other programs to prevent interruptions from messengers, pop-ups, feeds or other distractions and solely focus on the experiment. Also, the browser was set to full screen mode. Accompanying texts as in appendix 8.15 under the heading Experiment introduction and a graphic as seen in Figure 17 and were used to illustrate the expectations. This way, participants were less distracted and could focus even more on the experiment while mitigating biases.

2.4.2 Hearing test and level calibration

To assure that participants had a working stereo headphone setup and normal hearing, a hearing test was implemented. The hearing test was accompanied by a text requesting participants to put on their headphones and adjusting the volume to low. A noise sound with different embedded numbers for the right and left ear was played through means of a YouTube video (Google LLC, 2021).

A multiple-choice question was used to identify the participants with hearing problems or who had problems with their equipment. Only when a participant correctly identified the numbers the experiment continued for that participant. In all other cases, as shown in appendix Table 27, a message about the eligibility criteria was shown. The experiment would end for that participant. The full procedure of the hearing test is described in appendix 8.8.

When the hearing test was passed, a comparable perceived audio level in dB SPL over participants was to be achieved to prevent biases. A similar procedure as described on the hearingtest.online website (Pigeon, 2021) was followed by the participants where the level of rubbing hands in the headphones of the participant was to be calibrated with the level of the participants' rubbing hands. This was an iterative process, where the participant was to take off the headphones and then match the level of the rubbing hands of the recording by adjusting the volume in Windows. When the calibration was complete, the participant could click on the "Next" button. More details on the headphone level calibration can be found in appendix 8.9.

2.4.3 Color blindness test

Participants were only eligible to participate in the experiment if they had normal vision and were not color blind. This distinction between participants was made to prevent biases. The colors of the nature interventions would be perceived similarly to a real natural environment, but the effect of color blindness on restorativeness is not extensively researched. An Ishihara test was considered but advised against because low quality monitors do not have accurate color reproduction. Moreover, some participants might not know they are color blind. Participants were asked to report self-assessed color blindness using a simple question: "Are you color blind?". Answering options were "Yes / No / Not that I know of" in a multiple-choice format. Participants that answered "Yes" were not allowed to continue. More information on the background and the implementation of the color blindness measurement, see Appendix 8.12.

2.4.4 Stress measurement

Stress was measured before stress induction, after stress induction and after stress recovery in both the control or the experimental condition; see Figure 5 for reference where an asterisk "*" marks the moment a stress measurement was implemented. After entering the stress value reflecting the state of stress of the participant, one could only proceed by pressing the "Next" button after a value was given. More details on the implementation of stress measurement can be found in section 2.5.1.

2.4.5 Stress induction

During this experiment acute stress was induced twice with the Sustained Attention to Response Task (SART) procedure as described by (Robertson et al., 1997). The SART procedure was successfully used in several studies that were reviewed, such as Robertson et al. (1997), Berto (2005), Lee et al. (2015) and Shu & Ma (2019), needs very little explanation and is designed to cognitively fatigue directed attention. The SART procedure was also chosen for its cognitive nature, no affective stress was to be induced. The neural path of top-down processing is depleted with the SART procedure. The original description of the SART procedure as used by Robertson et al. (1997) was followed, including the number of presented numbers and target number. The script forced full screen mode in the browser to prevent distractions during this timing-sensitive task. Also, the mouse cursor was hidden during the SART test to heighten the focus on the keyboard actions needed to do this task.

Before the SART task was started, a short text was displayed in LimeSurvey (Schmitz, 2020): "In the next screen you will be asked to perform a task. This task will take about 5 minutes.". The text was displayed to be considerate of what was asked of the participants, since the SART task is rather demanding. Pressing the "Next" button would start the SART script with an overlaying iframe with a different style than LimeSurvey (Schmitz, 2020). Information was given in white letters on a black background. After the information was read, participants could start the SART practice round by pressing the "b". The same procedure as described in the Robertson et al. (1997) paper was then followed. A practice round started in which participants were to press the "space" bar when a number that was not "3" appeared on screen; the number "3" is referred to as the "target". During the practice round, feedback was given with the words "CORRECT" and "INCORRECT" in green and red color respectively. After the practice round with 18 numbers (of which 2 were targets), the full test with 225 numbers (25 targets) was administered after pressing the "b", leading to a total of 243 displayed numbers. No feedback was given on the user's input while sequentially displaying the 225 numbers. More detailed information on the stressor selection process and implementation of the SART script in LimeSurvey (Schmitz, 2020) is given in Appendix 8.10.

2.4.6 Stress recovery intervention

After each stress inducing manipulation, a consecutive de-stressing action was implemented. After the first trial, a control

Figure 6: Control de-stressor

stimulus only containing a progress bar filling up was displayed as shown in Figure 6. The color of the progress bar matched the green theme of LimeSurvey (Schmitz, 2020). No sound was present during the progress bar visualization. Before the progress bar was shown, a text was displayed: "In the next screen you have a short break. Relax while you keep looking at the screen. You will be taken to the next screen automatically". This text was implemented so participants knew what to expect and did not perceive a lack of control.

After the second stressor, one of the 9 experimental conditions (see Figure 3) was played in full-screen. The same text as earlier was displayed before continuing with the experimental condition.

2.4.7 Distractor task

Between the first de-stressor and the second stressor, a distractor task was implemented. At first, the fascination and soft fascination questionnaire were implemented at that place in the experiment. That deemed impractical, since there was nothing fascinating to see or hear when viewing the progress bar. The distractor task was needed since there was no time between the first de-stressor and the second stressor; a second measurement of stress level would then be ineffective.

Participants were instructed in the full-screen video to close their eyes and focus on their breathing with white text on a black background: "Please close your eyes and focus on your breathing. You will hear a sound when the next task starts.". At the end of the distractor task the sound of a bicycle bell was played to alert participants that the task was over.

2.4.8 Questionnaires

Multiple questionnaires were provided in this study. After the second stressor and concurrent nature intervention these questionnaires were implemented, as visualized in Figure 5 above. A fascination and soft fascination questionnaire were used to assess different qualities of the natural stimulus, where a media quality question was used to measure if biases were present. Please find the details on these questionnaires in section 2.5.1 to 2.5.4.

2.4.9 End of study — demographics and debriefing

At the end of the experiment participants were asked to answer demographic questions. More in-depth information is found in section 8.19. To inform participants that the experiment was finished and to thank them for participating, a debriefing text was displayed. After evaluation of data validity criteria as described in section 2.6.1, participants were compensated through Prolific. The debriefing text is displayed in Appendix 8.15.

2.5 Measures

The variables measured in this study are presented below. The goal was to measure the effect of a stress inducing task and consecutive de-stressing task on stress level in a control and an experimental condition. The measures are transformed and reduced in section 2.6. The aimed for statistical analyses are described in section 2.7.

2.5.1 Stress level measurement

To measure stress, a high resolution, numberless and stepless scale ranging from "Not stressed at all" to "Very stressed" is preferred. Participants would then not be distracted with interpretation of numbers or a scale ranging from 0-100 or -100-100, possibly biasing the participant. LimeSurvey (Schmitz, 2020) did not permit the use of a slider without numbers, so a slider with values from 0 to 100 was used to measure the stress level of participants. This procedure follows the Twedt et al. (2019) study, and their implementation of the measurement of perceived restorative potential, visual appeal, naturalness and presence of people.

In this study, the accompanying text was "Please rate how stressed you feel right now:". At the start, the slider was placed in the middle position, displaying a value of 50. A value of 0 corresponded to "Not stressed at all", where a value of 100 was linked to "Very stressed".

2.5.2 Fascination questionnaire

A subjective measure of the characteristics of the de-stressing experimental intervention was implemented. The fascination component from the Perceived Restorativeness Scale (PRS) questionnaire (Hartig et al., 1997) was used to measure fascination, a key component in ART.

For the fascination questionnaire a Likert scale ranging from 0 to 6 was used, following the original PRS procedure. An introductory text with an explanation was used for the fascination questionnaire: "Please rate your experience on the following criteria. There are no good or bad answers, please give the answer that best reflects your feeling about what you experienced. 0 = Not at all; 6 = Completely". To mitigate biases in the form of order effects, the items were presented in a randomized order.

The fascination items from the PRS are displayed in Table 3.

PRS item #	PRS Fascination questions
1	The setting has fascinating qualities
2	My attention is drawn to many interesting things
3	I would like to get to know this place better
4	There is much to explore and discover here
5	I want to spend more time looking at the surroundings

Table 3: PRS Fascination subscale

The questions should be applicable to auditory, visual as well as audiovisual modality presentations and consecutive experience. An issue that needs addressing when using the PRS is that item 5 explicitly states visual aspects of a stimulus: *"I want to spend more time **looking at** the surroundings"*. Since the current research aims to find differences between restorative potential of auditory and visual stimuli, a modification to also cater for auditory stimuli was needed. Item 5 was used in an adapted form, changing "looking at" to "exploring". To mitigate biases in the form of order effects, the items were presented in a randomized order. More details on the fascination questionnaire can be found in appendix 8.18.

2.5.3 Soft fascination questionnaire

Unfortunately, a validated questionnaire for soft fascination is not readily available in the scientific literature. Literature research and recommendations from papers lead to the fascination framework as described in section 1.2.4, and ultimately to the soft fascination questionnaire as presented in Table 4 below.

The PRS is a measure of restorative quality in environments, but a note should be made that the PRS contains elements measuring fascination, most of which do not explicitly discriminate between hard and soft fascination. Some items could lean more towards hard fascination or could be interpreted that way. Measuring soft fascination is thus not possible with the fascination scale of the PRS. Therefore, an analysis was made of the definition of soft fascination in literature, what elements it is comprised of and how that translates to a questionnaire that is easily administrable as well as usable for visual, auditory and audiovisual stimuli. The body of literature that gives indices for what is softly fascinating is discussed in depth in introduction section 1.2.3 and formed the basis for the soft fascination questionnaire. Joye et al. (2013) discusses the affective valence, attentional bias, and effortless attention dimension. The concept of mind wandering is an addition to soft fascination proposed by Williams et al. (2018). Lastly, Basu et al. (2019) proposes available mental bandwidth as a factor influencing soft fascination. To summarize:

Joye et al. (2013) examined the validity of fascination within ART theory and the affective valence, attentional bias and effortless attention dimensions that are fundamental building blocks of fascination. In their paper, Joye et al. propose questions to measure the affective dimension: “The scene was pleasant”; and the effort dimension: “There are many things here that attract my attention effortlessly” (Joye et al., 2013). The effort item is covered in PRCQ item #4 (Pals et al., 2009) and proposed by Joye et al. (2013) for measurement of the effort dimension. The affective and effort items are reflected on in the soft fascination questionnaire. The effort dimension is already measured in the fascination scale of the PRS and thus not reflected in the soft fascination questionnaire.

The concept of mind wandering (also called day dreaming) is introduced by Williams et al. (2018), next to soft fascination. Non goal-oriented mind wandering induces withdrawal from the immediate environment and task. In their paper, Williams et al. (2018) describe the experience of going from externally oriented soft fascination to internally oriented mind wandering and vice versa as an ebb and flow of attention. This text was adapted and added as item to the soft fascination questionnaire: “I experienced an ebb and flow of attention”.

Basu et al. (2019) add the concept of mental bandwidth by combining the dimensions of effortless attention and attentional bias into an attentional effort element. The concept of mental bandwidth that is proposed adds the premise of sufficient cognitive space in a persons’ head to process information. Adding the “I could still think of other things” question Basu et al. (2019) suggests gives an answer to the mental bandwidth dimension.

Check questions were developed and added to control for negative emotions and the theoretical opposite of soft fascination, which is hard fascination, since the PRS does not discriminate between hard and soft fascination. Negative affective states like irritation are usually mitigated by restorative nature experiences as they “involve a recovery from depleted cognitive resources and/or undo negative psychophysiological states” (Joye & Dewitte, 2018). Moreover, fascination with nature is considered “to have positive affective valence by being an aesthetically pleasurable experience” (Joye et al., 2013). The negatively coded item “I felt irritated” was added to the questionnaire.

Furthermore, if hard fascination was induced by either the control stimulus or experimental stimuli, a control question needs to measure exactly that. In the paper by Joye & Dewitte (2018), the Grand Canyon that provokes an awe response is mentioned as a hard fascination inducing environment (Joye & Dewitte, 2018). To control for an awe response to the carefully designed experimental stimuli that is to be avoided, the item “I was in awe” was added to the questionnaire.

The aforementioned items are all based on literature as summarized in Figure 2 and comprise the current soft fascination questionnaire. The attentional bias dimension Joye et al. (2013) discuss is already covered in the PRS, item number two: “My attention is drawn to many interesting things” (n.b. in the Hartig et al. (1996) paper this is item number four). It is therefore not added in the questionnaire, since the fascination part of the PRS is administered on the page directly before the soft fascination questionnaire. The complete soft fascination questionnaire is found in Table 4, accompanied by the measured dimension and reference to literature.

The same introductory text with explanation as with the fascination questionnaire was used for the soft fascination questionnaire. To match the seven point scale of the PRS and to not confuse participants, the soft fascination questionnaire was also administered using the same seven point scale. The soft fascination questionnaire was not pilot tested due to time constraints. To mitigate biases in the form of order effects, the items were presented in a randomized order.

In Table 4 below, the soft fascination questionnaire is displayed, along with the dimension from the framework that is measured.

Item #	**Soft Fascination questionnaire**	**Dimension**	**Reference**
1	The scene was pleasant	Affective valence	Joye et al. (2013) - study 1
2	There are many things here that attract my attention effortlessly	Effortless attention	Joye et al. (2013) - study 3
3	I experienced an ebb and flow of attention	Mind wandering	Williams et al. (2018)
4	I could still think about other things	Mental bandwidth	Basu et al. (2019)
5	† I felt irritated	Check question: Negative emotions	Joye & Dewitte (2018)
6	† I was in awe	Check question: Hard fascination	Joye & Dewitte (2018)

Table 4: Soft Fascination questionnaire items; † indicates an inversely coded item

The soft fascination items measure more of the affective, or emotional elements of soft fascination. The soft fascination items contain the words "felt" and "experienced", concerning more of the emotional and introspective experience the participant has. In contrast, the PRS Fascination scale described in Hartig et al. (1996) examines what the participant perceives or sees. The soft fascination questionnaire is also a means to determine if, and if so why, something is softly fascinating (Joye et al., 2013).

2.5.4 Media quality

To measure the aesthetic quality of the scene, a media quality check question was added. The aim was to be able to distinguish if a beauty bias was present. Factors at play could be a scene that is more beautiful, or otherwise has softly fascinating qualities. The question asked to the participants was: "Please rate the media quality of the scene you just experienced. There are no good or bad answers, please give the answer that best reflects your feeling about what you experienced. 0 = Low quality ; 6 = High quality".

2.5.5 SART errors and timings

The SART procedure is the only objective measure in this research. The presented numbers and accompanying reaction times translate to a measurement of correct omissions, omission errors, commission errors and correct omissions along with their timing. A mean time for omissions and mean time for commissions can be calculated from the data. These calculated data are used to identify non-compliant participants as discussed in more detail in section 2.6.3.

2.6 Data collection, transformation and reduction

After all stimulus material was produced, a means of delivering the stimuli and administering questionnaires to the participants was needed. Also, data produced during the experiment needed to be saved by participant number. For this study LimeSurvey (Schmitz, 2020) was used. LimeSurvey (Schmitz, 2020) does not require a participant to install software and can handle questionnaires and video integration well, which was of great influence for choosing this platform. Another pro was that the SART script could be integrated without going to an external website and returning; the whole experiment ran in one browser tab. Moreover, all data was directly stored in LimeSurvey (Schmitz, 2020), so no errors could emerge in data entry or from wrong copy-paste operations.

For more details on the questionnaire software, the software selection process and other questionnaire software that was evaluated, see Appendix 8.11.

2.6.1 Eligibility tests

Several tests for compliance to the experiments' requirements were implemented to make sure participants used the correct hardware, operating system and browser. A hearing test to verify a degree of good hearing and the use of working stereo headphones was implemented, as well as a color blindness test. These will be discussed below in order.

To test for hardware, operating system and browser compatibility, user agent string data was collected with a script in LimeSurvey (Schmitz, 2020). Post-experimentally, the stored data was used to identify participants that were not using a PC with Windows in combination with one of the specified browsers. Supported browsers were Chrome, Opera and Edge. The data for participants using other hardware, operating systems or browsers were checked for validity (i.e. question duration times were reasonable and all answers were filled in). Participants who showed unreasonably short question duration times skipped these elements and were excluded from the research. Unsupported hardware contained tablets and mobile phones. Unsupported operating systems included Linux, Apple and Android. The user agent string data was converted with a web converter to obtain readable data. More information on the implementation of acquirement of User Agent string information is found in Appendix 8.16.

A hearing test was implemented to check the working and correct placement of both earpieces of the mandatory headphones. Participants of which one earpiece was not working could not pass the hearing test. Also, when laptop speakers were used, the hearing test would likely fail because it is hard to distinguish left and right when listening to small signal differences in noise using laptop speakers. Participants who heard nothing or only heard sound from one earpiece could not finish the experiment and were sent to an exit page explaining the eligibility criteria. Their data would be incomplete and excluded from further analysis.

A question as part of the demographics section at the end of the research verified whether participants were using in-ear headphones, over ear headphones or speakers. Participants who filled in "speakers" were excluded from further analysis.

To implement a short color blindness test, participants were asked if they were color blind. If the answer was "yes", the participant would be sent to the exit page explaining the eligibility criteria that was also used for the hearing test. The participant could not finish the experiment, data would be incomplete and excluded from further analysis.

2.6.2 Stress level data collection

Stress level measurements were used to determine if stress level increased due to the stressor task and stress decreased due to the stress relieving stimuli. The Δ (i.e. differences) between stress levels in consecutive conditions was calculated and statistically analyzed. Here, there were no exclusion criteria, i.e. if a participant experienced less stress after being 'stressed', or experienced more stress after being 'stress relieved', the participants' data would still be included in the main analysis.

2.6.3 SART data

Data reduction was needed to transform the numeric data coming from the SART task to usable metrics (i.e. number of correct commissions, number of omission errors, number of commission errors, number of correct omissions, mean time for omissions and mean time for commissions). A correct commission is the pressing of the space bar when a number that was not "3" was displayed. A correct omission is the correct non-pressing of the space bar when a number "3" was displayed. The transformed SART data was used to discern the participants that completed the SART task with full attention and followed the instructions properly from the participants that showed give-up responses.

Data for participants who showed a lot of omission or commission errors were fully inspected. Many errors in a short time frame could indicate a give-up response. In literature, a response time of 300 ms between presentation and a button press is considered normal (Bridges et al., 2020, p. 4). Participants who had bursts of response times below 150 ms were not paying attention to the accuracy requirement, but rather pressed the space bar rhythmically. Bursts of response times larger than 1150 ms indicate a period of non-presses, which violates the requirement for sustained attention. Participants showing bursts of anomalous response times were excluded from data analysis. A give-up response is undesirable, since the start of restoring from stress starts during the task instead of after the task. The expected delta in stress level is lower than when the SART task was fully completed.

2.6.4 Fascination & Soft Fascination

The values of the different items of the fascination (PRS) and soft fascination questionnaire were averaged to get a single number value for the fascination part of restorativeness. This approach is also taken in the original PRS article Hartig et al. (1996). Participants who filled in all zeros for the fascination and soft fascination questionnaire were asked for comment on their input. If a participant admitted having filled in non-sensical data, the data was removed from further analysis.

2.6.5 Demographics and Fill in fields

The data for modal preference, computer type, screen size and headphone type were transformed to numerical data. Data that could be entered in fill in fields (i.e. Other) were checked. If non-applicable data was encountered the data was transformed.

At the end of the experiment one question was devoted to distractions during the experiment and reporting issues if there were any. In this fill in field several participants made mentions of eye strain, tearing eyes or blurry vision during the stressor task.

2.6.6 Question and group timings

Data for time spent on a question in LimeSurvey (Schmitz, 2020) were stored during the experiment and later analyzed. The outliers for time spent on a question were used to indicate participants that spent an unusually short or long time on a question. A short time can indicate a participant not reading the informed consent form, browser incompatibility or not filling in questionnaires with attention. Participants who filled in the fascination and soft fascination questionnaire in a really short time in most cases also did not follow instructions with the SART task truthfully and were excluded.

2.7 Statistical analyses

For all statistical testing, SPSS Statistics 25 (IBM, 2017) will be used. Participants with data points outside of 1st quartile – 1.5 * interquartile range to 3rd quartile + 1.5 * interquartile range are considered outliers.

2.7.1 Sample size calculation

The number of comparable studies in which the main effect of Modality (Audio / Visual / AudioVisual) and the interaction effect of Modality and Water level ($Water_{Low}$ / $Water_{Medium}$ / $Water_{High}$) is discussed is limited; little literature is available on this subject. An effect size estimate for the main effect was found in Wooller et al. (2018), which gives a lowest η^2 of 0.54 for an Auditory recovery condition on the Perceived Stress Scale used. The smallest effect size in the Wooller et al. (2018) study was an η_p^2 of 0.07. This translates to f ≈ 0.276, a medium effect size. A study discussing an interaction effect approaching similarity was found in Carles et al. (1999), who found an interaction effect between sounds and images with Sum of squares = 411.71, df = 25, Mean square = 16.47, F-value = 18.30, $p < 0.01$. This translates to a η_p^2 of 0.097 or an f which is between 0.25 and 0.30. Again, a medium sized effect.

The needed sample size for the main study was then calculated using G-power (Faul et al., 2020). The following settings to calculate the sample size were used: The test family was F-tests. The statistical test employed was ANOVA Fixed effects, special, main effects and interactions. An a-priori power test (A priori: Compute required sample size — given α, power and effect size) was chosen as power analysis type. The effect size of 0.276 was filled in in G-power (Faul et al., 2020), along with an α of .05, a power of .9, the numerator df was 4 and the Number of groups was 3. The calculated sample size following this input is 207. This sample size allows to detect an effect of the expected size, going by the evaluated literature. A pilot study was called for, so an extra 30 participants were calculated, bringing the total number of participants to 237. When stressing and stress relieving demonstrates to be functional in the pilot study, the pilot study data can be added to the main research. Then, less participants are needed.

2.7.2 Hypothesis testing

The statistical tests needed to test the hypotheses as stated in section 1.5 are described below. Prior to hypothesis testing, a manipulation check will be implemented to test whether all manipulations induce and relieve stress as intended. A series of dependent sample t-tests are conducted for normally distributed data, where Wilcoxon signed ranks tests are used in case of not normally distributed data. For both the control and experimental conditions, the baseline stress level will be compared with post-stressor stress level. Successively, post-stressor stress level is compared with post-stress-relief stress level. The differences in stress level for the distractor task as described in section 2.4.7 will also be tested.

Visual methods of normality checks were employed to test for normality. Assumption checks are done before running statistical analyses. Shapiro-Wilk tests were used to confirm the normality of variables for all analyses. The Shapiro-Wilk test was chosen because it is most powerful (Razali & Wah, 2011). When the Shapiro-Wilk test is significant, the data shows a departure from normality and non-parametric testing is needed.

1) Hypothesis 1a: "There is an effect of modality on perceived mental restoration and on fascination"

To examine the differences between modal groups on $\Delta_{de\text{-}stress\text{-}experimental}$, fascination and soft fascination, a one-way ANOVA will be employed when data is normally distributed with $\Delta_{de\text{-}stress\text{-}}$

experimental and stimulus modality as factors. Significant effects are reported when $p < 0.05$. Post-hoc t-tests discern which groups significantly differ. A Kruskal-Wallis test is to be used when data is non-normally distributed. A significance level of $p < 0.05$ is used. Post-hoc Mann-Whitney U tests show which groups significantly differ between them, in the case of a significant main effect of modality. In both cases, effect size is reported as η^2.

2) Hypothesis 1b: "Auditory stimuli induce higher mental restoration and fascination than Visual stimuli"

To examine the difference between the auditory and visual groups with respect to $\Delta_{de\text{-}stress\text{-}experimental}$, fascination and soft fascination, t-tests are used if $\Delta_{de\text{-}stress\text{-}experimental}$ data is normally distributed. In case of non-normal data, a Mann-Whitney U test will be employed.
Significant effects are reported when $p < 0.05$. The effect size will be reported as η^2 or *r*.

3) Hypothesis 1c: "Audiovisual stimuli induce higher mental restoration and fascination than both Auditory and Visual stimuli"

In the case of normally distributed $\Delta_{de\text{-}stress\text{-}experimental}$, fascination and soft fascination data, two t-tests with appropriate Bonferroni correction will be utilized. When $\Delta_{de\text{-}stress\text{-}experimental}$, fascination or soft fascination data is not normally distributed, non-parametric Mann-Whitney U tests will be used to calculate and compare the mean differences. If the audiovisual $\Delta_{de\text{-}stress\text{-}experimental}$, fascination or soft fascination mean is significantly larger than the means of both the auditory and visual condition, an additive effect is present. Effect sizes will be reported as η^2 or *r*.

4) Hypothesis 2: "A higher (accessible) water content in the scene leads to higher mental restoration and fascination"

To examine the difference between water levels in the scenes of the experimental condition, a one-way ANOVA will be employed when data is normally distributed with $\Delta_{de\text{-}stress\text{-}experimental}$, fascination, soft fascination and water level as factors. Significance is reported when $p < 0.05$, where effect size is reported as η_p^2. If a significant effect is found, a post hoc t-test determines which water levels differed from each other. In case of non-parametric testing, a Kruskal-Wallis test is performed. The effect size is then reported as η^2.

5) Interaction between modality & water level

To examine a possible interaction effect between stimulus modality and water level, a two-way ANOVA is performed with $\Delta_{de\text{-}stress\text{-}experimental}$, fascination, soft fascination, stimulus modality and water level as factors. Assumption checks for a two-way ANOVA are done, significance is reported when $p < 0.05$. Effect size is reported as η^2.
In case the data are not normally distributed, the same two-way ANOVA will be performed. Non-parametric interaction effect tests are not available, so a parametric test will be reported with the remark that data are not normally distributed and caution is to be observed when interpreting the outcome. Effect size η^2 will be reported only if a significant interaction effect is present.

Regarding effect sizes, η^2 is mostly used and calculated using Equation 1 below. Where: H is the Kruskal-Wallis test value; k is the number of groups; n is the total number of observations. Effect size η^2 is interpreted as: $\eta^2 = 0.01$ is a small effect; $\eta^2 = 0.06$ is a medium effect; $\eta^2 = 0.14$ is a large effect (Cohen, 1988).

$$\eta^2 = \frac{\mathrm{H} - \mathrm{k} + 1}{\mathrm{n} - \mathrm{k}}$$

Equation 1: Effect size η^2 calculation after Kruskal-Wallis H test

When a Mann-Whitney U or Wilcoxon signed ranks test is performed, effect size *r* is calculated with Equation 2. Where: Z is the output score from the Mann-Whitney U test; N is the sample size. Effect size *r* is interpreted as: *r* = 0.2 is a small effect; *r* = 0.5 is a medium effect; *r* = 0.8 is a large effect (LeCroy & Krysik, 2007).

$$r = \frac{\mathrm{Z}^2}{(\mathrm{N} - 1)}$$

Equation 2: Effect size *r* calculation after Mann-Whitney U or Wilcoxon signed ranks test

2.7.3 Exploratory analyses

To supplement the statistical analyses, Cronbach's alpha is calculated for both the fascination and soft fascination questionnaire. The fascination questionnaire was modified as explained in section 2.5.2. Therefore, internal consistency is checked. The soft fascination questionnaire is newly developed, so an analysis for the number of factors it measures was conducted. Furthermore, correlations between fascination and soft fascination are investigated, as well as a possible learning effect.

Factor analysis

Firstly, Cronbach's alpha is calculated for the fascination and soft fascination questionnaire. Extra analyses will be conducted if internal consistency is low, and question items that need to be removed to get an acceptable alpha are reported on. A factor analysis for the soft fascination questionnaire is conducted to test the number of factors it measures. When a large number of factors is measured, the items lack coherence. A one-factor solution that measures soft fascination is preferred. The percentage of variance explained by the factor(s) will be reported on, also in case questionnaire items are removed.

Learning effect

Counterbalancing for the control and experimental condition was not implemented (i.e. the control condition was always presented first, and the experimental condition was always last), so a possible learning effect is investigated. The means of $\Delta_{\text{stress-control}}$ and $\Delta_{\text{stress-experimental}}$ variables are compared. When parametric testing is indicated, a paired sample t-test will be used, whereas in non-parametric testing a Wilcoxon signed ranks matched pairs test is employed.

Mediation analysis

Using mediation analysis techniques it can be investigated if fascination is the main driver for mental restoration. If significant differences between the control and experimental condition are found a mediation analysis is pursued for fascination and soft fascination. The low water condition most probably will not elicit significant differences, so only the conditions that showed significant results will be analyzed. If data is normally distributed and other assumptions are not violated a Sobel test will be implemented after the a, b and c coefficients are determined. When data is not normally distributed the Sobel test will still be reported on but results need to be interpreted with caution.

Correlation

Correlation analyses show a relationship between the de-stressing effect and attention capturing through fascination and soft fascination. Spearman rank correlations between $\Delta_{de\text{-}stress\text{-}experimental}$ - fascination and $\Delta_{de\text{-}stress\text{-}experimental}$ - softfascination$_{Mean}$ are calculated and reported on.
Also, the correlation between the fascination and soft fascination questionnaire is examined; Spearman rank correlations between fascination and softfascination$_{Mean}$ are calculated and reported on.

3 Results

3.1 Pilot test

A pilot test with n=30 was conducted to verify the effectiveness of the stressor and de-stressor. A Wilcoxon signed ranks test was employed to measure the difference between pre-stressor and post-stressor stress level for all participants, as well as per experimental group. The results are shown in Table 5 below. The accompanying means and SDs per condition are displayed in Table 9.

	Control condition			**Experimental condition**	
	Pre-/post-stressor 1	**Post-stressor/post de-stressor 1**	**Post de-stressor 1 / pre stressor 2**	**Pre-/post-stressor 2**	**Post-stressor/post de-stressor 2**
Overall	< .001	< .001	< .001	< .001	< .001
Z-score	-9.882	-9.837	-6.606	-7.069	-10.387
Auditory	< .001	< .001	.003	< .001	< .001
Z-score	-6.404	-6.600	-2.972	-3.762	-5.757
Visual	< .001	< .001	.001	.001	< .001
Z-score	-4.990	-4.015	-3.182	-3.318	-5.724
Audiovisual	< .001	< .001	.001	.001	< .001
Z-score	-5.726	-6.057	-5.304	-5.118	-6.388
Water$_{Low}$	< .001	< .001	< .001	.005	< .001
Z-score	-5.216	-5.736	-3.592	-2.781	-5.865
Water$_{Medium}$	< .001	< .001	< .001	< .001	< .001
Z-score	-6.408	-5.562	-4.919	-5.864	-6.652
Water$_{High}$	< .001	< .001	.004	.002	< .001
Z-score	-5.474	-5.684	-2.919	-3.123	-5.389

Table 5: Pilot test effectiveness for stress and de-stress cycles

Participants reported significantly more stress after the stressor task was finished ($p < .001$ overall and $p < .005$ when viewed per condition). Also, overall the stress level dropped significantly after the progress bar de-stressor and the natural de-stressing intervention ($p < .001$). Moreover, the stress values were significantly lower in the experimental condition than in the control condition. No further inferences were made on differences between experimental conditions at this point because the sample was too small to draw meaningful conclusions. The data gathered during the pilot test was later added to the rest of the experimental data. In conclusion, the pilot test showed that both the stressor and de-stressor were effective in the control and experimental conditions.

3.2 Data integrity check

In this research no imputation of missing data was needed, nor exclusion of cases due to missing or erroneous data. Participants with data points outside of 1st quartile – 1.5 * interquartile range to 3rd quartile + 1.5 * interquartile range were considered outliers. No outliers were present in the measured stress level data as displayed in Figure 7. In $\Delta_{de\text{-}stress\text{-}experimental}$, eight outliers were present. The fascination data contained six outliers, where soft fascination only had one outlier. Outliers were excluded from analyses when they influenced the outcome of a statistical test.

Before doing an overall manipulation check on the stress level data, an overall normality check was done on stress level data. A Shapiro-Wilk test, accompanying histograms and QQ plots, showed a significant departure from normality, $W(227) = .861$, $p < .001$ for the overall normality test. The Wilk statistic lies between .861 and .975 for all Stress level, Δ_{stress} (difference between baseline stress level and post-stressor stress level) and $\Delta_{de\text{-}stress}$ (difference between post-stressor stress level and post de-stressor stress level) variables. After removal of data from participants that replied with zero scores for all Stress level measurements, a significant deviation from normality was still present, $W(218) = .870$, $p < .001$. Stress level data is not normally distributed. Non-parametric tests need to be used when testing the Stress level, Δ_{stress} and $\Delta_{de\text{-}stress}$ variables.

Normality analysis for mean fascination showed significant deviations from normality with $W(227) = .947$, $p < .001$. Non-parametric tests are to be used for analyzing the fascination questionnaire.

A normality analysis for mean soft fascination using all items showed a normal distribution with $W(227) = .994$, $p < .431$, warranting the use of an ANOVA analysis. After reliability analysis, as further described in section 3.9, the internal consistency of the full soft fascination questionnaire appeared to be low. Item one and six of the soft fascination questionnaire negatively influenced internal consistency. Normality was then also analyzed for the mean of the reduced soft fascination questionnaire (using only items two to five), which yielded a non-normal distribution $W(227) = .985$, $p < .015$. For completeness, the effect of stimulus modality on the overall mean of the full soft fascination questionnaire, and on the reduced soft fascination questionnaire will be examined separately. A Kruskal-Wallis H test will be performed over the mean of the reduced soft fascination questionnaire, analyzing statistically significant differences between the independent variable of modality and the dependent variable of soft fascination.

Further investigations included the field "Did you perceive any issues during the experiment? Did you encounter distractions or problems that could interfere with the results or do you have any comments?". This question was presented at the end of the demographics questionnaire and could give insight in possible biases induced during the time of the experiment.

A fear of jumpscares during the stressor task or the nature intervention was mentioned by several participants: "I was really scared of jumpscares in the video, but that's probably just me". One participant made a mention of added stress due to the nature intervention: "I don't like nature sounds. The wind and sea noises stress me out". This was also clearly visible in the stress level measurements for that participant. The nature intervention did not have the intended effect. Data from this participant were not used in most analyses. Another participant indicated that "Some of the issues are personal with some of the videos and pauses inciting stress onto my person with other media int he past influencing that stress". The data from this participant were excluded, also because of issues with the stressor data. Other issues mentioned included nervous ticks, personal issues or environmental factors. Data for these participants was still usable. Some examples are displayed below:

"No issues at all. I believe I did better after the breathing exercise, and I realized how anxious I was when the loading bar showed up. It triggerd this nervous tick that I have, which is similar to biting one's nails, but I tear down irregularities on my lips."

"Yes, but it was due to my loudy neighbours"

Finally, a participant filled in "1920 x 1080" for screen size. Since this is non-critical data, the rest of the data was still used.

3.3 Factor analysis for Soft Fascination Questionnaire

The soft fascination questionnaire was designed for this experiment and is described in more detail in the Method section, sections 2.5.3 and 2.6.4. Before running the analysis, item 5 ("I felt irritated") and item 6 ("I was in awe") were re-coded, these questions were negatively coded. An 'awe' response is undesirable when soft fascination is to be induced, as described in section 4.2 (Why does hard fascination preclude reflection?) of (Joye & Dewitte, 2018). No score card for this scale is available yet. Therefore, a factor analysis was performed on the Soft Fascination questionnaire, revealing two factors that explain 58.74% variance. A one factor solution with a higher percentage of explained variance would be more desirable. A varimax rotation was attempted but did not lead to a higher percentage of explained variance.

Questionnaire item	Component 1	Component 2	Question (0 = Not at all, 6 = Completely)
1	-0.261	0.174	I could still think about other things
2	0.775	0.317	The scene was pleasant
3	0.448	-0.663	I experienced an ebb and flow of attention
4	0.828	-0.104	There are many things here that attract my attention effortlessly
5	0.477	0.755	† I felt irritated
6	-0.729	0.244	† I was in awe

Table 6: Factor analysis on Soft Fascination questionnaire data — all items included — † indicates re-coded items.

Items two to five load positively on component 1, where only questionnaire item 1, 2, 5 and 6 positively load on component 2. From these factor loadings, questionnaire items 1 and 6 show low loadings. Removing them and re-running the factor analysis revealed that 76.57% variance is explained by 2 factors.

Questionnaire item	Component 1	Component 2	Question (0 = Not at all, 6 = Completely)
2	0.845	-0.150	The scene was pleasant
3	0.407	0.816	I experienced an ebb and flow of attention
4	0.803	0.238	There are many things here that attract my attention effortlessly
5	0.617	-0.642	† I felt irritated

Table 7: Factor analysis on Soft Fascination questionnaire data — item 2-5 included — † indicates re-coded item.

3.4 Manipulation check

To verify that participants overall were more stressed after the two stressor tasks (Δ_{stress}: pre- and post-stressor measurement) and sequentially relieved of stress ($\Delta_{de\text{-}stress}$: post stressor and post de-stressor) in both the control and experimental conditions, Wilcoxon signed ranks tests were performed on the non-normal stress level data (every interval was measured, as well as all conditions). The Wilcoxon signed ranks test results are in Table 8 and the means for stress levels are displayed in Figure 7 (means are displayed numerically under the measurement). The results of the accompanying statistical tests are in Table 8 below.

Difference over all participants n=227	Pre-/post-stressor 1	Post-stressor/post de-stressor 1	Post de-stressor 1 / pre stressor 2	Pre-/post-stressor 2	Post-stressor/post de-stressor 2
Z	-9.882	-9.837	-6.606	-7.069	-10.387
p	< .001	< .001	< .001	< .001	< .001

Table 8: Differences over stress intervals — Over all participants

A Wilcoxon signed-ranks test showed that the stressing and de-stressing manipulations did elicit a statistically significant change in stress level. Most important for this research is the de-stress analysis in the experimental condition ($Z = -10.387$, $p < .001$) with a small effect size of $r = .082$. Median stress level rating was 38 when stressed and 30 after de-stressing. In Figure 7, an asterisk (or "*") indicates a statistically significant difference between the conditions.

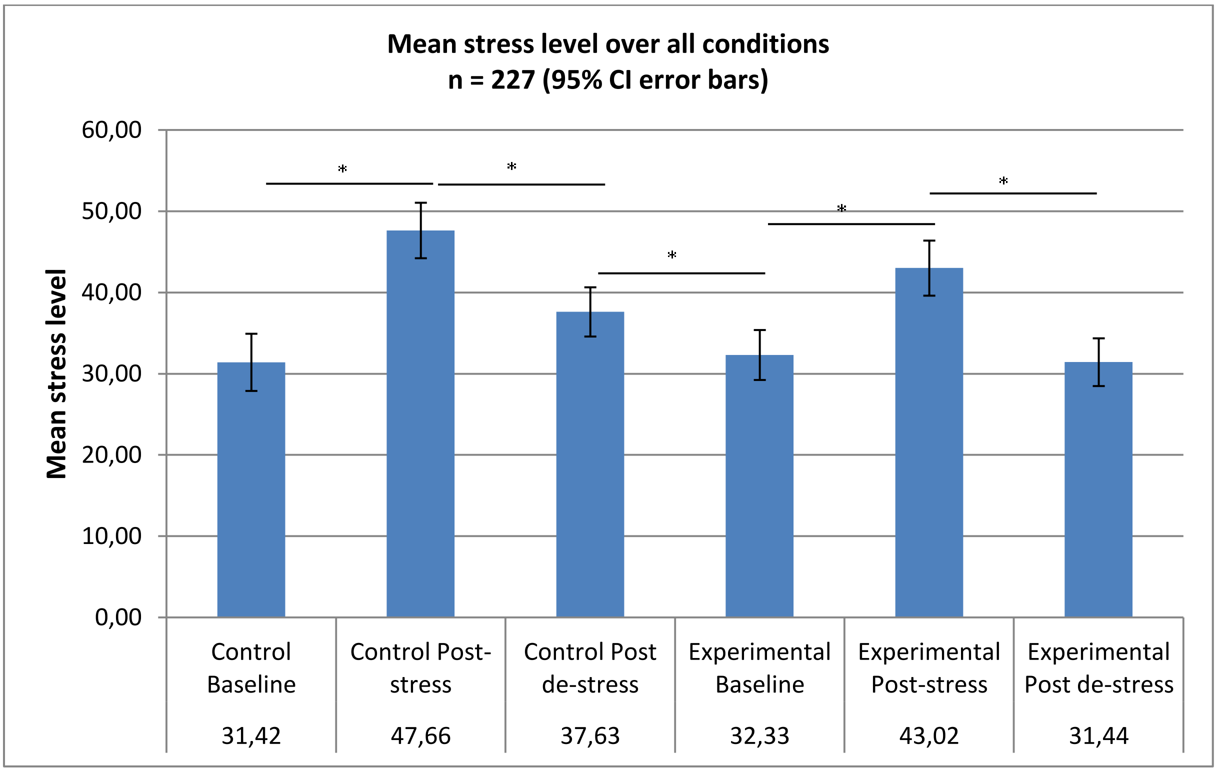

Figure 7: Mean stress levels on measurement points in the experiment (n=227)

Figure 7 shows a clear rise in mean stress levels after the stress tasks (Post-stress), as well as a decline in stress level after the de-stressor tasks (Post de-stress). Also visible is a decline of stress level in the distractor task (between Control Post de-stress and Experimental Baseline) in which participants were asked to close their eyes and focus on their breathing. A larger stressing effect during the control condition can be observed, an analysis will follow later.

The results in Table 8 show that stressing participants overall has the intended effect, and that the induced stress can also be mitigated (with the stimuli) during the experiment. The manipulation was thus proven effective for the stressor and de-stressor in both the control and experimental conditions.

3.5 Effectiveness of de-stressor

The de-stressor appears to be effective in the control and in the experimental condition. However, it is not yet clear whether the effectiveness of the de-stressor differs statistically between the control and experimental condition. A differentiation can be made between the idea that de-stressing is just a matter of time, or that a natural stimulus is more effective in de-stressing the participants as hypothesized. The de-stressor data is the difference Δ (delta) between Post-stress and Post de-stress variables and will be analyzed using an outlier analysis, overview of descriptive statistics and lastly with statistical testing.

First, the data was inspected more closely for erroneous data, after which an outlier analysis was conducted. No outliers were present in the stress level data. Furthermore, eight participants reported zero on all their stress level responses. This can imply something is wrong with the question asked or with the interpretation of the participant. In the discussion section, more attention will be paid to this observation.

In Table 9, an overview of the mean and SD data over conditions for the stress level data and stress level delta data is displayed. A small mean is observed for the de-stressing effect in the experimental condition for Video, $Water_{Low}$ and $Water_{High}$. $Water_{Medium}$ achieves the highest decrease in stress, where the means of Audio and Audiovisual are higher than the overall mean. A higher $\Delta_{de\text{-}stress}$ value indicates more stress relief.

		Control			Experimental			Control		Experimental	
		Baseline	Stressed	De-stressed	Baseline	Stressed	De-stressed	Δ_{stress}	$\Delta_{de\text{-}stress}$	Δ_{stress}	$\Delta_{de\text{-}stress}$
Overall	Mean	32.32	49.19	38.75	33.23	44.42	32.36	16.87	-10.44	11.19	**-12.06**
n=218	SD	26.51	24.77	22.15	22.80	24.96	21.84	22.40	16.20	21.32	16.18
A	Mean	29.22	48.14	36.31	32.82	43.09	29.57	18.92	-11.83	10.27	**-13.52**
n=77	SD	25.88	25.89	22.95	24.58	26.44	22.69	23.30	15.15	21.53	19.08
V	Mean	30.66	44.43	37.03	32.59	39.94	31.01	13.76	-7.40	7.35	-8.93
n=68	SD	26.49	23.58	21.32	23.02	22.60	20.30	21.17	16.82	17.52	12.89
AV	Mean	37.14	54.74	42.92	34.26	49.99	36.55	17.60	-11.82	15.73	**-13.44**
n=73	SD	26.86	23.88	21.76	20.83	24.74	21.98	22.53	16.51	23.65	15.37
$Water_{Low}$	Mean	31.33	46.36	36.24	31.29	37.82	28.18	15.03	-10.13	6.53	-9.64
n=72	SD	26.51	25.43	22.17	22.22	25.58	22.34	21.76	12.48	17.21	12.77
$Water_{Medium}$	Mean	25.07	46.10	35.42	27.42	46.13	29.86	21.03	-10.68	18.70	**-16.27**
n=71	SD	24.01	25.06	21.78	20.19	25.40	20.07	23.97	17.97	22.98	18.61
$Water_{High}$	Mean	40.13	54.84	44.31	40.59	49.13	38.73	14.71	-10.53	8.55	-10.40
n=75	SD	26.98	23.14	21.69	23.96	22.82	21.79	21.18	17.74	21.61	16.06

Table 9: Mean and SD stress level per modal condition and per scene (water) condition — bold green values performed better than Overall $\Delta_{de\text{-}stress}$

For the fascination, full soft fascination and reduced soft fascination questionnaires the means and SDs are displayed in Table 10 below. For completeness, Cronbach´s α is also reported at the bottom of the table.

De-stress experimental	**Fascination**			**Full soft fascination questionnaire**			**Reduced soft fascination questionnaire**		
	n	Mean	SD	n	Mean	SD	n	Mean	SD
Audio	74	3.82	1.30	76	2.97	0.77	76	2.99	0.86
Video	73	4.03	1.16	73	3.11	0.70	73	3.05	0.80
Audiovisual	72	4.01	1.20	77	3.00	0.75	77	2.97	0.83
$Water_{Low}$	73	4.15	1.18	74	3.09	0.73	74	3.08	0.87
$Water_{Medium}$	71	3.90	1.19	75	2.96	0.75	75	2.93	0.84
$Water_{High}$	75	3.81	1.27	77	3.03	0.74	77	3.02	0.78
Total	219	3.95	1.22	226	3.03	0.74	226	3.01	0.83
Cronbach's α	.90			.11			.67		

Table 10: Mean, SD and Cronbach's α data for the fascination and soft fascination questionnaire

In the next section, the hypotheses of this study will be tested.

3.6 Hypothesized effects — hypothesis testing

Now the stressor and de-stressor are proven effective, an analysis to test the restorative effect over all experimental conditions can be conducted. A normality test for the mental restoration dimensions in the experimental condition was conducted first. A Shapiro-Wilk test showed a significant departure from normality on $\Delta_{\text{de-stress-experimental}}$ with $W(218) = .870$, $p < .001$. A normality test per modality was also performed and revealed that none of the data are normally distributed (all $p < .05$) so non-parametric tests are to be used in hypothesis testing for $\Delta_{\text{de-stress-experimental}}$ data.

Regarding the fascination and soft fascination questionnaires, the mean of all fascination (PRS) items and soft fascination items was calculated and used as described in section 3.10. This practice proved adequate in previous research by Berto (2005), Sona et al. (2019) and Jahncke et al. (2015). The mean fascination data was also tested for normality and did not meet the criteria for a normal distribution, $W(227) = .947$, $p < .001$. Soft fascination was normally distributed only when the full questionnaire was used, $W(227) = .994$, $p = .431$. When using the reduced soft fascination questionnaire (only items two to five) to increase internal consistency (see section 3.9 for more details) a non-normal distribution for soft fascination was observed, $W(227) = .985$, $p = .015$. After removal of outliers and zero scores for fascination and soft fascination the outcomes of normality testing were unchanged.

3.6.1 Hypothesis 1a: "There is an effect of modality on perceived mental restoration and on fascination"

The first hypothesis regards the effects of modality on mental restoration and is divided into three sub-hypotheses which are presented below. The statistical tests encompass testing for $\Delta_{\text{de-stress-experimental}}$, fascination and soft fascination. All numerical mean and SD data for $\Delta_{\text{de-stress-experimental}}$, fascination and soft fascination are displayed in Table 9 and Table 10 respectively.

In Figure 8, the $\Delta_{\text{de-stress-experimental}}$ information over modal conditions is presented visually. Please note that the stress levels decrease when de-stressing, therefore the displayed means are negative.

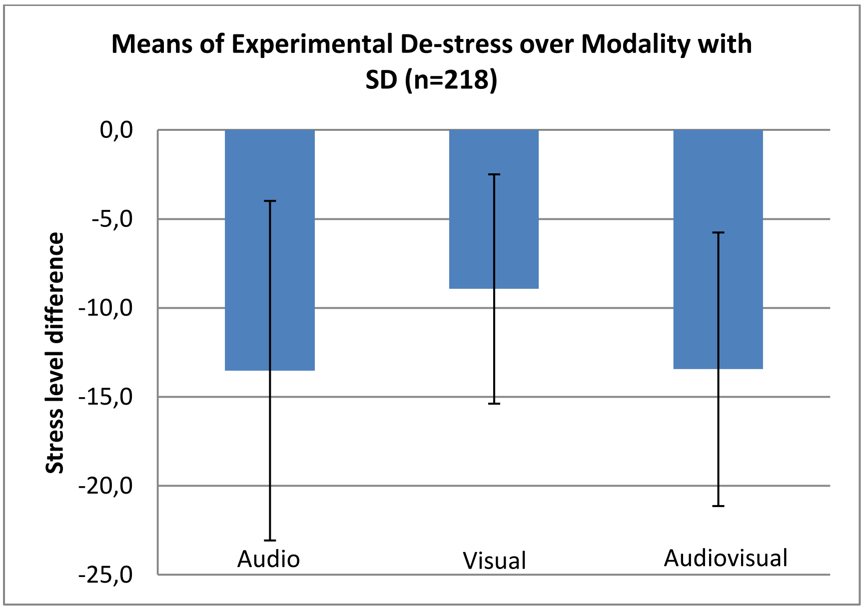

Figure 8: Stress restoration effect of modality with SD error bars

The $\Delta_{de\text{-}stress\text{-}experimental}$ data were found to be non-normal as discussed in section 3.2, so a Kruskal-Wallis H test was performed to examine the effect of modality on $\Delta_{de\text{-}stress\text{-}experimental}$. An important assumption with the Kruskal-Wallis H test is that the distribution of the variables that are compared have the same shape; a histogram shows that the shape of the distribution is similar. Outliers were present in the $\Delta_{de\text{-}stress\text{-}experimental}$ data due to zero scores. When $\Delta_{stress\text{-}experimental}$ and $\Delta_{de\text{-}stress\text{-}experimental}$ were zero, no stressing or stress relieving effect was present and the manipulations did not have the intended effect. Therefore, eight participants of which the stress level scores were zero on all conditions were excluded from this part of the analysis.

	Experimental $\Delta_{de\text{-}stress}$
Kruskal-Wallis H	3.849
df	2
Asymp. Sig.	.146

Table 11: Kruskal-Wallis H test for modal comparison

Hypothesis 1a-1: Effect of modality on stress restoration

The Kruskal-Wallis H test, shown in Table 11, showed that no statistically significant difference in $\Delta_{de\text{-}stress\text{-}experimental}$ between the different modalities was present, $\chi^2(2) = 3.849$, $p = .146$, with a mean rank stress level of 117.96 for Visual, 112.86 for Auditory and 98.07 for Audiovisual. The effect size η^2 for $\Delta_{de\text{-}stress\text{-}experimental}$ over all modalities was calculated to be 0.008. This is a very small effect.

Hypothesis 1a-2: Effect of modality on fascination

Fascination data was found to be non-normally distributed as discussed in section 3.2. A Kruskal-Wallis H test was performed to investigate differences in fascination scores between experimental groups. The Kruskal-Wallis test showed that no statistically significant difference in perceived fascination between the different modalities was present, $\chi^2(2) = 0.766$, $p = .682$. The effect size η^2 for fascination over all modalities was calculated to be 0.008, a very small effect size.

Hypothesis 1a-3: Effect of modality on soft fascination

Soft fascination data for the mean of all items was normally distributed as discussed in section 3.2. An ANOVA analysis was conducted to examine the effect of modality on the mean full (items one to six — see Table 6) soft fascination construct. Simple main effects analysis showed no statistically significant effect of modality on soft fascination F (2, 217) = .717, *p* = .490. This indicates that no differences were present between soft fascination outcomes for the different groups. Effect size η_p^2 for soft fascination was calculated to be 0.006, a very small effect.

The soft fascination mean data for item two to five (reduced soft fascination questionnaire) showed a significant deviation from normality as discussed in section 3.2. A Kruskal-Wallis H test was used to investigate the effect of modal conditions on the mean of the reduced (items two to five — see Table 7) soft fascination questionnaire. The Kruskal-Wallis test showed that no statistically significant difference in soft fascination was present between modal conditions, with $\chi^2(2) = 0.322$, *p* = .851. Effect size was again very small: η^2 is 0.008.

From these test outcomes it can be concluded that there is no statistically significant difference between the modality groups. The modality in which the experimental stimuli (Audio, Visual or Audiovisual) are presented did not make a difference in restorative effect with respect to $\Delta_{de\text{-}stress\text{-}experimental}$, fascination or soft fascination. Therefore, hypothesis 1a is rejected.

3.6.2 Hypothesis 1b: "Auditory stimuli induce higher mental restoration and fascination than Visual stimuli"

Since no statistically significant difference for $\Delta_{de\text{-}stress\text{-}experimental}$, fascination or soft fascination was present between modalities, the predicted effect of auditory stimuli having a higher restorative potential than visual stimuli was not observed and hypothesis 1b was rejected.

To calculate the effect size of stress relief, Z scores were obtained by performing Mann-Whitney U tests on $\Delta_{de\text{-}stress\text{-}experimental}$, fascination and the two variants of the soft fascination questionnaire. The Mann-Whitney U test on $\Delta_{de\text{-}stress\text{-}experimental}$ confirmed that no statistically significant difference in stress reduction between the Audio and Video modality was present, U = 2642, *p* = .526. The effect size *r* when comparing $\Delta_{de\text{-}stress\text{-}experimental}$ for the visual and auditory modality was calculated to be 0.0019. The Z scores and effect size data for all restoration variables is displayed in Table 12. As expected from the results in hypothesis 1a, no statistically significant differences were found and all effect sizes are very small to small.

Contrast	Z	*p*	Effect size *r*
Audio – Video (n = 145)			
$\Delta_{de\text{-}stress\text{-}experimental}$	-.635	.526	0.0019
Fascination	-1.190	.234	0.0098
Soft fascination (Items 1 to 6)	-1.518	.129	0.016
Soft fascination (Items 2 to 5)	-.553	.580	0.0021

Table 12: Audio-Video contrasts and effect sizes

The last hypothesis test regarding modality is covered in hypothesis 1c and is discussed below.

3.6.3. Hypothesis 1c: "Audiovisual stimuli induce higher mental restoration and fascination than both Auditory and Visual stimuli"

As can be observed in Figure 8 the Video condition shows the lowest restoration, where the Audio and Audiovisual means are very similar, with overlapping error bars. It was hypothesized that "Audiovisual stimuli induce higher mental restoration and fascination than both Auditory and Visual stimuli", however hypothesis 1a showed that a significant difference between modalities was not present. The additive effect for Audiovisual stimuli is therefore not present and hypothesis 1c must also be rejected.

To calculate the effect size of modality on $\Delta_{\text{de-stress-experimental}}$, fascination and soft fascination, the Mann-Whitney Z statistics are needed. Therefore, Mann-Whitney U tests were performed to compare the restorativeness of the stimuli. In Table 13 the accompanying statistics are displayed, confirming what was found in Hypothesis 1a. No statistically significant effects of modality on mental restoration, fascination or soft fascination were found. For the audio-audiovisual contrast a value of p = .035 is reached for $\Delta_{\text{de-stress-experimental}}$. Note that a Bonferroni correction is needed, and the alpha-level needed to reach significance is p = .017.

Contrast	**Z**	***p***	**Effect size *r***
Audio — Audiovisual (n = 150)			
$\Delta_{\text{de-stress-experimental}}$	-.856	.392	0.0049
Fascination	-.376	.707	0.00095
Soft fascination (full)	-.645	.519	0.0028
Soft fascination (reduced)	-.069	.945	0.00032
Video — Audiovisual (n = 141)			
$\Delta_{\text{de-stress-experimental}}$	-2.106	.035	0.032
Fascination	-.668	.504	0.0032
Soft fascination (full)	-.864	.388	0.0053
Soft fascination (reduced)	-.553	.580	0.0022

Table 13: Audio-Audiovisual and Video-Audiovisual contrasts and effect sizes

Summary hypothesis 1

To summarize, the mentally restorative effects found in literature regarding the differing modality of stimuli were not found in the current study. The hypotheses regarding sensory modality are all rejected. In Figure 10 an overview of the findings of statistical tests is presented. The Audio, Visual and Audiovisual effects are considered very small (Ferguson, 2009) and are lower than expected. Additionally, all experimental conditions were perceived as approximately equally fascinating, given the similarity between means scores for all conditions. The second hypothesis regarding water content in the natural stimuli is discussed below.

3.6.4. Hypothesis 2: "A higher (accessible) water content in the scene leads to higher mental restoration and fascination"

The water content in the experimental scenes was manipulated on three levels with $\text{Water}_{\text{Low}}$, $\text{Water}_{\text{Medium}}$ and $\text{Water}_{\text{High}}$ conditions. The factorial design allowed for manipulation of water level over modal conditions of Audio, Video and Audiovisual. It was hypothesized that "A higher (accessible) water content in the scene leads to higher mental restoration". Data for nine participants were excluded from the analysis because their stress level responses were all zero, or because nature was reported to be distressing for the participant. The same shape distribution assumption was tested; histograms show that the shape of the distributions is similar. In Table 9 the mean and SD data on $\Delta_{de\text{-}stress\text{-}experimental}$ is displayed. In Table 10 the means and SDs are displayed for fascination and the two variants of soft fascination data. Normality of the dependent variables is discussed in section 3.2, and statistical tests will be used accordingly.

Hypothesis 2-1: Effect of water level on stress restoration

A Kruskal-Wallis H test (n = 218) showed that an effect nearing statistical significance was present for $\Delta_{de\text{-}stress\text{-}experimental}$ between the different water levels, $\chi^2(2) = 5.328$, $p = .07$, with a mean rank $\Delta_{de\text{-}stress\text{-}experimental}$ level of 117.26 for $Water_{Low}$, 95.37 for $Water_{Medium}$ and 115.43 for $Water_{High}$. An overview of the means of $\Delta_{de\text{-}stress}$ in the experimental group is shown in Figure 9. The stress levels decrease when de-stressing; please note that the displayed means are negative for this reason. The effect size η^2 for $\Delta_{de\text{-}stress\text{-}experimental}$ is 0.015.

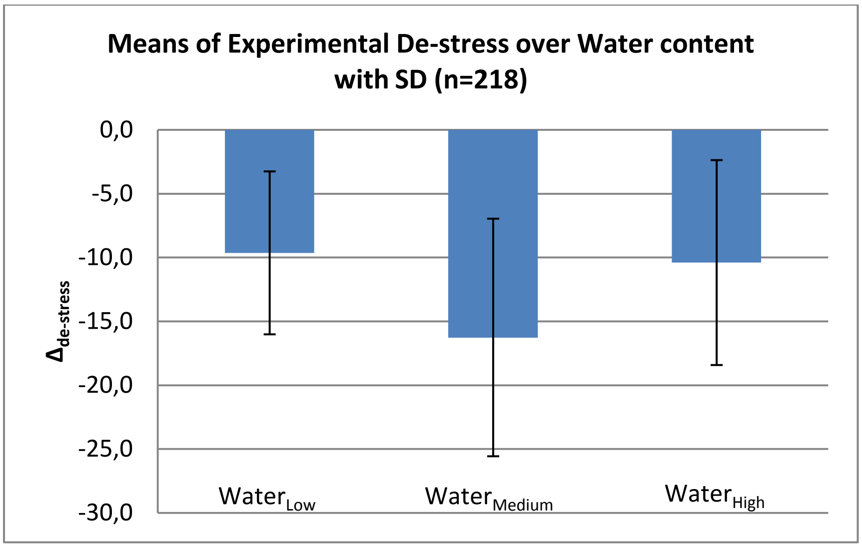

Figure 9: De-stressing effect of water content with SD error bars

Hypothesis 2-2: Effect of water level on fascination

A Kruskal-Wallis H test (n = 226) was performed to test an effect of scene water content on mean fascination data. Mean data and SDs are displayed in Table 10 above. The Kruskal-Wallis H test showed that no statistically significant difference in perceived fascination between the different water levels was present, $\chi^2(2) = 2.696$, $p = .260$ with a mean rank fascination of 119.77 for $Water_{Low}$, 106.65 for $Water_{Medium}$ and 103.65 for $Water_{High}$. This finding indicates that no difference was present between water levels with respect to perceived fascination. The effect size η^2 for fascination over different water conditions is 0.0031.

Hypothesis 2-3: Effect of water level on soft fascination

An ANOVA (n = 226) with soft fascination as dependent variable and the factor of water content was performed on the mean of the complete soft fascination questionnaire. All assumptions with ANOVA testing were met after exclusion of one participant because it was an outlier. Analyses with and without outlier did not show different statistical outcomes. The ANOVA analysis that examined the effect of water level on the mean soft fascination construct was performed on the data without the outlier. Simple main effect analysis showed no statistically significant effect of water content on soft fascination F (2, 217) = .631, p = .533. The effect size η_p^2 was calculated to be 0.006, a very small effect.

A Kruskal-Wallis H test (n = 226) with water as independent variable and soft fascination as dependent variable was performed for the mean of the reduced soft fascination questionnaire. The outcomes showed no statistically significant differences in perceived restoration between water levels, $\chi^2(2) = 1.083$, p = .582 with a mean rank soft fascination of 119.39 for $\text{Water}_{\text{Low}}$, 108.34 for $\text{Water}_{\text{Medium}}$ and 112.87 for $\text{Water}_{\text{High}}$ content in the displayed scene. Effect size η^2 for soft fascination over the three water levels is 0.0041.

The effect sizes are all very small and are summarized in Table 14 below.

Mental restoration	**n**	***p***	**Effect size η^2**
$\Delta_{\text{de-stress-experimental}}$	218	.07	0.015
Fascination	226	.260	0.0031
Soft fascination (full)	226	.533	0.006 — η_p^2
Soft fascination (reduced)	226	.582	0.0041

Table 14: Water related *p*-values and effect sizes; η^2 and η_p^2

Summary hypothesis 2

The hypothesis that an increasing water content inverse-proportionally decreases stress needs to be rejected. In Figure 10 below an overview of the findings of statistical tests is presented. Please note that in Figure 10 the upper two analyses regard the hypothesis testing, where the lower three analyses regard the exploratory analyses discussed in section 3.7. Although no statistically significant effects were found, means for $\Delta_{\text{de-stress-experimental}}$ follow the hypothesized direction between the $\text{Water}_{\text{Low}}$ and $\text{Water}_{\text{Medium}}$ conditions, but this is not the case between $\text{Water}_{\text{Medium}}$ and $\text{Water}_{\text{High}}$, as clearly visible in Figure 9. For the fascination and soft fascination measurements this trend is not visible, since $\text{Water}_{\text{Medium}}$ scored lowest. Effect sizes η^2 (for $\Delta_{\text{de-stress-experimental}}$, fascination and the reduced soft fascination) and η_p^2 (for soft fascination containing all items) are considered very small, see Table 14. Differences between water content in the stimulus materials were therefore not observed in the data and effect sizes are very small.

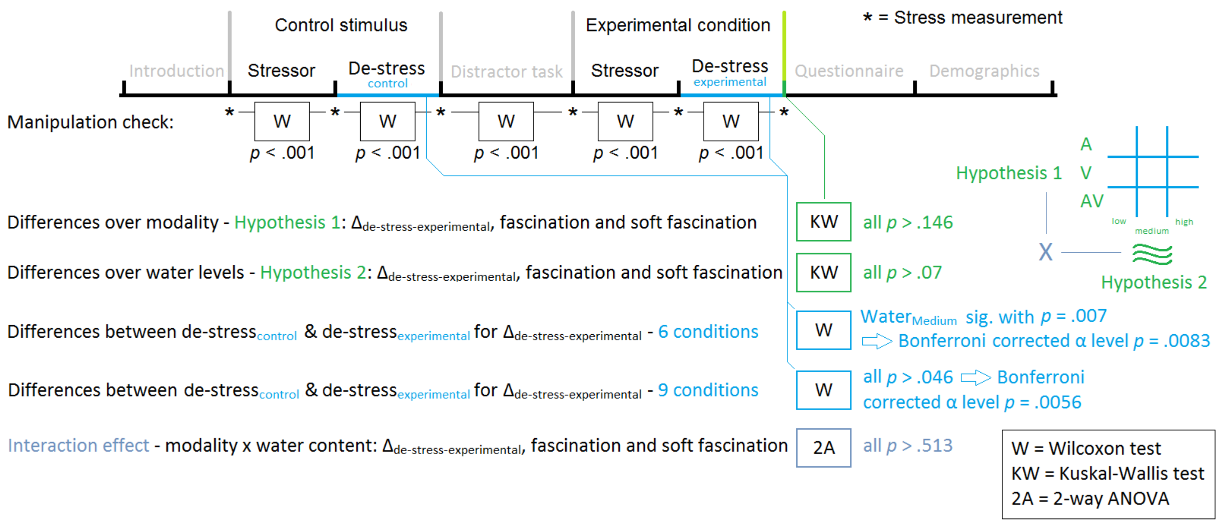

Figure 10: Overview of results for hypothesis 1, hypothesis 2 and explorative analyses

To check if not only main effects were present and since water showed a non-linear relationship with restoration, an interaction effect might be present between modality and water content.

3.7 Exploratory analyses on the interaction between modality and water level

Although there were no significant main effects, inspection of the means might indicate that an interaction effect between modality and water content was present. A higher mean restoration was only observed for the Audio, Audiovisual (Figure 8) and $Water_{Medium}$ conditions (Figure 9). More specifically, for $\Delta_{de\text{-}stress\text{-}experimental}$ the auditory modality and $Water_{Medium}$ scene show the largest restoration. For fascination the most restorative scene was $Water_{Low}$, with the visual modality as most restorative condition (see Table 15). For the reduced soft fascination questionnaire the video modality was perceived as most restorative, with the $Water_{Low}$ scene as the most restorative condition. A possible explanation for these findings could be an interaction effect, which will be pursued exploratively below. First, exhaustive interaction tests on $\Delta_{de\text{-}stress\text{-}experimental}$ are pursued, since an interaction effect can explain the lack of significant results in hypotheses testing. For this, selections on participants were made, with increasing levels of participant exclusion. Finally, testing on fascination and soft fascination will be discussed.

Interaction analysis (modality x water content) for $\Delta_{de\text{-}stress\text{-}experimental}$

For the exploratory analysis of $\Delta_{de\text{-}stress\text{-}experimental}$ a two-way ANOVA was conducted to investigate interaction effects of the modality and water factors on $\Delta_{de\text{-}stress\text{-}experimental}$. For clarity, $\Delta_{de\text{-}stress\text{-}experimental}$ is described by the difference between post-stressor experimental and post-restoration experimental stress level as also discussed in manipulation check section 3.4.

In Table 15, the means and standard deviations of the $\Delta_{de\text{-}stress\text{-}experimental}$ data are summarized. The audiovisual modality with a medium water level (Audiovisual-$Water_{Medium}$) experimental condition shows the largest restoration. The data are also displayed graphically in Figure 11.

Modality	Water level	Mean	Std. Deviation	n
Visual	Low	-7.14	7.939	21
	Medium	-5.95	5.817	20
	High	-6.05	7.380	20
	Total	-6.39	7.020	61
Auditory	Low	-4.71	7.663	21
	Medium	-5.82	8.278	17
	High	-4.67	7.045	21
	Total	-5.02	7.519	59
Audiovisual	Low	-8.14	8.039	21
	Medium	**-11.89**	8.429	18
	High	-7.77	8.788	22
	Total	-9.11	8.485	61
Total	Low	-6.67	7.888	63
	Medium	-7.85	7.915	55
	High	-6.19	7.781	63
	Total	-6.86	7.846	181

Table 15: two-way ANOVA means and SDs for $\Delta_{de\text{-}stress}$ data of the experimental condition

The data is not normally distributed, but a non-parametric test alternative for the ANOVA that measures interaction effects is unfortunately not available. First, outliers needed to be removed since the ANOVA test is quite sensitive to outliers. The assumptions that need to be fulfilled for a two-way ANOVA are not all met, since the data are not normally distributed. In addition, Levene's test for equal error variance of the dependent variable across groups is significant, indicating that the assumption of equal variances is violated. Nevertheless, to test whether an interaction effect could explain the results found in the hypothesis testing section, a two-way ANOVA was performed. A total of 37 outliers, eight participants with zero scores and one participant for which nature was distressing were removed from the analysis (total n = 181).

A two-way ANOVA was conducted to examine the effect of modality and water level on $\Delta_{de\text{-}stress\text{-}experimental}$ and the interaction between the two factors. Table 16 shows the results of this analysis. No statistically significant interaction was observed between the effects of modality and water on $\Delta_{de\text{-}stress\text{-}experimental}$, F (4,172) = 36, p = .661. Contrary to our findings in hypothesis 1, modality shows a significant effect F(2,172) = 4,6, p = .011 on $\Delta_{de\text{-}stress\text{-}experimental}$ in the parametric test. Nevertheless, considering a parametric test was not appropriate for reasons of violations of assumptions necessary to do a proper two-way ANOVA analysis, this finding should be approached with caution. Effect size η^2 or Cohen's f were not calculated since the data are not normally distributed and the used test was not appropriate.

Test variable: $\Delta_{de\text{-}stress=experimental}$	**df**	**F**	***p***
Modality	2	4.60	**.011**
Water level	2	0.76	.471
Modality * Water level	4	0.60	**.661**
Error	172		
Total	181		

Table 16: Interaction effect of $\Delta_{de\text{-}stress\text{-}experimental}$ between Modality and Water content

In Figure 11 the means and 95% CIs for all experimental conditions are presented in a bar plot. No apparent disordinal (cross-over) interactions are visible in the graph.

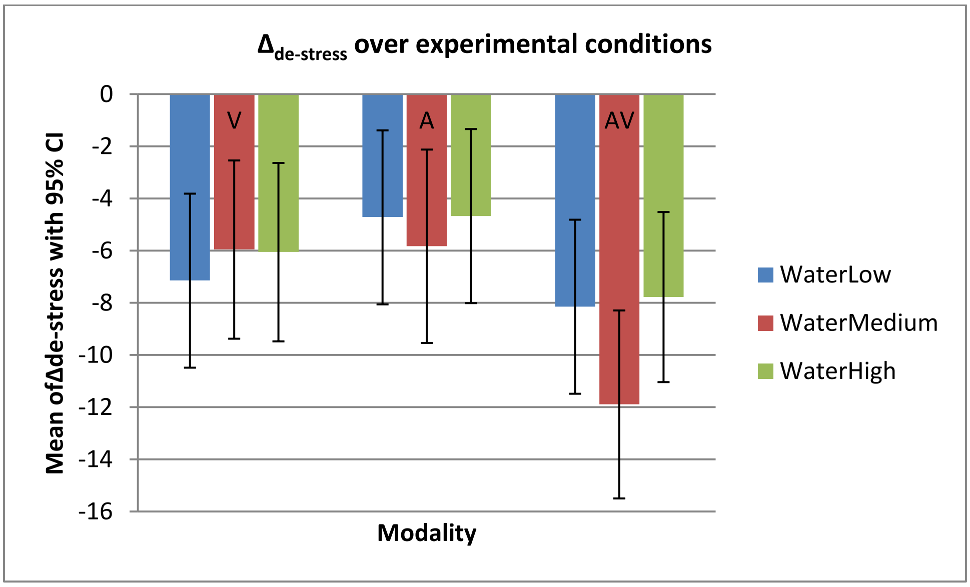

Figure 11: $\Delta_{de\text{-}stress}$ over experimental conditions, binned by modality, n = 181

The two-way ANOVA was repeated three more times. One time the interaction analysis was run with most participants in, including outliers and zero scoring participants. A second time with most participants included in the analysis, only excluding the eight zero score participants. The third time including only the participants that show the hypothesized stress relieving effect.

When selecting most participants, the *p*-value obtained from the two-way ANOVA was higher than the earlier found *p* = .661. Further testing with and without the 8 participants with zero scores on the experimental de-stress condition and outliers in the data revealed that an even higher *p* value for the interaction term was obtained. To exhaustively examine possible interaction effects, the selection with only participants that showed the anticipated effect (i.e. participants scored higher on stress level after the stressor, and scored lower on stress level after the de-stressor) was tested. Data was not normally distributed, although the rest of the assumptions were met. Even with this restricted sample, no statistically significant interactions between the effects of modality and water level was present on $\Delta_{de\text{-}stress\text{-}experimental}$, *F* (4,110) = .823, *p* = .513.

Interaction analysis (modality x water content) for fascination

A two-way ANOVA analysis was conducted that examined the effect of modality and water level with mean fascination as the dependent variable, as well as the interaction between modality and water level. No statistically significant interaction between the effects of modality and water content on the mean fascination construct was observed, $F(4, 214) = .516$, $p = .724$. The effect size η^2 for the interaction term is 0.006.

Interaction analysis (modality x water content) for soft fascination

A two-way ANOVA analysis was conducted that examined the effect of modality and water level with as dependent variable the full soft fascination construct, and the interaction term of modality and water content. No statistically significant interaction between the effects of modality and water content on the mean soft fascination construct was observed, $F(4, 217) = .821$, $p = .513$. The effect size η^2 for the interaction term is 0.007.

In conclusion, no significant interaction effects were observed. In all described variants of exploratory testing of $\Delta_{de\text{-}stress\text{-}experimental}$, no interaction effect between modality and water level was found. The former analyses also showed no interaction effect of modality, nor water content in the scene, on fascination or soft fascination. Also, from the data in Table 10, Table 11 and in Figure 11 no significant interaction effects are expected.

3.8 Comparing restorative effect of stimulus against control

In the manipulation checks, the overall and sequential stressing and de-stressing response in the control and experimental condition showed significantly different results. Further analyses test whether the experimental stimuli with elements stemming from nature were more restorative than control. First, the modal and water conditions are tested with respect to control. The data is not normally distributed, so Wilcoxon signed ranks tests were used. The results of the Wilcoxon signed ranks tests per modality and per water level are summarized in Table 17. For the A, V, AV, $Water_{Low}$, $Water_{Medium}$ and $Water_{High}$ results, a Bonferroni correction was implemented since six conditions are compared. The α-level under which a difference is statistically significant is then $p = .0083$.

Condition	n	Z	sig ($p < .0083$)	Median $\Delta_{de\text{-}stress}$ control	Median $\Delta_{de\text{-}stress}$ experimental
Overall	218	-1.160	.246	-7	-7
Audio only	77	-0.691	.490	-7	-6
Video only	68	-0.873	.383	-3.5	-6
Audiovisual	73	-0.534	.593	-10	-11
$Water_{Low}$	72	-0.190	.850	-7	-6.5
$Water_{Medium}$	71	-2.693	.007*	-7	-11
$Water_{High}$	75	-0.860	.390	-7	-7

Table 17: Comparison of $\Delta_{de\text{-}stress}$ of control and experimental conditions - * indicates a sig. effect

For the $Water_{Medium}$ factor a statistically significant difference is observed for $\Delta_{de\text{-}stress}$ between the control and experimental condition ($Z = -2.693$, $p = .007$). Median scores were -7 for control and -11 for experimental. Thus, for the $Water_{Medium}$ factor only, the experimental condition induced a significantly larger restorative effect then the control condition. The effect size for the $Water_{Medium}$ factor comparison is small, with $r = 0.1$. None of the other factors showed a statistically significant difference (all *p*'s > .0083).

The difference of $\Delta_{de\text{-}stress}$ between control and each of the nine experimental conditions was also tested with a Wilcoxon signed ranks test for each of the nine experimental conditions. The Bonferroni corrected alpha level was $p = .0056$. None of the nine experimental conditions showed significant differences with the control condition. This shows that none of the experimental conditions was able to restore participants better from stress than the control condition.

Modality	Water level	n	Z	*p*-value	Median $\Delta_{de\text{-}stress}$ control	Median $\Delta_{de\text{-}stress}$ experimental
Audio only	$Water_{Low}$	25	-1.279	.201	-5	-5

Audio only	$Water_{Medium}$	26	-1.659	.097	-8	-11,5
Audio only	$Water_{High}$	26	-1.792	.073	-10	-5,5
Video only	$Water_{Low}$	21	-0.967	.334	-6	-7
Video only	$Water_{Medium}$	23	-1.040	.260	-3	-6
Video only	$Water_{High}$	24	-1.158	.247	-1,5	-7
Audiovisual	$Water_{Low}$	26	-.077	.939	-15,5	-8
Audiovisual	$Water_{Medium}$	22	-1.998	.046	-9	-14
Audiovisual	$Water_{High}$	25	-1.010	.312	-9	-7

Table 18: Wilcoxon signed ranks test on $\Delta_{de\text{-}stress}$ between control and experimental

To resume, the only condition in which the experimental intervention was more effective in lowering stress levels in participants than control was $Water_{Medium}$ ($Z = -2.693$, $p = .007$). A small effect size of $r = 0.1$ was observed. In the next sections the fascination and soft fascination questionnaire data are examined for internal consistency as well as compared with respect to the means.

3.9 Internal consistency α for fascination and soft fascination

The PRS is a validated measurement tool; the fascination scale was used in the current thesis. As described in the Method section, question five was modified from "I want to spend more time *looking at* the surroundings" to "I want to spend more time *exploring* the surroundings" to also fit the Auditory conditions. The internal consistency could have changed, albeit little, by this modification. Therefore, Cronbach's alpha was calculated for the fascination questionnaire.

The fascination scale originating from the PRS consisted of five items (see section 2.5.2) with reliability $\alpha = .907$. Internal consistency is very high. The Soft Fascination questionnaire (see section 2.5.3) items were positively and negatively stated as indicated with a dagger symbol in Table 6. Questions five and six were re-coded to represent the correct values.

The Soft Fascination questionnaire consisted of six items with reliability $\alpha = .11$. Removing item six ("I was in awe") from the Soft Fascination questionnaire increases internal reliability to $\alpha = .496$. Also removing the first item from the Soft Fascination questionnaire ("I could still think about other things") increased reliability to $\alpha = .670$. This is an medium to acceptable internal reliability. For soft fascination two versions of the questionnaire will be statistically tested, named by the items included in the analysis: the full soft fascination questionnaire with items one to six and the reduced soft fascination questionnaire containing items two to five.

3.10 Means for fascination

The fascination data was analyzed for normal distribution. The data is not normally distributed. The Shapiro-Wilk test showed a significant departure from normality for all items ($W(227) = .905$, $p < .001$). The W statistic for all PRS items lies between .905 and .913, with all $p < .001$. Figure 12 below indicates the means and 95% CIs for the individual PRS question numbers. Please note that the Y-axis starts at three for improved resolution.

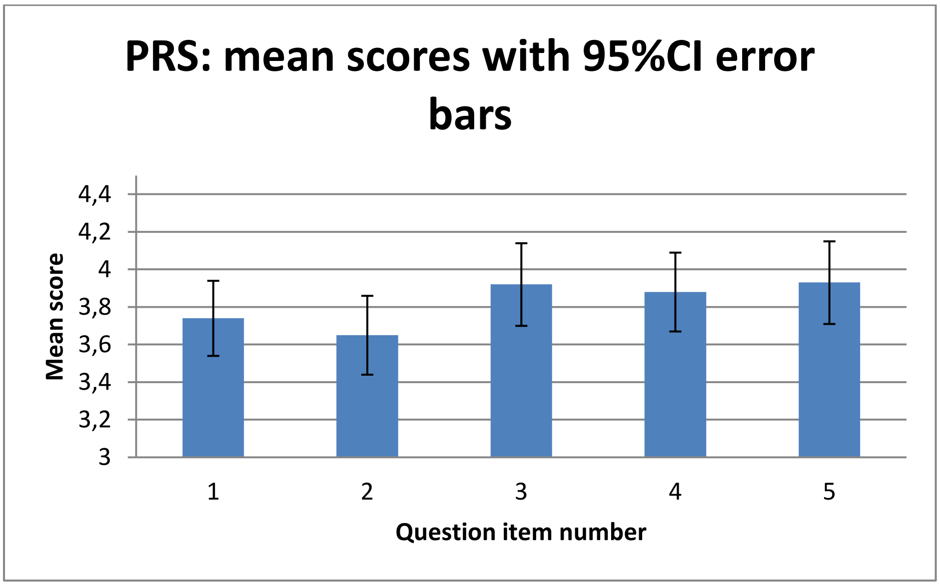

Figure 12: PRS mean scores with 95% CI, grouped by question number

To calculate an overall restorativeness score and to be able to compare the PRS scores with the Soft Fascination questionnaire and calculate their correlation, a mean for all items was calculated per participant. This practice has proved its benefit in research by Berto (2005), Sona et al. (2019) and Jahncke et al. (2015). A Shapiro-Wilk test again showed a significant departure from normality ($W(223) = .957$, $p < .001$).

3.11 Means for Soft Fascination

The Soft Fascination data was not normally distributed. The Shapiro-Wilk test had significant results for all items. The W statistic for all PRS items lies between .792 and .949, with all $p < .001$. Please note that items five and six are shown in their re-coded form (for an explanation, see section 3.3). Figure 13 displays the mean scores with 95% CIs. Please note that the Y-axis starts at three for improved resolution. With all items included internal consistency was low ($\alpha = .11$). By removing items one and six, an internal consistency with alpha level of $\alpha = .670$ was achieved.

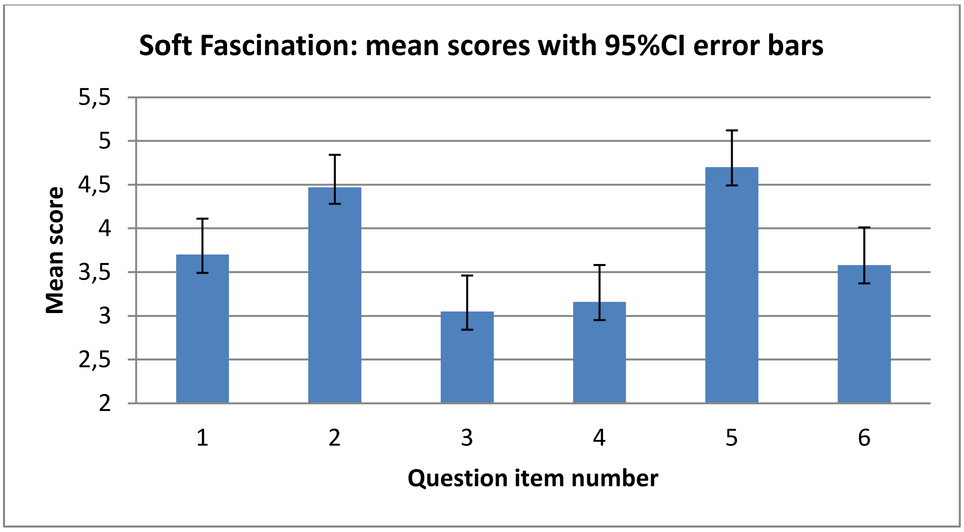

Figure 13: Soft Fascination mean scores with 95% CI

As final analyses, a mediation analysis was carried out and the correlation between the subjective measures of $\Delta_{de\text{-}stress\text{-}experimental}$, fascination and soft fascination is pursued.

3.12 Mediation analysis on the $Water_{Medium}$ condition

Since fascination is considered the main driver of mental restoration (Joye et al., 2013), a mediation analysis was carried out for fascination. Only data from the $Water_{Medium}$ condition was analyzed since it showed a significant effect when comparing stress relief measurements for the control and experimental condition. Soft fascination is an extension of fascination, so the soft fascination measure was also included in the mediation analyses. The analyses below used the experimental stress delta as input variable and the experimental stress relief delta as output variable, with fascination or soft fascination as the mediator. For mediation analysis linearity, normality, homogeneity of error variance and independence of errors is assumed. Please note that assumption violations (i.e. normality) are present, so the results should be interpreted with caution.

With mean fascination as mediator, the model coefficients were a = 0.0014 (SE = 0.0072), b = -3.5921 (SE = 1.0483), c = -0.5484 (SE = 0.0641). The output of the Sobel test showed a non-significant effect, Z = -.194, SE = 0.0259, $p = 0.846$ with lower limit CI = -.0697 and upper limit CI = .0758. This implies that fascination was not a mediator for mental restoration.

With the mean of soft fascination as mediator, the model coefficients were a = 0.0040 (SE = 0.0038), b = -4.0737 (SE = 2.0642), c = -0.5373 (SE = 0.0678). The output of the Sobel test showed a non-significant effect, Z = -.928, SE = 0.0175, $p = 0.353$ with lower limit CI = -.0565 and upper limit CI = .0163. No significant effect was found, so soft fascination was not a mediator for mental restoration. The complete soft fascination questionnaire showed low internal validity, so the test was repeated with the reduced soft fascination questionnaire.

With the mean of reduced soft fascination (items two to five) as mediator, the model coefficients were a = 0.0074 (SE = 0.0042), b = -3.9199 (SE =1.820), c = -0.5244 (SE = 0.0685). The output of the Sobel test showed a non-significant effect, Z = -1.36, SE = 0.0212, p = 0.172 with lower limit CI = -.0815 and upper limit CI = .0093. Here also, soft fascination was not a mediator for mental restoration.

None of the mediation analyses indicated a significant effect. The confidence intervals all cross zero, which also indicates that no significant mediation effects were present for fascination and soft fascination. Although the assumptions with mediation analysis were violated and results should be interpreted with caution, fascination and soft fascination most probably are not mediators influencing mental restoration.

3.13 Correlation between dependent variables

A high correlation between $\Delta_{de\text{-}stress\text{-}experimental}$ and $fascination_{mean}$ illustrates a link between the dimensions of the fascination questionnaire and a de-stressing response. Furthermore, a meaningful correlation between the fascination and soft fascination questionnaire can underline the usefulness of adding the soft fascination questionnaire in future research. Spearman correlations are used because of the non-parametric nature of the data.

$\Delta_{de\text{-}stress\text{-}experimental}$: fascination$_{mean}$

To calculate the correlation between $\Delta_{de\text{-}stress\text{-}experimental}$ and fascination$_{mean}$ all assumptions were checked first. A marginal monotonic relationship between the variables was present and data were non-normally distributed. Furthermore, the $\Delta_{de\text{-}stress\text{-}experimental}$ data was very noisy, so many outliers were present. It was chosen to remove a reasonable amount of the outliers and run a Spearman rank-order correlation analysis with n = 172. Outliers were removed after following criteria discussed in section 2.7 using an iterative process of printing boxplots and then removing outliers from the analysis.

A Spearman's rank-order correlation was conducted to determine the relationship between 172 $\Delta_{de\text{-}stress}$ responses and mean fascination scores. A significant, negative correlation between $\Delta_{de\text{-}stress\text{-}experimental}$ and fascination$_{mean}$ was observed ($r_s(170) = -.152$, $p = .047$). The correlation between $\Delta_{de\text{-}stress\text{-}experimental}$ and fascination is 15.2 percent, where a higher number indicates more overlap between the tested variables.

$\Delta_{de\text{-}stress\text{-}experimental}$: reduced softfascination$_{mean}$

Checks for assumptions needed for the correlation analysis were carried out. Softfascination$_{mean2to5}$ and $\Delta_{de\text{-}stress\text{-}experimental}$ were not normally distributed. Outliers were still present in $\Delta_{de\text{-}stress\text{-}experimental}$. A Spearman's rank-order correlation was run to determine the relationship between 166 $\Delta_{de\text{-}stress}$ responses and mean reduced Soft Fascination scores (items two to five). There was a negative correlation between $\Delta_{de\text{-}stress\text{-}experimental}$ and reduced softfascination$_{Mean}$ which was statistically significant ($r_s(164) = -.156$, $p = .045$). The correlation between $\Delta_{de\text{-}stress\text{-}experimental}$ and softfascination$_{mean2to5}$ is 15.6 percent.

fascination$_{mean}$: full softfascination$_{mean}$

Again, the assumptions accompanying the Spearman rank-order correlation were checked. Of the data, 219 data points were non-normally distributed, with no outliers present. A scatter plot showed a very good agreement in monotonic relationship between the variables. The Spearman rank-order correlation was run to determine the relationship between 219 mean fascination and mean soft fascination scores. There was a positive correlation of 64.9 percent between fascination$_{mean}$ and full softfascination$_{mean}$ which was statistically significant ($r_s(217) = .649$, $p < .001$).

The negative correlations found in the first two analyses are expected, since $\Delta_{de\text{-}stress\text{-}experimental}$ is coded negatively, where fascination$_{mean}$ and full softfascination$_{mean}$ are coded positively. Both correlations regarding $\Delta_{de\text{-}stress\text{-}experimental}$ are small and only correlate around 15%, which is a small effect. This implies that the stress measurement and the fascination measurements measure a different concept, which could be explained by the cognitive nature of the stressor and accompanying measurement scale, and the contrast with the affective measurement of fascination and soft fascination. It is interesting though that both these correlations are similar within 0.5%. Furthermore, a positive correlation between the fascination and soft fascination questionnaire was expected because soft fascination is a subset of fascination. A large correlation of 65% was found between the fascination and soft fascination questionnaire. These questionnaires measure a similar concept, although soft fascination measures a more specific part of the fascination concept.

3.14 Order effect for Δ_{stress}

In the manipulation check, a test was conducted between all sequential conditions to verify that the stressor was effective twice. The stressor was effective twice, but, as is clearly visible in Figure 7 and in the Δ_{stress} columns in Table 9, the stressor induced more stress in the control condition than in the experimental condition. This could be the result of an order effect. An analysis was done with a Wilcoxon signed ranks matched pairs test to examine a potential order effect. The $\Delta_{stress\text{-}control}$ and $\Delta_{stress\text{-}experimental}$ conditions are compared. Data for nine participants that had zero scores on all stress level measurements were excluded. The results show that a significant difference was present between the $\Delta_{stress\text{-}control}$ and $\Delta_{stress\text{-}experimental}$ condition ($Z = -4.311$, $p < .001$, $r = .082$). Median stress ratings were 11.0 and 5.5 respectively. The stressor did not change in form or intensity during the first and second condition, as the stressor was the same. An order effect for undergoing the stressor was therefore present, although it should be noted that the stressor was effective twice. Due to design limitations and choices made when setting up the experiment, it was not designed to prevent an order effect, or be able to determine whether an order effect is present. In the discussion section the implications of this finding are extended on.

4 Discussion

Regarding mediated environments, little attention has been paid to the relative impact of natural auditory and visual stimuli on stress restoration, which leaves a gap of knowledge on the mechanisms of technology assisted stress restoration. In addition to the aforementioned gap of knowledge, further investigation into the effect of the presence of water on mental restoration and fascination was warranted. The main aim of the current thesis was to investigate the relative differences between perceived stress reduction with respect to different perceptual modalities (audio, video, audio and video) and the amount of water (low water, medium water and high water) present in a mediated natural stimulus. Following the experimental conditions, a between-subjects experimental setup was employed in an online setting. Mental restoration and fascination was assessed with measurements of stress relief, fascination and soft fascination, the latter constructed based on current literature on soft fascination. The following sections discuss the main findings, limitations and future research and the implications of the research findings.

4.1 Overview of main findings

To investigate the effects of modality on restorativeness three hypotheses were tested. One general hypothesis to test whether an effect of modality on stress level, fascination and soft fascination is present, and two more specific hypotheses to test if the effects of modality were as predicted according to the literature (see section 4.1.1). In section 4.1.2 the hypothesis regarding the effect of water level in a stimulus on stress level, fascination and soft fascination is discussed. Figure 10 shows an overview of the statistical tests and main findings. The findings of the current thesis show no statistically significant differences in restorative effect of modality or for different levels of water content on the outcome variables of stress, fascination and soft fascination. Statistical testing for differences between the water levels neared significance though. A significant difference in stress relieving performance between the control and experimental conditions was found for the medium water condition. The newly developed soft fascination questionnaire showed a moderate internal validity, measuring two factors that are most probably indicative of soft fascination and daydreaming.

4.1.1 Effect of modality on mental restoration and on fascination

Hypothesis 1a stated: "There is an effect of modality on perceived mental restoration and on fascination". After testing, hypothesis 1a was rejected since no statistically significant evidence for differences in mental restoration and fascination between the modalities could be found for the $\Delta_{\text{de-stress-experimental}}$, fascination and soft fascination measurements. This is also apparent when inspecting the mean and SD data for $\Delta_{\text{de-stress-experimental}}$, fascination and soft fascination in Table 9 and Table 10. The distributions for experimental conditions are overlapping, indicating that no significant differences between the modalities are to be observed. The SDs are larger than the means for $\Delta_{\text{de-stress-experimental}}$ which indicates that data was noisy and showed a non-normal distribution. In the current thesis the hypothesized effect of modality on $\Delta_{\text{de-stress-experimental}}$, fascination and soft fascination (indicating mental restoration) was not found.

The study by Wooller et al. (2018) has a very comparable design, with the recovery condition as between-subjects factor and the pre- and post-stressor measurement as within-subjects factor. Wooller et al. (2018) found similar results for differences between restorative stimuli in what they call a group main effect. No differences between experimental groups with respect to stress reduction or affective state were found. However, a significant trial main effect of conditions was found indicating that the stressor increased stress and that experimental conditions reduced stress (J. J. Wooller et al., 2018). In the current study this finding corresponds to the manipulation check. An additive restorative trend for audiovisual stimuli was visible for perceived stress scale, although statistical testing failed to reach significance when comparing experimental condition groups (J. J. Wooller et al., 2018). Analyses showed significant interactions between experimental group and trial in the Wooller et al. (2018) study, indicating that possibly an order effect was present, with some combinations being more restorative than others. An ANCOVA analysis was used to correct for baseline stress level in the Wooller et al. (2018) study. The possible presence of an order effect is not discussed by Wooller et al. (2018). Differences in methodology are also observed; the Wooller et al. (2018) study was carried out in a lab, giving more control over external variables, effectively lowering noise. Finally, participants in the study by Wooller et al. (2018) performed exercise in all conditions with mediated nature, increasing arousal in the participants. This could have given rise to a confound, finding significant results and larger effect sizes.

For the current study, the *p* values and effect sizes are summarized for comparing different modal presentations in Table 12 and Table 13. Differences between conditions for stress relief and fascination were all small, as were the observed effect sizes. The induced stress was very fugacious where the removal of the stressor produced the largest effect. Since effect sizes are small, adding nature to relieve stress instead of simply stopping the stressor elicited very small benefits. Wooller et al. (2018) found medium size effects when comparing groups. In the current research a similar pattern is observed with small effects for comparisons between experimental conditions, and slightly larger effects for the manipulation check. In this regard, results are comparable, although Wooller et al. (2018) found larger effect sizes. Interpreting the results and lack of found significant effects, a likely cause for the large amount of noise on the data was caused by the online nature of the study, which is more susceptible to noise than a lab study. More control over the environmental variables is present in a lab. A smaller stress difference was observed for the experimental conditions than for the control condition, reducing the restorative potential of the experimental stimuli. This was most probably caused by an order effect. Most probably, finding smaller or no effects was due to comparing a restorative stimulus with another restorative stimulus. Noise due to the online nature of the experiment was also presumed to be a large factor. Both mentioned issues will be discussed more thoroughly in section 4.1.3 and section 4.2 respectively.

Hypothesis 1b: Auditory stimuli are more restorative and fascinating than visual stimuli

Since there were no statistically significant differences between the modal presentations, the hypothesis that "Auditory stimuli induce higher mental restoration and fascination than Visual stimuli" was rejected. The auditory modality did perform marginally better with respect to mean stress reduction than the visual modality as displayed in Table 9. This is in line with research by Ma & Shu (2018), who found that sounds have a greater effect on psychological restoration in comparison with purely visual scenes. Contrasting research by Wooller et al. (2018) found that the visual modality showed a larger mean restorative effect than the auditory modality. Preis et al. (2015) also reports effects of audio on comfort assessment, which were not found for the visual part of an audiovisual sample. Concluding, findings in the scientific literature are contradicting. The different findings could be caused by a relatively large influence of individual differences between participants, possibly impacted by unknown or overlooked variables in the content of the stimuli. Stimuli are not always well documented, making it hard to detect the source of the differences. Moreover, differences in stress reduction between modalities are hard to measure since a comparison is made between restorative stimuli, not a restorative and a non-restorative stimulus as in most restoration research. Another factor convoluting these comparisons is that the largest restorative effect comes from stopping the stressor.

For fascination and soft fascination no significant differences were observed between modalities as is visible in Table 13. The fascination results were therefore in line with research by Wooller et al. (2018). Lindquist et al. (2016) found a similar effect of visual condition on realism and preference, as well as a relatively strong interaction between view and sound. For soft fascination, no comparable research is available. The effect sizes of the current study are reported in Table 12, and are comparable with the mostly small reported effect sizes for modality in literature. As with stress relief comparisons, a comparison was made between stimuli that are regarded as positive, or restorative. Small differences are to be expected, making it hard to find significant differences. The exclusive use of fascinating stimuli most probably skewed results, leading to the non-normal distribution of the data. These findings also imply that the stimuli are all experienced as affectively positive.

Hypothesis 1c: Audiovisual stimuli are more restorative and fascinating than auditory and visual stimuli

Finally, for the modality testing, hypothesis 1c is examined: ”Audiovisual stimuli induce higher mental restoration and fascination than both Auditory and Visual stimuli”. In the current study, the audiovisual stimulus did not perform significantly better with respect to mental restoration and fascination than either the auditory or visual modality. Small effect sizes were found, as displayed in Table 12. The restorative effect for audiovisual stimulation on $\Delta_{de\text{-}stress\text{-}experimental}$ was larger than in the video only condition, where for the audio only condition an even larger but non-significant mean restoration was found. The effect thus partially goes in the hypothesized direction which is also in line with research by Deng et al. (2020), who found that audiovisual stimuli stemming from nature are rated as having better health benefits than visual stimuli alone. From a small sample size study (n=26) Liu et al. (2019) concluded that the addition of sound to pictures positively influences involvement and immersion while reducing mental load, showing that on subjective measurements a restorative influence of audiovisual stimuli is present. Sona et al. (2019) add that holistic sensory impressions are relevant and foster recovery, encouraging the use of audiovisual or other multimodal restoration schemes for mental restoration.

Finally, Annerstedt et al. (2013) found that adding auditory stimuli to a VR environment significantly adds to restorativeness, using the objective measures of saliva cortisol concentration and heart rate data supplemented with heart rate variability. Lastly, effect sizes reported by Annerstedt et al. (2013) were much higher than in the current study. These findings can be explained by the difference in experimental setup. The Annerstedt et al. (2013) study was carried out in a lab, where the current study was online. Also, in the Annerstedt et al. (2013) study, the experimental condition was implemented as within-subjects factor, compared to a between-subjects design in the current study. Also, a 3D VR environment can be perceived as more immersive than the videos used in the current study, influencing the being away component as described by Kaplan (1995).

To summarize, the hypothesized effect of modality on mental restoration and fascination did not show statistical significance, so no support for an influence of stimulus modality – comparing the experimental stimuli – on any of the mental restoration measurements including fascination was found in the data from the current thesis. Moreover, effect sizes of differences between conditions were smaller than expected from literature research. A similarly designed lab study by Wooller et al. (2018) also compared restorative conditions, finding non-significant but larger effects than in the current online study. Removing the stressor induced the largest restorative effect and adding a natural stimulus only adds a small benefit, if any. The restorative effect induced by natural stimuli is small, so differences between natural stimuli are hard to measure, especially in a non-lab experiment. Apart from modality, water content was also manipulated in the stimulus material, which is discussed in the next section.

4.1.2 Effect of water on mental restoration and on fascination

The effect of water content on mental restoration and fascination was tested in hypothesis two; "A higher (accessible) water content in the scene leads to higher mental restoration and fascination". The difference in experimental stress relief between the (experimental) water groups showed a near significant effect with a small effect size. The medium water condition showing the highest restoration mean. The effect of water level on fascination and soft fascination was found to be non-significant. In Table 14, an overview of effect sizes regarding water is displayed. Figure 9 and Table 9 illustrate that the low water and high water scene show similar or even lower mean stress relief values when compared with the control condition. The medium water condition showed better mean experimental stress relief values than the control condition but failed to show significantly better mental restoration when testing against the control condition. Also, no statistically significant differences between the conditions were present as tested in hypothesis 2, although the predicted direction of a higher restorative effect when more water is present in a stimulus is observed between the low water and medium water scene for experimental stress relief. Perceived restorativeness for the high water scene was much lower than initially expected. The trend on experimental stress relief as visible in Figure 9 is in line with what was found in research by White et al. (2010), who showed that more water in a picture leads to better subjective restorativeness, although a fully aquatic environment leads to a lower restorative effect. The finding that more water is only more restorative up until a certain threshold is in opposition of Ulrich et al. (1991), who states that humans have an unlearned predisposition to respond positively to water. The findings in the current study indicate that other factors than water presence most probably influenced the observed lack of restorative value of the stimuli. Individual differences are mentioned by Gidlow et al. (2016) as factors influencing restorativeness and could account for the lack of differences between conditions that were observed. Wilkie & Stavridou (2013) found that the type of water surface influences the restoration judgment, with differences observed between for example coastlines and murky (inaccessible) water. In the current research only stagnant water in a natural lake was displayed, so the water type was similar over stimuli. Participants could vary greatly in their perception of how accessible (i.e. clean, or ready to drink) the water in the scenes was, influencing the restorativeness of the water elements, leading to large individual differences and thus noise in the data. Wilkie & Stavridou (2013) comment justly that unpleasant water elements can be perceived as less restorative than stimuli containing no water elements at all. In this light, water in the high water condition was most probably not perceived as unpleasant, since it did not perform significantly worse than control. No significant differences were observed between the control and any of the experimental conditions.

Park et al. (2020) found significant differences on psychological and physiological responses between a visual and audiovisual condition, where the audio contained flowing water sounds. This finding is clearly not present in the current thesis. The study design of Park et al. (2020) had a within-subjects design with 10 conditions, leading to less noise in the data and larger measured effects, but also a learning effect. Barton & Pretty (2010) found in their meta study that the presence of water generated greater effects on self-esteem and mood in real nature combined with exercise. In the current thesis no exercise component was present, since that would also induce undesirable arousal. The work by Barton & Pretty (2010) measures subjective parameters on how a participant feels, where the current thesis measured subjective stress level, fascination and soft fascination.

Research by Watts et al. (2009), Ma & Shu (2018), Shu & Ma (2019) and Deng et al. (2020) made use of streaming water sounds. These are, among others, waterfalls, fountains and streams. Watts et al. (2009) adds that high frequency variable water sounds, for instance water falling onto small boulders, are rated highest. Ulrich et al. (1991) presented streaming water sounds as well as streaming water visuals. In the current thesis the use of streaming water sounds was not suited because the video material showed small lakes containing stagnant water. Calm water surfaces are discussed by Ulrich (1983), who makes mention of landscapes that contain calm water surfaces are lowest in tension. As discussed in appendix 8.1, the video material that was recorded for a different experimental setup limited the flexibility to change the stimulus to flowing water images and sounds. In a study by Wang & Zhao (2019) a congruent auditory and visual stimulus improved aesthetic experience. To still honor the congruency criterion as also discussed in Ma & Shu (2018) and Sona et al. (2019) it was chosen not to include streaming water sounds, only sounds that could have been made by water dropping and fish jumping from the water were included. The duration of water sounds is quantified in Table 1. The higher restorative effect of streaming water sounds could be explained from a survival perspective and the concept of accessible water as discussed in section 1.2 under the heading Water; i.e. streaming water usually contains less contaminants, bacteria and viruses than stagnant water (Ashbolt, 2004). Not using flowing water sounds could have been a source of low restorativeness ratings. Another explanation for the failure to obtain significant differences is given by Watts et al. (2009), who states that individual differences in responses to water sounds should be considered, as well as the context in which the sounds are heard. Because of the online nature of the current research, the context could vary greatly, giving rise to larger individual differences than in a lab study. As visible in Table 9, the SDs are larger than mean $\Delta_{de\text{-}stress\text{-}experimental}$, indicating very noisy data. Again, not all environments are suitable restorers for everybody, and it appears that individual differences can have a larger influence on the direction of the effects than what can be compensated for in a mediated environment. Moreover, little to no control was present over what the participants in the audio only conditions were gazing at due to the online nature of the research. If participants were distracted by something in their environment there was no way to control for that (i.e. one participant made a comment about distracting loud neighbors). Therefore, a lab study controlling for environmental variables that researches the role of streaming water on subjective and objective measures could be interesting to pursue.

The findings illustrate that not all environments containing water are equally restorative or suitable for mental restoration. For fascination and soft fascination, the "more is better" effect was not found in the current thesis, which is contradictory to White et al. (2010). The study by White et al. (2010) describes three levels of water content in a within-subjects design and measured preference, affect and perceived restorativeness. In literature, streaming water sounds, images and video are considered as restorative. Even though the current study contained stagnant water, a near-significant effect was present. Future research could take this finding into account.

Fascination was measured during the experimental condition only, so the current study cannot show an increase or decrease of fascination between the control and experimental condition. A comparison of means and SDs for fascination shows that similar results for the nature condition were found by Lee et al. (2015) and Sona et al. (2019). Both studies used the PRS with the same scale as the current research. From Table 10, the measured mean fascination for all stimuli in the current study shows that the stimuli were perceived as fascinating.

The effect sizes obtained in the current study are small when compared with other research. A-priori power analysis showed a sample size of 93 participants needed to find the effect of water on mental restoration. This effect size estimate was based on a study by Carles et al. (1999) investigating the influence of water content and modality on restoration, in which a η_p^2 of 0.15 (translating to $f = 0.42$) was found. The minimum sample size is much lower than the 227 participants that were included in the current research. Finding a near-significant difference between the water conditions for $\Delta_{de\text{-}stress\text{-}experimental}$ indicates that the effect size is much lower than expected, within the context and limitations of the current study. A possible cause is found in the low restorative power of the splashing water sounds used in the current study, where in previous research, flowing water sounds and high frequency water sounds – water falling on small boulders for example – were observed to be very restorative. Discussion on the type of water sounds is needed, since this seems to be a fairly recent development and mixed results are present in literature for both subjective and objective measures.

In summary, a near significant difference was found when testing the differences in restoration between water levels. All water conditions failed to better restore stress in the participants than in the control condition although a near-significant effect was found for the audiovisual-medium water condition. The scene containing a medium amount of water showed the highest mean restorativeness, while contrary to initial expectations a large amount of water was counterproductive and led to less restoration from stress.

4.1.3 Additional effects found in the data

Interaction between modality and water content

The possibility of an interaction between modality and water content in the stimuli was tested; no interaction between modality and water content was observed for both stress relief and fascination. Observed effect sizes were small. In effect, this does not give an alternative explanation why the hypothesized effects were not found. The outcomes of this test should be handled with caution, since several assumptions with ANOVA testing were violated, as described in section 3.7. The ANOVA was repeated with a reduced sample but again the interaction failed to reach significance. Figure 11 also shows that no apparent interaction effects are present in the data. An alternative explanation why the manipulation was not as effective as hypothesized is to be found. The observed order effect is investigated first.

Order effect

Due to an experimental design choice and technical limitations of the LimeSurvey (Schmitz, 2020) web interface the order of control and experimental stimulus were not randomized. As a result, the order of the experiment was the same for all participants and no comparison between control and experimental conditions on fascination and soft fascination could be made. The control condition was always presented to the participants first, where the experimental condition was always the second condition. An order effect was therefore anticipated on, which shows in the Δ_{stress} values in Table 9 and the analysis in section 3.14. The stress inducing cycles were the same for the control and experimental conditions. A significant difference was found between stressing participants in the control and experimental condition, indicating that an order effect was present. For the second stressor cycle the Δ_{stress} values were lower, indicating that the second SART stressor condition induced less stress than the first SART stressor condition, which corresponds to the experimental and control condition respectively.

The study by Wooller et al. (2018) also employed a design in which the control condition was measured first, and at a second visit the experimental condition was measured. This could also have caused an order effect, although this is not reflected on in their paper. The theory and findings formulated in other studies might not give robust predictions for other studies. In contrast to the current online study, the Wooller et al. (2018) study was carried out in the controlled environment of a lab. Evidently, when the induced stress is lower in the second trial due to an order effect, the expected level of mental restoration for $\Delta_{de\text{-}stress\text{-}experimental}$ is also lower; i.e. less range was available to indicate a de-stressing effect in the experimental condition. The $\Delta_{de\text{-}stress}$ values were mostly larger during the experimental de-stress condition, indicating that the experimental conditions have a larger restorative effect than in the control condition. The observed order effect could have led to smaller differences between $\Delta_{de\text{-}stress\text{-}control}$ and $\Delta_{de\text{-}stress\text{-}experimental}$, although it is not known if there is a linear relationship between stressing and de-stressing. In Table 9 overall $\Delta_{stress\text{-}control}$ is 1.51 times larger than $\Delta_{stress\text{-}experimental}$, while $\Delta_{de\text{-}stress\text{-}control}$ is 1.16 smaller than $\Delta_{de\text{-}stress\text{-}experimental}$. The impact of the stressor (i.e. SART task) on stress level was significantly different between the control and experimental condition, with less induced stress in the experimental condition, as was tested in section 3.13. In contrast, the difference between the de-stressors (i.e. the stimuli) was not significant, as tested in hypothesis 1a and hypothesis 2. This is most probably a consequence of the observed order effect. If the order effect is mitigated, a re-test on the hypotheses can be performed. Interpretation of the results should be handled with caution, since the relation between stressing and de-stressing is most probably non-linear. In this respect, it is important to note that stressing participants twice was a within-subjects factor. A randomization of order for control and experimental conditions could have mitigated this identified bias. The experimental condition though, would still be the only condition in which the fascination and soft fascination questionnaire would be administered. The progress bar displayed in the control condition was not inherently fascinating. It could induce confusion in the participants if the experimental condition with fascination and soft fascination questionnaire are administered first, and then the control condition would have been administered without questionnaires. A learning effect might also have resulted if the fascination and soft fascination questionnaire were also adopted after the control condition.

Differences between the control and experimental conditions

As an exploratory analysis a test whether the natural stimuli were perceived as more restorative than the control condition was conducted. Two separate tests were carried out in section 3.8, with the mean data for stress restoration displayed in Table 9. The medium water condition was significantly different from the control condition, showing that over auditory, visual and audiovisual conditions the medium water condition was more restorative than the progress bar (i.e. control). Other conditions yielded non-significant results. In Table 18 the differences between control and experimental conditions for stress restoration are displayed for six of the experimental conditions.

A second test analyzed the restorativeness of the individual nine conditions of the three (modality) by three (water content) matrix against the control condition. After Bonferroni correction, no statistically significant effects between the individual nine conditions and the control condition were found. In conclusion, none of the experimental conditions was able to restore participants better from stress than the control condition, although the audiovisual - medium water condition showed most promise.

Differences between experimental conditions

In hypothesis 1 and hypothesis 2 the differences between experimental conditions were investigated between modal conditions and water level conditions, both with non-significant results. As an exploratory analysis, the effect of individual conditions (combinations of modality and water level) on stress restoration, fascination, and soft fascination was examined and yielded no significant differences. This finding indicates that the individual experimental conditions did not differ in perceived restorativeness, so all experimental conditions were perceived as equally restorative.

4.1.4 Fascination and soft fascination

The role of fascination and soft fascination in the current study was smaller than expected, since both tested hypotheses and the exploratory analyses showed no significant differences between conditions. The ART theory by Kaplan & Kaplan (1989) suggests that mental restoration is influenced by fascination, although the definition of fascination and what makes a stimulus fascinating is still debated on.

In Table 10 the mean fascination scores are displayed and show that the means for all factors are above the average of three (scores from 0 = Not at all; 6). This implied that all experimental conditions were perceived as restorative. When comparing with research by Sona et al. (2019) mean scores for fascination and being away were between 3.12 (SD 1.28) and 3.74 (SD 0.94), although in the Sona et al. (2019) study no distinction is made between fascination and being away. Lee et al. (2015) reports a Cronbach's alpha of 0.76 for the PRS and a normal distribution with a mean of 3.46 (S.E. = .10). The mean fascination (PRS) scores in the current research were higher, indicating a higher degree of fascination through the stimulus material while displaying a comparable internal reliability. The fascination questionnaire showed a high internal consistency of $\alpha = .907$, which is in line with research by Sona et al. (2019) reporting an internal consistency of $\alpha = .90$. Differences between the experimental conditions were not significant as discussed in section 4.1.1 and section 4.1.2. If it was not fascination that influenced the restorativeness of the nature intervention, it is likely that one of the other (not measured) components of Kaplan's (1995) theory influenced the restorativeness of the experimental stimuli. Moreover, the fascination and soft fascination questionnaire were only administered in the experimental condition, so they could not be compared to the control condition. This is also considered a limitation of the current study. As discussed in section 2.5.3, the fascination questionnaire is a measure of restorative quality in environments, although the question items make no discrimination between soft and hard fascination.

From factor analysis a two-factor solution explaining 58.75% variance emerged for the soft fascination questionnaire, where initially a one-factor solution was expected. The two factors encompass soft fascination and mind wandering and explain 76.57% of the variance. Mind wandering, also called daydreaming, is discussed in section 1.2.3 and is considered a separate mechanism by Williams et al. (2018) that can add to the mentally restorative effect. Neurally distinct pathways are believed to be responsible for soft fascination and mind wandering (Williams et al., 2018). Since a two-factor solution emerged from the data, the idea that neurally distinct pathways are responsible for soft fascination and mind wandering is thus supported by the data. More research would be needed to investigate the process of mind wandering and the distinct neural pathways. For the full soft fascination questionnaire, the initial alpha level was determined at a low $\alpha = .11$. With item one and six (see Table 7) removed from the analysis an internal consistency of $\alpha = .670$ was achieved, which indicates a moderate but acceptable reliability.

A literature study into soft fascination showed the components that encompass soft fascination (see Figure 2). In the current study definitions for fascination and soft fascination were postulated, and a first attempt at measuring soft fascination with a newly developed questionnaire was undertaken. Drawing conclusions on soft fascination is difficult because the soft fascination questionnaire is not validated yet; a sample size of n = 227 is too small to validate the soft fascination questionnaire. Instead, a first attempt at subjectively measuring soft fascination was undertaken. A moderate reliability for the questionnaire was reached. Improvements to the questionnaire can improve reliability in future research. Objectively measuring soft fascination is not possible yet, and although fMRI methods can give objective results in the distinct neural pathways discussion, the interpretation of the data is still subjective (Williams et al., 2018). Since objective measurement of soft fascination is deemed impossible, the best subjective alternative to measure soft fascination was pursued. Improving the reliability will be further discussed in section 4.2.

Correlations between dependent variables

Correlation analysis between fascination and soft fascination showed significant agreement, which shows that the fascination and soft fascination questionnaire are in line with one another. A high correlation was expected, and shows that the current soft fascination questionnaire is a good starting point to measure soft fascination. From previous literature it was expected that when fascination rises, more stress is reduced. If the correlation between stress reduction and fascination was very low, either stress is not reduced as is observed in the current thesis, or stress reduction is not caused by fascination. Research by Kaplan (1995) indicates that stress is not only reduced through fascination, as displayed in Figure 2. The exploratory correlation analyses between the dependent variables illustrate the relation between perceived stress and the fascination and soft fascination questionnaire. A significant negative correlation of 15.2% and 15.6% was found between $\Delta_{de\text{-}stress\text{-}experimental}$ - $fascination_{mean}$ and $\Delta_{de\text{-}stress\text{-}experimental}$ - reduced $softfascination_{mean}$ respectively. This negative correlation was expected, since stress is a negative concept and restoration is positive. The low correlations found between stress reduction and (soft) fascination confirm that no significant restorative effects were found in hypothesis 1 and 2. Another factor possibly playing a role is that stress was induced cognitively, where the restoration was measured on an affective dimension. It would be surprising if $\Delta_{de\text{-}stress\text{-}experimental}$ and (soft) fascination would completely coincide, i.e. it is almost a requirement they do not measure the same phenomenon. For $fascination_{mean}$ - reduced $softfascination_{mean}$ a moderately significant correlation of 64.9% was observed, indicating the concepts of fascination and soft fascination do measure restoration in a corresponding direction. This finding was expected, as soft fascination is a subset of fascination as explained in Figure 2. The moderate internal validity of the soft fascination questionnaire should also be considered in its validity, improvements are discussed in section 4.2.

4.2 Limitations and future research

The results of the current thesis need to be understood within the context of the recognized limitations. Data was collected between December 17th, 2020 and January 18th, 2021. This was a time where a strict lockdown due to the COVID-19 virus was enforced in many countries across the world. The study changed from a lab study to an online paradigm where measurement of objective measures was not possible. This also had an influence on the stimulus materials that were selected after they were recorded, as well as the experimental setting the research was conducted under. The original setup of this thesis is discussed in section 8.1. The construct of being away, one of Kaplan's (1995) elements influencing restorativeness, was less apparent when participants are at home instead of in a lab, especially in a context in which COVID measures prevented people to go outside.

Because of the online nature of the experiment, difficulties were observed to control for all factors or check all confounding variables, although much effort went into mitigating biases. The online experiment was constructed to be an integrated experience without distractions, although technical conditions and effects needing preventative attention are discussed below. In particular, this section will discuss observed limitations of the online research paradigm, non compliant participant behavior, noise on the data, individual differences, limitations in the stimulus material, representativeness of the sample coupled to covariate analyses and finally design limitations.

Online research paradigm

The online nature of the experiment gave little to no control over extraneous factors that could have influenced the results. Therefore, some unforeseen circumstances arose. As many factors as possible were controlled for and checked, such as the browser used, the time participants used for a group of questions and the playback volume at the start of the experiment.

A request to not use a program that changes the color temperature of the screen (for example f.lux, Iris or SunsetScreen) was also included. Screen brightness was a known factor influencing restorative results, but no control for this factor could be implemented. This leads to added noise, although the results should average out over the participants in a group. No control over the room lighting or time of day the experiment was carried out in was present. This could have confounded the results, since environmental lighting conditions can influence the affective response to stimuli, with lighter scenes and a brighter environment being rated more positively (Lakens et al., 2013). In a lab study, the room lighting would be standardized over participants. Also, no control was present over the size, type (i.e. TN, IPS, OLED) and color reproduction properties of the screen a participant used. The metrics of screen size, settings and type of screen are extensively discussed in research by Kimpe & Marchessoux (2016). Eight or higher bits per color screens were preferred, although information about the type of computer screen is not available to most people. When using a five bit per color monitor un-natural looking Mach bands can be present, resulting in a bias.

In lab research, usually a calibrated headphone setup is used which is equally loud and sounds the same for all participants. The frequency response and sensitivity of headphones varies greatly; where some headphones will play some frequencies louder, other headphones will sound softer. In the current research a different approach was needed. Therefore, a similar perception of loudness was pursued and an attempt to match auditory levels over participants was undertaken to decrease noise in the data. It was not possible to check for equal loudness, although the acoustic levels were assumed to be more similar than when not matching them. An open question about what the participants heard was omitted. Background sound level could not be controlled for, where in a lab study this is usually presented as a dBA measurement. Other limitations were also present, like time spent in online research.

The amount of time a participant can participate in an online experiment is constrained, leading to limitations. A limited number of questions was asked in the demographics questionnaire. For example, participants were asked in which country they lived, but not in what type of surrounding (i.e. urban or rural). Also, a question indicating personal differences in connectedness to nature (like with the CNS scale) was not part of the questionnaire, which could bias results as discussed in research by Mayer & Frantz (2004) and Menardo et al. (2021). This is also a measure of individual differences that could not be taken into account due to time constraints and the online nature of the experiment. Thirst, which can have an influence on ratings of water perception (Aarts et al., 2001), was also not measured due to time constraints. External factors are always present in research, especially in an online paradigm. A few examples influencing the ecological validity of the current research are a partner, children or a pet entering the room which could diminish concentration. Other sources of distractions could be traffic or construction noises, neighbors making noise or the time of day the experiment was conducted. The time of day could have influenced the mental state of depletion of the participant. Internet speed or connection dropouts could also have an influence on the outcomes. Furthermore, in a lab setting participants could have asked a question if anything was unclear. This could have prevented some of the discussed biases.

The relatively stronger effects that were found in literature were not found in the current research. The lack of significant restorative effects in the current study could be due to the online experiment environment. Most research in the literature was carried out in the controlled environment of a lab which demonstrated stronger effects and significant differences. This implies that the restorative effect could be less robust when you take the participant out of the controlled environment. Distractions are much more likely in the home situation, where in the lab environment distractions are absent. Participants were not explicitly asked about distractions during the experiment, although some participants indicated that distractions were present. A follow-up study in a lab could introduce different levels of distraction and test whether the restorative effect is still found with increasing distraction.

Finally, the online setting of the experiment prevented the use of objective measures. The current study exclusively measured subjective stress and subjective fascination and soft fascination. More studies containing objective data can help to find patterns that influence restorativeness. In the literature research, 15 studies using objective measures were identified. Studies that directly measured the regeneration of attentional capacity were studies by Hartig et al. (1991), Hartig et al. (1996), Tennessen & Cimprich (1995), Hartig et al. (2003), and Laumann et al. (2003). Objective measures have the advantage that an interpretation by participants is not needed, although for example psychophysiological data still needs to be interpreted after data analysis. Also, objective measures can uncover data on the state of a participant the participant does not have conscious access to. Different results can be reached with subjective measures, although sometimes results of objective and subjective measures can give mixed results, as in de Kort et al. (2006). Studies using fMRI can identify the activation of brain regions under states of stress and restoration coupled to certain functions. Hunter et al. (2010), Martínez-Soto et al. (2013) and Gould van Praag et al. (2017) all found differences but can only connect what is known about a function of a brain region to the variables tested in their research. Therefore, fMRI measurements are not yet conclusive for researching mental restoration and fascination in the field of environmental psychology.

Gidlow et al. (2016) investigated the influence of adding water to a natural environment for a restorative experience using objective psycho-physiological measures and used an unmediated within-subjects design. Results showed a significant main effect of cortisol concentration on restoration with a medium effect size, where Heart Rate Variability showed inconclusive results. Psychophysiological measurements (such as PPG measurement with a phone camera) were considered for the current thesis but deemed unfeasible. Mixed results for psychophysiological measurements are present in the literature. Park et al. (2020) for example, found no significant differences for psychophysiological measurements, where psychological questionnaires did show significant differences. Research by Gidlow et al. (2016) found that psychophysiological measurements did show significant differences. Finally, the advantage of psycho-physiological measures is their objective nature. The recordings continue even if a participant is non-compliant.

Non-compliant participant behavior

As it was not always possible to verify if participants did what was asked of them, various attempts were undertaken to control for and identify non-compliant behavior in participants. Participants could have changed the volume during the experiment, leading to variations in the acoustical level over participants. Participants were instructed to be rested before partaking in the experiment. Unfortunately, controlling for stressed, tired or already depleted participants was not possible.

Some participants took unusually long to complete the SART procedure. To account for the observed long times for SART completion, participants could have waited to press the letter 'b' to start the explanation and later the main SART procedure, increasing time on the SART task. A rationale for this behavior could be requiring more time to recuperate from completing the first SART procedure or a distraction that was present in the environment of the participant. Another explanation could be participants switching between browser tabs (using Ctrl+Tab or Ctrl+Shift+Tab), as the SART test would pause and on return continue. Some participants showed a ‘give-up’ response during the SART procedure, not even trying to avoid pressing the space bar when the number 3 was presented, or not pressing the space bar at all. The problem with these participants was that they started restoring before the end of the SART procedure, leading to lower restoration differences in the data and participant exclusion. Participants also did not always interpret the directions with the SART correctly, leading to unusually short response times on all numbers presented. These participants showed low effort timing scores on their SART test and were rejected.

A loss of resolution was induced by data of eight participants where the stress scores were all zero. This implies that something is wrong with the question asked or with the interpretation of the word "stress" by the participants. The word “stress” as a state of being emotionally fatigued could have been misinterpreted for a state of life-threatening danger. This could lead to low scores on stress measurements. In future research a definition of the measured type of stress could be given before measuring stress. The data for these participants was analyzed with the participants in and out but did not lead to full data exclusion. Participants were rejected if they used speakers instead of headphones, participants using a phone or tablet and participants using the wrong browser. Lack of clarity about the expectations and goal of the experiment was most probably also large in participants who clicked through all text parts in the experiment without reading instructions. This was indicated by short group times and could have led to noise in the data. In a lab environment this is less prone to happen. Finally, in a non-lab environment, participants are not always eager to fill in honest or useful answers. The larger social distance in online research is also indicative of the large observed individual differences between participants. The consequences of this finding are discussed under 'Noisy data' below.

Noisy data

The Δ_{stress} data as displayed in Table 9 deviates largely over conditions, while the same stressor was used for all conditions. SD values for Δ_{stress} and $\Delta_{de\text{-}stress}$ are all larger than the accompanying means. This suggests that the stress level data was very noisy and the observed large amount of noise made for that no significant restorative effects over modality or water content were found in the data. Comparable studies were all lab studies instead of online studies. Control over distractions and participants following instructions was very low for the current study, leading to higher noise in the data. Several cases of non-compliant participants were detected, in some cases leading to exclusion. Further observations included that not all participants used the full scale of the stress level measurement, where other participants most probably did not start the experiment with a relaxed state of mind. This could have caused an inaccurate stress level baseline measurement.

A singular question on stress level was used, as in Beute & de Kort (2018). Their question was "How much stress do you experience at this moment" (Beute & de Kort, 2018). Beute & de Kort (2018) also pointed out that participants differed in their general stress levels, where SE levels are also relatively high and 95% CIs wide. Opposed to the singular question “Please rate how stressed you feel right now” as used in the current study, a validated way to measure perceived stress is with the Perceived Stress Scale (PSS) (Cohen, 1988). The PSS was successfully used by Gidlow et al. (2016) and Wooller et al. (2018), among others, and might result in less noise on the stress data. Compared to the mean, lower SDs are reported in lab studies by Gidlow et al. (2016) and Wooller et al. (2018). Consequently, total experiment time would increase by introducing the PSS. Using the PSS could also prevent participants interpreting the word stress wrongly or giving zero scores on all stress measurements. The fascination data was more consistent over participants. In Table 10 the means and SDs for fascination and soft fascination are displayed per modality and water level. The means with SDs are overlapping, so no statistically significant differences between modalities or water manipulations for fascination and soft fascination data in the current thesis were found.

The observed non-normal distribution of stress level data in the current study was suspect of an order effect, large individual differences and stress measurement issues. The sample of 227 was large enough to show a normal distribution, but the data was markedly not normally distributed. Suggestions for mitigating noise can be using a more controlled setting such as a lab, reducing distractions and creating a similar environment for all participants. Changing the experimental setup from a between-subjects to a within-subjects design or using a more homogeneous sample could also help mitigate noise. A possible other cause for noise in the stress level data was the way stress was assessed. The order effect as discussed in section 4.1.3 could also cause deviations in stress level data, but only for the experimental condition. Data containing the discussed levels of noise can indicate that the stimulus was not restorative or that individual differences were a large source of noise.

Individual differences

In the current study responses to — for example — the stressor varied greatly over participants. Individual differences between participants were larger than expected, which had an influence on the lack of finding significant effects. As mentioned earlier, the online paradigm could have amplified noise or added individual differences due to non-standardized hardware, time of day or other factors. A study by Menardo et al. (2021) suggests that “identification of environmental or individual variables that influence the perception of restorativeness" should be researched (Menardo et al., 2021). Influences as age and gender were measured in the current study, but level of education was not. In the current study differences between participants were identified influencing the results. Some of the participants indicated they were afraid of jumpscares during the experimental condition, which was also reported in the study by Annerstedt et al. (2013). The fear of jumpscares also showed in the post-stimulus stress level scores of the current study and very low answers on the fascination or soft fascination questionnaire (indicating low fascination).

When examining the issues participants reported with the experiment, some indicated they expected a scene with sound and vision, not just with sound, which created an incongruent experience. This congruency issue was also discussed by Annerstedt et al. (2013). Therefore, the experience was not as fascinating as expected. The expectations of a participant were not measured and are very difficult to control for, giving rise to larger individual differences and noise in the measurements. Ma & Shu (2018) also suggest that restorative environments are strongly linked to subjective preference. A measurement that was implemented to find these types of mismatches was the scene quality, although interpretations can vary.

One participant was identified where the nature intervention led to stress instead of relaxation, also reflected in the stress scores. This participant was excluded in most analyses. When checking for color blindness, some participants might not have been aware of having color blindness. The choice option "Not that I know of" was added to the color blindness questionnaire and selected by eight participants. The interpretation of this data did not lead to problems, since noise in the Δ_{stress} and $\Delta_{de\text{-}stress}$ variables was very large. It is assumed that the larger part of that noise had a different cause and a latent color blindness added very little to that noise.

A final individual difference to discuss is the mental state of the participants. From literature it is known that people with depression show larger recovery after nature exposure (Beute & de Kort, 2018). The mental state of the participants was not queried but could have influenced the outcomes of the research. Concluding, interpersonal differences are under-investigated. Menardo et al. (2021) suggest to use measures that actually measure restoration.

Limitations in the stimulus material

The stimulus material can be subdivided into stressing and stress relieving material. First the stressing material is discussed. The SART task was cognitively easy to execute, although inhibition and focus might be perceived as difficult, since several participants indicated watery eyes from staring at the fixation cross. A possible issue with the SART procedure could be that it was not very cognitively taxing. The premise of a sufficiently taxing cognitive task was discussed in research by Joye et al. (2013) when reflecting on the effort dimension (see Figure 2). If a task is not cognitively stressing, but only depleting for attention-holding-circuits in the brain, the induced stress could be short lived. An example is found in research by Valtchanov et al. (2010), where a stress-inducing math quiz and writing task were "not cognitively taxing enough to result in cognitive fatigue and loss of focus, leaving participants at near-baseline cognitive ability with little room for recovery" (Valtchanov et al., 2010). A comparable situation could have emerged when participants were stressed for the second time in the current study. Participants were significantly more stressed, although the task was very simple in nature. In future research the typical time span for relaxation after stressing could be investigated.

The stress relieving stimulus material was matched as closely as possible on a large variety of variables as described in sections 2.3.2, 2.3.3 and 2.3.5, such as video brightness, color temperature and sound level. The medium water condition invoked the highest mean with regard to restorativeness, which was initially unexpected. The higher mean score could be explained by the restorativeness of water content in a stimulus following an inverted U-shape, as described in research by White et al. (2010). An alternative explanation for the higher restorative mean of the medium water scene is the color temperature of the scene, which is one of the characteristics of light discussed in the study by Menardo et al. (2021). Color differences were present in the visual stimuli, as discussed in section 2.3.3 and appendix 8.4. The medium water scene had a more red hue and performed better with respect to mental restoration than the other two scenes (low water and high water).

The low water scene contained no animals where the medium water and high water scene did. This could have introduced a bias, although it is not likely that the lack of animals resulted in a lower restorativeness in the low water scene. Finally, the biodiversity (see Figure 4) appeared a little more salient in the medium water scene when compared to the low and high water scene, which could have played a role. No trees are present in the background of the low water condition, possibly limiting the restorativeness.

Another identified factor influencing the results was the interpretation of what participants heard. Control for what the participants heard was not implemented by using a check question for example. In mental restoration research, incoherent sounds, noise or not hearing anything is considered negative, or at least less restorative. Beautiful nature, insects, wind sounds or water sounds are considered positive, or at least more restorative.

Representative sample and covariate analyses

The distributions for gender, age groups, fixed PC or laptop, over ear headphones or ear buds and screen size were not indicative of biases as discussed in section 2.1. An even distribution of participants over measured characteristics was present. Variables measured during the experiment are displayed in Table 34.

Screen size was shown to influence the experience of presence and immersion as researched by de Kort et al. (2006), possibly influencing the results of the current research. Analyses investigating the mentioned factors as a covariate were not carried out in the current study. Liu et al. (2019) also researched immersion and realism, which were not measured in the current research, other than with a scene quality question. This item could have also helped in identifying a beauty bias, which is discussed in Twedt et al. (2019). Confounds influencing familiarity (Kaplan, 1995), complexity (Ulrich, 1983), safety (Ulrich, 1983) and depth (Ulrich, 1983; Menardo et al., 2021) were not measured in the demographics or other questionnaires but could influence restorativeness. Due to time constraints these questions were not added.

Most of the participants in the sample came from Western cultures. Little research is available on the effect of cultural differences on the perception of restorative environments, although Huai & Van de Voorde (2022) point out that across a Chinese and European city similar categories contribute to positive perceptions of parks. The level of computer skills was not measured in the current study but could have led to biases in novice users. For example, if a novice user refused to use the full-screen setting in the browser this could have led to more distraction. From all reviewed literature, no factors were found influencing the results where a more homogeneous sample could help in reducing noise. Missing a part of the population making the sample non-representative is considered highly unlikely.

The only consistent factor influencing the outcomes of the experiment was the online nature of the experiment. Selecting only the participants in which the restorative effect went in the expected direction revealed that even then the anticipated restorative effect was not statistically significant. This finding is discussed in section 3.7. With this final test using a selection of the sample, it was considered reasonable to assume that covariate analyses would not uncover significant restorative effects.

Design limitations

After acute stress is induced, it dissipates very quickly, possibly because the stressor is simply absent. Since acute stress is so fugacious, the addition of a natural stimulus could have had little impact on stress relief. The induction of acute (short-term) stress proved to be viable and is ethically justifiable. Increasing the stressor time, with the goal to increase stress in participants and consequently prolong the restoration time, could give more deterministic results with respect to differences between the restoring stimuli.

Other types of induced stress are also to be considered, as “future studies might investigate different types of stress (e.g., vocational and interpersonal stress) separately” (Sona et al., 2019). To expand on this notion, research into which types of stress can be distinguished, which brain areas they fatigue, how long it takes before fatigue sets in and how long that type of stress persists could advance research on mental restoration and the field of environmental psychology. Inducing chronic stress in participants is unethical and time consuming, although it would be interesting to measure if a natural stimulus has a similar effect on chronic stress as on acute stress.

Social stressors such as the Trier Social Stress Test were effectively used in research by Annerstedt et al. (2013) and are more affective in nature. In a new study an affective stressor could be tested, for example letting participants think of sad moments in their lives. In the current study, purely cognitive stress was induced, where later the affective response is assessed. The mapping of stressor and de-stressor, or stressor and questionnaire, could be incorrect although the impact of this mismatch is unclear since no comparable research addressing this issue was found. The only measure in the current study that directly maps onto cognitive stressing is the stress level measurement, depending on the interpretation of the participant. The affective measurement in the fascination and soft fascination questionnaire did not correspond to cognitive stress. From a more clinical perspective, nature is presumed to always be restorative. That includes restoration by nature after cognitive stressing. From the literature, it is expected that “Nature, which is filled with intriguing stimuli, modestly grabs attention in a bottom-up fashion, allowing top-down directed-attention abilities a chance to replenish” (Berman et al., 2008). Many studies covered in the literature research used a cognitive stressor, but did not use affective measurements to assess stress reduction, making a comparison with the current study difficult.

As in general mental restoration research, Bratman et al. (2015) discussed the use of two very different environments (urban with high traffic density – rural quietness). In contrast, the environments used in the current research were exclusively rural of nature and thus very similar. A negative or non-restorative condition was not added in the stimuli. The measured differences in restorativeness between the stimuli then are expected to be accordingly small: i.e. differences between modality and water levels were measured, but all environments were considered restorative. This observation is in line with research by Ulrich et al. (1991), who found no differences in restorativeness between water and vegetation, where Skin Conductance Response and lower Pulse Transit Times showed slightly higher but non-significant differences in restoration when comparing water and vegetation.

Recommendations for future research

The experiment was conducted using an online paradigm, although it would have been preferred to carry out the experiment in the more controlled conditions of a lab. A lab study can uncover if noise was present because of the online nature of the current study, because of individual differences between participants, or both.

Future research could also focus on the medium water condition and the audiovisual - medium water conditions, since these conditions showed the largest restorative mean as shown in Table 15, as well as a near-significant effect in the results of Hypothesis 2.

External factors such as distractions could have influenced the results. In future research, it is paramount to have the participants distracted only by the stimuli. Additional analyses could be conducted using induced stress level, gender, age, indicated modal preference, screen size or perceived scene quality as covariates.

To investigate more of the individual differences the demographics questionnaire can be expanded in future research with questions about nature connectedness or indications of a rural or urban living area for example. The use of validated measures where available is recommended to increase validity and comparability with other research. The order of the control and experimental condition is to be randomized to prevent the significant order effect as described in section 3.14.

The baseline stress level measurement for control and experimental appeared similar in Figure 7. Therefore, the baseline level was not indicative of biases. The diminished stressing capacity of the second stressor however, was problematic. No statistical test was used to verify significant differences between pre-stressor stress-control and pre-stressor stress-experimental.

A covariate analysis on the current $\Delta_{\text{de-stress}}$ data would be instrumental to correct for the within-subjects factor of stress induction, since the second time participants were stressed with the SART test they were less stressed as explained in section 3.14 about order effects. A correction caused by the observed order effect could be applied to the $\Delta_{\text{de-stress-experimental}}$ data, after which measurement resolution could increase and more of the experimental conditions could show a significantly better restoration than the control condition. The latter would most probably improve the manipulation check. Further analyses would then investigate if differences between the experimental factors of modality and water can be observed.

Mediation analysis for the effect of fascination on mental restoration (see section 3.12) can be expanded on with non-parametric mediation analyses. Joye et al. (2013) consider fascination as the main driver of mental restoration. A mediation analysis could be used to confirm and show how much of the model is explained by the mediator. For this analysis to be feasible, a validated fascination and soft fascination questionnaire are needed. A large source that negatively influenced significant results of tests was the fairly large amount of noise in the mental restoration data. The effects of this noise are difficult to mitigate, although noise can be decreased by increasing the sample size. When calculating the needed sample size, effect sizes from literature describing between subjects design lab studies by Annerstedt et al. (2013) with n=30 and Wooller et al. (2018) with n=50 were used as reference material. The effect sizes found were larger than in the current study with n = 227. The current research had a mixed design with the natural stimuli presented as between subjects variable.

The effects for restoration through natural stimuli found in literature differ considerably over papers. Explanations for the differences in restoration literature can be found in differences in the stimuli, differences in stressing material, differences in experimental setup and individual differences in participants (for example the affective response, interpretation of natural stimuli, modal preference, stimulus sensitivity, or culture). To further the field of environmental psychology, more research is needed to investigate which differences influence the results most.

The perception of natural stimuli may be subject to large individual differences so characteristics of people and their former experiences should also be considered in future research. Measurement tools should be developed to identify “environmental or individual variables that influence the perception of restorativeness” (Menardo et al., 2021). Finally, meta-analysis could shed more light on which experimental manipulations worked in past research, possibly uncovering previously overlooked dimensions. Then, effect sizes need to be published or calculated in order to make comparisons. Furthermore, it is recommended that future research covers “measures of actual restoration rather than perceived restorative potential” (Twedt et al., 2019). Doing so, the results of research do not rely on self-report, which is also one of the criticisms of Joye et al. (2013). Furthermore, more extensive fMRI research could unravel the neurobiological mechanisms or pathways underlying the behavioral and physiological responses associated with restorative stimuli, and in effect, why nature induces a restorative effect (Gould van Praag et al., 2017).

Fascination is assumed by Joye et al. (2013) to be the main driver of mental restoration, with soft fascination as an integral part of fascination. In the current study the effort, extent, compatibility and being away dimensions Kaplan (1995) theorizes (see Figure 2) were not researched or captured in a questionnaire. If these were measured, it could give more insight in why all environments in the current research were statistically equally restorative, or which environmental characteristics are more restorative than others and if they are interchangeable. In this regard, Petersen & Posner (2012) note that environments that are difficult to comprehend by being low in coherence or extent will activate the executive network as the individual seeks to make sense of the surroundings, lowering restoration. Therefore, also testing the current stimuli for effort, extent, compatibility and being away would add to the current body of knowledge.

Fascination and soft fascination data were obtained in the experimental condition only, so a comparison between fascination and soft fascination data for control and experimental data was not possible with the current experimental design. In future research, the role of the affective dimension of fascination and soft fascination can be further investigated by manipulating how fascinating a stimulus is and testing for mental restoration. Sufficient distance between how fascinating the stimuli are perceived is needed to observe significant differences. In case of the current research, the relative mental restoration distance between the progress bar and the natural stimuli was small.

Items one and six were excluded after factor analysis because the factor loadings were low. These questions did not measure what was expected they would measure. For item one “The scene was pleasant” (see Table 4), it could have been unclear to participants what was meant with ‘scene’, especially in the audio only condition, because only a white screen was displayed. Changing the word ‘scene’ could also improve the reliability. Moreover, in the video only conditions a non-congruent stimulus was offered, possibly also lowering restorativeness. Participants went through effort to calibrate their headphones, and probably also expected to hear something with the video material. This could have lowered the pleasantness as measured in item one of the soft fascination questionnaire. Item six of the soft fascination questionnaire encompassed a check question for the inducing of undesired hard fascination: “I was in awe”. Rephrasing this question to “I was overwhelmed” might be more applicable, especially for non-native English speakers. Furthermore, item three of the soft fascination questionnaire read “I experienced an ebb and flow of attention” and was suggested by Williams et al. (2018). For participants an ‘ebb and flow of attention’ might not be clear and give room for different interpretations, leading to a lower reliability and thus more noise on the data. Although the factor loading (see Table 7) was high on factor two (daydreaming), a not-negligible loading on factor one (soft fascination) is also observed. A rephrase of this item to “I was daydreaming for some of the time” is suggested. Future research can also assess whether daydreaming is a form of being away and whether daydreaming induces a restorative effect more powerful than soft fascination by a natural stimulus. Daydreaming is a short lasting phenomenon, theoretically alternated with soft fascination. It is still unclear how effective daydreaming is in stress relief when experiencing acute stress. The attentional bias dimension Joye et al. (2013) discusses is already covered in the fascination questionnaire under item number two: “My attention is drawn to many interesting things”. Therefore, it was not included again in the soft fascination questionnaire. In new research where Kaplan's (1995) components are administered with the full PRS questionnaire, this question could be added to the soft fascination questionnaire.

The effect sizes found in the current thesis are smaller than in reviewed literature, and thus not in line with research by Shu & Ma (2019) and White et al. (2010) who both used a within-subjects design. The current thesis used a between-subjects design with respect to the experimental conditions. In a within-subjects design larger effect sizes are to be expected.

Finally, scene A instead of scene F (see section 8.2 — Figure 14) could be tested as high water condition. A comparable amount of water is present, but the color temperature of the video material is lower than in scene F. Another approach could be using specialized image processing software to give the low and high water scenes a warmer hue without changing the color of the water. This can test the hypothesis that scene color positively influences restorativeness.

4.3 Implications of research findings

The current research found a near significant effect of the audiovisual-water medium condition in a between-subjects test. The medium water condition showed significantly better mental restoration than in the control condition. The stimulus material from the medium water condition can be used in future research or applications in which restoration is sought, the workplace (N. Zhao et al., 2017) or a waiting room for example. Nature sounds could be added to softly draw attention to the visual presentation and create a coherent experience. Relaxation in a classroom could also be considered, as described by Shu & Ma (2019). Especially in spaces without windows, the addition of a screen that gives the illusion that a person can look outside can be beneficial as described by Sona et al. (2019). In a medical environment, the medium water scene could be used as described in Diette et al. (2003).

The literature suggests that inducing measurable acute stress in five minutes is viable (Robertson et al., 1997). New is that stress was induced twice with the SART task. Reducing acute stress to baseline levels is viable in one to two minutes (Sun et al., 2018; Wang & Zhao, 2019). Restoring with nature is beneficial to people since nature can amplify the restorative effect as well as shorten the time needed to mentally restore. The restorative effect however is not always present, even with nature stimuli that are considered to be restorative and were pilot tested. The nature dose is dependent on individual differences, although an optimum for visual tree percentage (Jiang et al., 2014) and percentage of water content (White et al., 2010) was found in literature. Most probably, an optimum for more factors can be found in literature or after research.

The study largely followed the experimental design of Wooller et al. (2018) and showed that — excluding the exercise component — the restorative effect could be less robust when induced in an online setting.

5 Conclusion

Using mediated natural environments is promising for stress restoration at home in the absence of real nature, but little is known about the effect of modality and water content on mental restoration and fascination. Soft fascination is of particular interest in mental restoration research because the hard fascination is not restorative, where the construct of fascination is suggested to be the main driver of mental restoration. The current study examined the effect of stimulus modality and water content on mental restoration, fascination and — for the first time — soft fascination by using restorative stimuli stemming from nature in an online setting. A total of nine stimuli were created for the purpose of stress restoration with modal and water level differences. The combination of three modal conditions (audio, video, audiovisual) and three water level conditions (with three orders of magnitude of water content: no water, medium water content and high water content) were tested in a between subjects experiment. Audio and video were congruent in the audiovisual conditions. Participants completed a depleting go/no-go task to induce stress before being exposed to a control and to the experimental condition. After the second go/no-go task participants were exposed to one of nine experimental stimuli. Evaluations of stress restoration, fascination, and soft fascination were measured, the latter with a questionnaire developed in this thesis.

Contrary to the hypotheses, neither the modality of — nor the water content in the stimuli — or their interaction had a significant effect on mental restoration and fascination. The medium water experimental condition was more effective in stress relief than the control condition. The audiovisual - medium water condition was found to be more restorative than control, but failed to meet the Bonferroni-corrected significance threshold. In the literature similar results were found, although with higher effect sizes. The soft fascination questionnaire shows promise and can be further developed.

Exploratory analyses comparing the stimulus materials with the control condition showed that the audiovisual - medium water condition was most restorative and a near-significant difference with the control condition was observed. The low water condition was most fascinating. Although great care was taken to mitigate possible biases, high levels of noise were observed for stress level measurement, indicated by SD values larger than the means and non-normal distribution of the data. Possible explanations for this finding were investigated, which could explain the high levels of noise. Some explanations also illustrate differences between findings in literature and the current study. Removing a stressor most probably has the largest effect. The induced stress response is very fugacious, meaning that acute stress will diminish fast after the stressor stops. Adding a restorative natural stimulus probably adds little to the restorative effect. Individual differences regarding perception of the stimulus materials are large, also explaining low effect sizes in literature. In the current study, restorative stimuli were compared instead of negative and positive stimuli, leading to smaller expected differences. An order effect was observed when analyzing the stress level measurements, leading to diminishing stress induction in the experimental condition and consequently lower resolution in the stress relief measurements. Some participants showed low commitment with the experiment, as indicated by short response times and low effort responses to questionnaires, possibly influenced by strict Corona regulations at the time of the experiment. External distractions also created noise on the stress level data. The mentally restorative effect of nature could also be not robust enough for online research. In resume, several variables played a role in adding noise. The experimental stimuli were perceived as fascinating when compared to the literature, although no statistically significant differences for fascination between modality or water content of a stimulus were observed. Also, no differences for soft fascination over experimental conditions were observed, i.e. all conditions were equally fascinating and softly fascinating. The modification of the PRS fascination items did not negatively influence the internal validity. The soft fascination questionnaire showed a moderate internal validity and loaded on two factors, most probably indicative of the mental states of soft fascination and mind wandering.

Although the study had limitations due to the online paradigm, the current study induced stress and measured the effects of modality and water content on mental restoration, fascination and — for the first time — soft fascination in an online experiment.

Although no significant effects of modality and water content on mental restoration, fascination and soft fascination were observed, of the tested stimuli only the audiovisual - medium water stimulus showed promising restorative potential and could be used in future restoration research. For follow-up research, a lab study with a within-subjects design is recommended, with objective measurements measuring actual stress restoration. A promising first soft fascination questionnaire was developed. After reflection on the results, modifications to the soft fascination questionnaire were proposed. Future research is encouraged to test the validity further with a larger sample and use additional stimuli.

6 Acknowledgements

First and foremost I would like to thank my supervisors Kynthia Chamilothori and Wijnand IJsselsteijn for their flexibility, ideas and guidance during the process of graduation. Especially in the last weeks of my thesis they were flexible and committed to give necessary feedback. Furthermore, this research has greatly advanced with the help of HTI staff members, friends and family.

Videos were recorded and edited with help from former HTI staff member Wouter van den Hoogen, who also founded Wouter van den Hoogen fotografie. The necessary permits to enter the protected nature reserve of the Strabrechtse heide were arranged by Wouter, as well as all materials needed to record the video material; like memory cards, lenses and gradual grey square filters. His expertise on photography, video making and editing was indispensable for making the natural stimuli.

The Zoom H4n used for recording the audio was provided by Martijn Willemsen from the HTI group.

The microphones used for recording the mid-side channel audio were provided by Ilja Roest. He also provided expertise on the audio recordings and gave information on how to implement matrix processing on the recorded audio in post-production.

Help with programming the SART script was given by Mike Wilmer, Sjaak Veraa and Nasir Abed.

Nasir Abed also helped a great deal with the API interface of LimeSurvey and researching the possibilities of implementing the experiment in Millisecond Inquisit, Lab.js and PsychoPy.

The implementation of the SART script and the progress bar in the LimeSurvey framework was facilitated by Chris Kruining. Storing the values of variables in LimeSurvey was also accomplished with the help of Chris.

Pre-pilot testing the four 1-minute videos for suitability and restorative qualities was done by Willemien Aalbers, Roel Grootenboer, Victor Lochmans, Jennifer Melskens and Daniel de Haas.

For the implementation of storing the user agent string and measuring the color and brightness levels of the scenes, I am grateful to have had the help of Michiel Hazelhof. Michiel programmed several scripts using the command line interface of ffmpeg and ImageJ to make the low-level measurements needed for comparing brightness and color of the videos possible in Lab color space.

I would also like to say thanks to the ESSC capoeira group for their advice and mental support during the course of my thesis.

Proof reading of manuscript, structure and dotting proverbial i's: Esther Smit, Daniel de Haas and Bridget Spoor. Esther was a great sparring partner and source of mental support during the course of report writing and preparing the presentation.

7 References

Aarts, H., Dijksterhuis, A., & Vries, P. (2001). On the psychology of drinking: Being thirsty and perceptually ready. *British Journal of Psychology*, *92*(4), 631–642. https://doi.org/10.1348/000712601162383

Adobe. (2004). *Photoshop* (CS) [Computer software]. Peachpit Press.

Adobe. (2017). *Premiere* (12.0) [Computer software]. Peachpit Press.

Alvarsson, J. J., Wiens, S., & Nilsson, M. E. (2010). Stress Recovery during Exposure to Nature Sound and Environmental Noise. *International Journal of Environmental Research and Public Health*, *7*(3), 1036–1046. https://doi.org/10.3390/ijerph7031036

Annerstedt, M., Jönsson, P., Wallergård, M., Johansson, G., Karlson, B., Grahn, P., Hansen, Å. M., & Währborg, P. (2013). Inducing physiological stress recovery with sounds of nature in a virtual reality forest—Results from a pilot study. *Physiology & Behavior*, *118*, 240–250. https://doi.org/10.1016/j.physbeh.2013.05.023

Ashbolt, N. J. (2004). Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology*, *198*(1–3), 229–238. https://doi.org/10.1016/j.tox.2004.01.030

Bailey, H. (2020). *OBS studio* (25.0.8) [Computer software]. OBS studio.

Barton, J., & Pretty, J. (2010). What is the Best Dose of Nature and Green Exercise for Improving Mental Health? A Multi-Study Analysis. *Environmental Science & Technology*, *44*(10), 3947–3955. https://doi.org/10.1021/es903183r

Basu, A., Duvall, J., & Kaplan, R. (2019). Attention Restoration Theory: Exploring the Role of Soft Fascination and Mental Bandwidth. *Environment and Behavior*, *51*(9–10), 1055–1081. https://doi.org/10.1177/0013916518774400

Bellard, F. (2020). *Ffmpeg* (2.8.16) [Computer software]. https://ffmpeg.org

Berman, M. G., Jonides, J., & Kaplan, S. (2008). The Cognitive Benefits of Interacting With Nature. *Psychological Science*, *19*(12), 1207–1212. https://doi.org/10.1111/j.1467-9280.2008.02225.x

Berto, R. (2005). Exposure to restorative environments helps restore attentional capacity. *Journal of Environmental Psychology*, *25*(3), 249–259. https://doi.org/10.1016/j.jenvp.2005.07.001

Berto, R. (2014). The Role of Nature in Coping with Psycho-Physiological Stress: A Literature Review on Restorativeness. *Behavioral Sciences*, *4*(4), 394–409. https://doi.org/10.3390/bs4040394

Beute, F., & de Kort, Y. A. W. (2014). Natural resistance: Exposure to nature and self-regulation, mood, and physiology after ego-depletion. *Journal of Environmental Psychology*, *40*, 167–178. https://doi.org/10.1016/j.jenvp.2014.06.004

Beute, F., & de Kort, Y. A. W. (2018). The natural context of wellbeing: Ecological momentary assessment of the influence of nature and daylight on affect and stress for individuals with depression levels varying from none to clinical. *Health & Place*, *49*, 7–18. https://doi.org/10.1016/j.healthplace.2017.11.005

Bovik, A. (2005). *Handbook of Image and Video Processing—Perceptual Criteria for Image Quality Evaluation*. Elsevier Academic Press.

Bowler, D. E., Buyung-Ali, L. M., Knight, T. M., & Pullin, A. S. (2010). A systematic review of evidence for the added benefits to health of exposure to natural environments. *BMC Public Health*, *10*(1), 456. https://doi.org/10.1186/1471-2458-10-456

Bratman, G. N., Daily, G. C., Levy, B. J., & Gross, J. J. (2015). The benefits of nature experience: Improved affect and cognition. *Landscape and Urban Planning*, *138*, 41–50. https://doi.org/10.1016/j.landurbplan.2015.02.005

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). *The timing mega-study: Comparing a range of experiment generators, both lab-based and online*. School of Psychology, University of Nottingham.

Byron, K., Khazanchi, S., & Nazarian, D. (2010). The relationship between stressors and creativity: A meta-analysis examining competing theoretical models. *Journal of Applied Psychology*, *95*(1), 201–212. https://doi.org/10.1037/a0017868

Cacioppo, J., Tassinary, L. G., & Berntson, G. G. (2007). *Cacioppo—Handbook of Psychophysiology.pdf* (Third edition). Cambridge University Press.

Carles, J., Barrio, I., & Lucio, J. de. (1999). Sound influence on landscape values. *Landscape and Urban Planning*, *43*, 191–200. https://doi.org/10.1080/01426399208706361

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Crocker, M. J. (2007). *Hearing Thresholds, Loudness of Sound, and Sound Adaptation*. Wiley Online Library.

Danielsson, N. S., Harvey, A. G., MacDonald, S., Jansson-Fröjmark, M., & Linton, S. J. (2013). Sleep Disturbance and Depressive Symptoms in Adolescence: The Role of Catastrophic Worry. *Journal of Youth and Adolescence*, *42*(8), 1223–1233. https://doi.org/10.1007/s10964-012-9811-6

de Kort, Y. A. W., Meijnders, A. L., Sponselee, A. A. G., & IJsselsteijn, W. A. (2006). What's wrong with virtual trees? Restoring from stress in a mediated environment. *Journal of Environmental Psychology*, *26*(4), 309–320. https://doi.org/10.1016/j.jenvp.2006.09.001

Deng, L., Luo, H., Ma, J., Huang, Z., Sun, L.-X., Jiang, M.-Y., Zhu, C.-Y., & Li, X. (2020). Effects of integration between visual stimuli and auditory stimuli on restorative potential and aesthetic preference in urban green spaces. *Urban Forestry & Urban Greening*, *53*, 1–13. https://doi.org/10.1016/j.ufug.2020.126702

Diette, G. B., Lechtzin, N., Haponik, E., Devrotes, A., & Rubin, H. R. (2003). Distraction Therapy With Nature Sights and Sounds Reduces Pain During Flexible Bronchoscopya. *Chest*, *123*(3), 941–948. https://doi.org/10.1378/chest.123.3.941

Dirac Research AB. (2021). *Dirac* (6.0) [Computer software]. Dirac Research AB.

Dopko, R. L., Capaldi, C. A., & Zelenski, J. M. (2019). The psychological and social benefits of a nature experience for children: A preliminary investigation. *Journal of Environmental Psychology*, *63*, 134–138. https://doi.org/10.1016/j.jenvp.2019.05.002

Ekman, P., Levenson, R., & Friesen, W. (1983). Autonomous nervous system activity distinguishes among emotions. *Science, New Series*, *221*(4616), 1208–1210.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the Efficiency and Independence of Attentional Networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2020). *GPower* (3.1.9.6) [Computer software]. Heinrich Heine Universität.

Ferguson, C. J. (2009). Is Psychological Research Really as Good as Medical Research? Effect Size Comparisons between Psychology and Medicine. *Review of General Psychology*, *13*(2), 130–136. https://doi.org/10.1037/a0015103

Fletcher, H., & Munson, W. A. (1933). Loudness, Its Definition, Measurement and Calculation. *Journal of Acoustics Society America*, *12*, 377–430. https://doi.org/10.1002/j.1538-7305.1933.tb00403.x

Foy, T., Dwyer, R. J., Nafarrete, R., Hammoud, M. S. S., & Rockett, P. (2019). Managing job performance, social support and work-life conflict to reduce workplace stress. *International Journal of Productivity and Performance Management*, *68*(6), 1018–1041. https://doi.org/10.1108/IJPPM-03-2017-0061

Gidlow, C. J., Jones, M. V., Hurst, G., Masterson, D., Clark-Carter, D., Tarvainen, M. P., Smith, G., & Nieuwenhuijsen, M. (2016). Where to put your best foot forward: Psycho-physiological responses to walking in natural and urban environments. *Journal of Environmental Psychology*, *45*, 22–29. https://doi.org/10.1016/j.jenvp.2015.11.003

Glaser, D., N., Tatum, B. C., Nebeker, D. M., Sorenson, R. C., & Aiello, J. R. (1999). Workload and Social Support: Effeects on Performance and Stress. *Human Performance*, *2*(2), 155–176.

Gould van Praag, C. D., Garfinkel, S. N., Sparasci, O., Mees, A., Philippides, A. O., Ware, M., Ottaviani, C., & Critchley, H. D. (2017). Mind-wandering and alterations to default mode network connectivity when listening to naturalistic versus artificial sounds. *Scientific Reports*, *7*(1), 45273. https://doi.org/10.1038/srep45273

Goutier, P. (2007). *Wavelab* (6.10 (build 340)) [Computer software]. Steinberg GmbH.

Grellier, J., White, M. P., Albin, M., Bell, S., Elliott, L. R., Gascón, M., Gualdi, S., Mancini, L., Nieuwenhuijsen, M. J., Sarigiannis, D. A., van den Bosch, M., Wolf, T., Wuijts, S., & Fleming, L. E. (2017). BlueHealth: A study programme protocol for mapping and quantifying the potential benefits to public health and well-being from Europe's blue spaces. *BMJ Open*, *7*(6), e016188. https://doi.org/10.1136/bmjopen-2017-016188

Gudi, A., Bittner, M., Lochmans, R., & van Gemert, J. (2019). Efficient Real-Time Camera Based Estimation of Heart Rate and Its Variability. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 1570–1579. https://doi.org/10.1109/ICCVW.2019.00196

Gunnarsson, B., Knez, I., Hedblom, M., & Sang, Å. O. (2017). Effects of biodiversity and environment-related attitude on perception of urban green space. *Urban Ecosystems*, *20*(1), 37–49. https://doi.org/10.1007/s11252-016-0581-x

Hartig, T., Evans, G. W., Jamner, L. D., Davis, D. S., & Gärling, T. (2003). Tracking restoration in natural and urban field settings. *Journal of Environmental Psychology*, *23*(2), 109–123. https://doi.org/10.1016/S0272-4944(02)00109-3

Hartig, T., Korpela, K., Evans, G. W., & Gärling, T. (1996). Validation of a Measure of Perceived Environmental Restorativeness. *Göteborg Psychological Reports*, *26*(7), 1–64.

Hartig, T., Korpela, K., Evans, G. W., & Gärling, T. (1997). A measure of restorative quality in environments. *Scandinavian Housing and Planning Research*, *14*(4), 175–194. https://doi.org/10.1080/02815739708730435

Hartig, T., Mang, M., & Evans, G. W. (1991). Restorative Effects of Natural Environment Experiences. *Environment and Behavior*, *23*(1), 3–26. https://doi.org/10.1177/0013916591231001

Hartig, T., Mitchell, R., de Vries, S., & Frumkin, H. (2014). Nature and Health. *Annual Review of Public Health*, *35*(1), 207–228. https://doi.org/10.1146/annurev-publhealth-032013-182443

Hockey, R. (1983). *Stress and fatigue in human performance* (1st ed.). Wiley.

Houtsma, A. J. M., Durlach, N. I., & Braida, L. D. (1980). Intensity perception XI. Experimental results on the relation of intensity resolution to loudness matching. *The Journal of the Acoustical Society of America*, *68*(3), 807–813. https://doi.org/10.1121/1.384819

Huai, S., & Van de Voorde, T. (2022). Which environmental features contribute to positive and negative perceptions of urban parks? A cross-cultural comparison using online reviews and Natural Language Processing methods. *Landscape and Urban Planning*, *218*, 104307. https://doi.org/10.1016/j.landurbplan.2021.104307

Hunter, M. D., Eickhoff, S. B., Pheasant, R. J., Douglas, M. J., Watts, G. R., Farrow, T. F. D., Hyland, D., Kang, J., Wilkinson, I. D., & Horoshenkov, K. V. (2010). The state of tranquility: Subjective perception is shaped by contextual modulation of auditory connectivity. *NeuroImage*, *53*(2), 611–618. https://doi.org/10.1016/j.neuroimage.2010.06.053

IBM. (2017). *IBM SPSS Statistics for Windows* (25.0) [Computer software]. IBM Corp.

Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive Affect Facilitates Creative Problem Solving. *Journal of Personality and Social Psychology*, *52*(6), 1122–1131.

Jahncke, H., Naula, S., & Eriksson, K. (2015). The effects of auditive and visual settings on perceived restoration likelihood. *Noise and Health*, *17*(74), 1. https://doi.org/10.4103/1463-1741.149559

James, W. (1892). The Stream of Consciousness. *Psychology: Briefer Course*, 179–202. https://doi.org/doi.org/10.1037/11630-011

Jiang, B., Chang, C.-Y., & Sullivan, W. C. (2014). A dose of nature: Tree cover, stress reduction, and gender differences. *Landscape and Urban Planning*, *132*, 26–36. https://doi.org/10.1016/j.landurbplan.2014.08.005

Joye, Y., & Bolderdijk, J. W. (2015). An exploratory study into the effects of extraordinary nature on emotions, mood, and prosociality. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01577

Joye, Y., & Dewitte, S. (2018). Nature's broken path to restoration. A critical look at Attention Restoration Theory. *Journal of Environmental Psychology*, *59*, 1–8. https://doi.org/10.1016/j.jenvp.2018.08.006

Joye, Y., Pals, R., Steg, L., & Evans, B. L. (2013). New Methods for Assessing the Fascinating Nature of Nature Experiences. *PLoS ONE*, *8*(7), e65332. https://doi.org/10.1371/journal.pone.0065332

Joye, Y., & van den Berg, A. (2011). Is love for green in our genes? A critical analysis of evolutionary assumptions in restorative environments research. *Urban Forestry & Urban Greening*, *10*(4), 261–268. https://doi.org/10.1016/j.ufug.2011.07.004

Kaplan, R., & Kaplan, S. (1989). *The Experience of Nature—A Psychological Perspective*. Cambridge University Press.

Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology*, *15*(3), 169–182. https://doi.org/10.1016/0272-4944(95)90001-2

Kerr, D., Stenson, R., Duba, J., & Dayan, N. (2014). *Midside Matrix* (3.5.1 r200102) [Computer software]. Goodhertz, Inc.

Kimpe, T., & Marchessoux, C. (2016). *4.5: Important Differences Between Medical Displays and Normal Desktop Displays and Underlying Reasons*.

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, *28*(1–2), 76–81. https://doi.org/10.1159/000119004

Kjellgren, A., & Buhrkall, H. (2010). A comparison of the restorative effect of a natural environment with that of a simulated natural environment. *Journal of Environmental Psychology*, *30*(4), 464–472. https://doi.org/10.1016/j.jenvp.2010.01.011

Krzywicka, P., & Byrka, K. (2017). Restorative Qualities of and Preference for Natural and Urban Soundscapes. *Frontiers in Psychology*, *8*, 1705. https://doi.org/10.3389/fpsyg.2017.01705

Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J., & Midden, C. J. H. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion*, *27*(7), 1225–1246. https://doi.org/10.1080/02699931.2013.781501

Larkin, K. T., & Chantler, P. D. (2020). Stress, depression, and cardiovascular disease. In *Cardiovascular Implications of Stress and Depression* (pp. 1–12). Elsevier. https://doi.org/10.1016/B978-0-12-815015-3.00001-5

Laumann, K., Gärling, T., & Stormark, K. M. (2003). Selective attention and heart rate responses to natural and urban environments. *Journal of Environmental Psychology*, *23*(2), 125–134. https://doi.org/10.1016/S0272-4944(02)00110-X

Leather, P., Pyrgas, M., Beale, D., & Lawrence, C. (1998). Windows in the workplace—Sunlight, view and occupational stress. *Environment and Behavior*, *30*(6), 739–762.

LeCroy, C. W., & Krysik, J. (2007). Understanding and Interpreting Effect Size Measures. *Social Work Research*, *31*(4), 243–248. https://doi.org/10.1093/swr/31.4.243

Lee, K. E., Williams, K. J. H., Sargent, L. D., Williams, N. S. G., & Johnson, K. A. (2015). 40-second green roof views sustain attention: The role of micro-breaks in attention restoration. *Journal of Environmental Psychology*, *42*, 182–189. https://doi.org/10.1016/j.jenvp.2015.04.003

Lin, H.-P., Lin, H.-Y., Lin, W.-L., & Huang, A. C.-W. (2011). Effects of stress, depression, and their interaction on heart rate, skin conductance, finger temperature, and respiratory rate: Sympathetic-parasympathetic hypothesis of stress and depression. *Journal of Clinical Psychology*, *67*(10), 1080–1091. https://doi.org/10.1002/jclp.20833

Lindquist, M., Lange, E., & Kang, J. (2016). From 3D landscape visualization to environmental simulation: The contribution of sound to the perception of virtual environments. *Landscape and Urban Planning*, *148*, 216–231. https://doi.org/10.1016/j.landurbplan.2015.12.017

Liu, Y., Hu, M., & Zhao, B. (2019). Audio-visual interactive evaluation of the forest landscape based on eye-tracking experiments. *Urban Forestry & Urban Greening*, *46*(126476), 1–11. https://doi.org/10.1016/j.ufug.2019.126476

Lombard, M., & Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2).

Luppa, M., Heinrich, S., Angermeyer, M. C., König, H.-H., & Riedel-Heller, S. G. (2007). Cost-of-illness studies of depression. *Journal of Affective Disorders*, *98*(1–2), 29–43. https://doi.org/10.1016/j.jad.2006.07.017

Ma, H., & Shu, S. (2018). An Experimental Study: The Restorative Effect of Soundscape Elements in a Simulated Open-Plan Office. *Acta Acustica United with Acustica*, *104*(1), 106–115. https://doi.org/10.3813/AAA.919150

Marci, C. D., Glick, D. M., Loh, R., & Dougherty, D. D. (2007). Autonomic and prefrontal cortex responses to autobiographical recall of emotions. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(3), 243–250. https://doi.org/10.3758/CABN.7.3.243

Martínez-Soto, J., Gonzales-Santos, L., Pasaye, E., & Barrios, F. A. (2013). Exploration of neural correlates of restorative environment exposure through functional magnetic resonance. *Intelligent Buildings International*, *5*(S1), 10–28. https://doi.org/10.1080/17508975.2013.807765

Mayer, F. S., & Frantz, C. M. (2004). The connectedness to nature scale: A measure of individuals' feeling in community with nature. *Journal of Environmental Psychology*, *24*(4), 503–515. https://doi.org/10.1016/j.jenvp.2004.10.001

McCraty, R., Atkinson, M., Tiller, A., Rein, G., & Watkins, A. D. (1995). The Effects of Emotions on Short-term Power Spectrum Analysis of Heart Rate Variability. *The American Journal of Cardiology*, *76*, 1089–1093.

McGonagle, K. A., & Kessler, R. C. (1990). Chronic stress, acute stress, and depressive symptoms. *American Journal of Community Psychology*, *18*(5), 681–706. https://doi.org/10.1007/BF00931237

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual Profile Of Mood States*. Education and Industrial Testing Service.

Menardo, E., Brondino, M., Hall, R., & Pasini, M. (2021). Restorativeness in Natural and Urban Environments: A Meta-Analysis. *Psychological Reports*, *124*(2), 417–437. https://doi.org/10.1177/0033294119884063

Mills, A. W. (1960). Lateralization of High-Frequency Tones. *The Journal of the Acoustical Society of America*, *32*(1), 132–134. https://doi.org/10.1121/1.1907864

Mokrzycki, W., & Tatol, M. (2011). *Mokrzycki 2011 DeltaE- Survey-2.pdf*. Machine Graphics and Vision. 20. 383-411.; Machine Graphics and Vision. 20. 383-411. https://www.researchgate.net/publication/236023905_Color_difference_Delta_E_-_A_survey

Moreno, A., Baker, S., Varey, K., & Hinze-Pifer, R. (2018). Bringing attention restoration theory to the classroom: A tablet app using nature videos to replenish effortful cognition. *Trends in Neuroscience and Education*, *12*, 7–21. https://doi.org/10.1016/j.tine.2018.07.002

Nickerson, D. (1936). The Specification of Color Tolerances. *Textile Research Journal*, *6*(12), 505–514. https://doi.org/10.1177/004051753600601202

Nikolic, J. (2020). *Youlean Loudness Meter 2* (2.4.2 build 344) [Computer software]. Youlean.

Nikolic, Y.-J. (2020, November 13). *How to Edit a Video to Achieve Good Audio Loudness on YouTube?* Youlean. https://youlean.co/how-to-edit-a-video-to-achieve-good-audio-loudness-on-youtube/

Nilsson, M. E., & Berglund, B. (2006). *Soundscape Quality in Suburban Green Areas and City Parks* (Vol. 92 (2006) 903-911). Acustica.

Ojala, A., Korpela, K., Tyrväinen, L., Tiittanen, P., & Lanki, T. (2019). Restorative effects of urban green environments and the role of urban-nature orientedness and noise sensitivity: A field experiment. *Health & Place*, *55*, 59–70. https://doi.org/10.1016/j.healthplace.2018.11.004

Pals, R., Steg, L., Siero, F. W., & van der Zee, K. I. (2009). Development of the PRCQ: A measure of perceived restorative characteristics of zoo attractions. *Journal of Environmental Psychology*, *29*(4), 441–449. https://doi.org/10.1016/j.jenvp.2009.08.005

Park, S. H., Lee, P. J., Jung, T., & Swenson, A. (2020). Effects of the aural and visual experience on psycho-physiological recovery in urban and rural environments. *Applied Acoustics*, *169*(107486), 1–14. https://doi.org/10.1016/j.apacoust.2020.107486

Parsons, R., Tassinary, L. G., Ulrich, R. S., Hebl, M. R., & Grossman-Alexander, M. (1998). The View From the Road: Implications for Stress Recovery and Immunization. *Journal of Environmental Psychology*, *18*(2), 113–140. https://doi.org/10.1006/jevp.1998.0086

Pasini, M., Berto, R., Brondino, M., Hall, R., & Ortner, C. (2014). How to Measure the Restorative Quality of Environments: The PRS-11. *Procedia - Social and Behavioral Sciences*, *159*, 293–297. https://doi.org/10.1016/j.sbspro.2014.12.375

Petersen, S. E., & Posner, M. I. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience*, *35*(1), 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525

Pierce, J. R. (1983). *The Science of Musical Sound*. Scientific American Press.

Pigeon, D. I. S. (2021, November 11). *Online Hearing Test & Audiogram Printout*. https://hearingtest.online/

Preis, A., Kociński, J., Hafke-Dys, H., & Wrzosek, M. (2015). Audio-visual interactions in environment assessment. *Science of The Total Environment*, *523*, 191–200. https://doi.org/10.1016/j.scitotenv.2015.03.128

Publications Office of the European Union. (2014). *Psychosocial risks in Europe—Prevalence and strategies for prevention*. Publications Office. https://data.europa.eu/doi/10.2806/70971

Rasband, W., S. (2019). *ImageJ* (1.52n) [Computer software]. U.S. National Institutes of Health. http://imagej.nih.gov/ij

Ratcliffe, E., Gatersleben, B., & Sowden, P. T. (2013). Bird sounds and their contributions to perceived attention restoration and stress recovery. *Journal of Environmental Psychology*, *36*, 221–228. https://doi.org/10.1016/j.jenvp.2013.08.004

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.

Ritchie, H., & Roser, M. (2018). Urbanization. *Our World in Data*. https://ourworldindata.org/urbanization

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). `Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-3932(97)00015-8

Rummukainen, O. S., Robotham, T., & Habets, E. A. P. (2021). Head-Related Transfer Functions for Dynamic Listeners in Virtual Reality. *Applied Sciences*, *11*(14), 1–16. https://doi.org/10.3390/app11146646

Salvagioni, D. A. J., Melanda, F. N., Mesas, A. E., González, A. D., Gabani, F. L., & Andrade, S. M. de. (2017). Physical, psychological and occupational consequences of job burnout: A systematic review of prospective studies. *PLOS ONE*, *12*(10), e0185781. https://doi.org/10.1371/journal.pone.0185781

Schlittenlacher, J., Hashimoto, T., Kuwano, S., & Namba, S. (2017). Overall judgment of loudness of time-varying sounds. *The Journal of the Acoustical Society of America*, *142*(4), 1841–1847. https://doi.org/10.1121/1.5003797

Schmitz, C. (2020). *LimeSurvey* (3.22.15+200505) [Computer software]. LimeSurvey GmbH. http://www.limesurvey.org

Shanahan, D. F., Bush, R., Gaston, K. J., Lin, B. B., Dean, J., Barber, E., & Fuller, R. A. (2016). Health Benefits from Nature Experiences Depend on Dose. *Scientific Reports*, *6*(28551), 1–10. https://doi.org/10.1038/srep28551

Shu, S., & Ma, H. (2019). Restorative Effects of Classroom Soundscapes on Children's Cognitive Performance. *International Journal of Environmental Research and Public Health*, *16*(2), 293. https://doi.org/10.3390/ijerph16020293

Sona, B., Dietl, E., & Steidle, A. (2019). Recovery in sensory-enriched break environments: Integrating vision, sound and scent into simulated indoor and outdoor environments. *Ergonomics*, *62*(4), 521–536. https://doi.org/10.1080/00140139.2018.1491643

Sony. (2012). *Vegas Pro* (Version 12) [Computer software]. Sony Creative Software.

Spielberger, C. D. (1983). *State-Trait Anxiety Inventory for Adults (STAI-AD)* [Data set]. American Psychological Association. https://doi.org/10.1037/t06496-000

Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1228–1248. https://doi.org/10.1016/j.neubiorev.2012.02.003

Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—Or Why the *P* Value Is Not Enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Sun, M., Herrup, K., Shi, B., Hamano, Y., Liu, C., & Goto, S. (2018). Changes in visual interaction: Viewing a Japanese garden directly, through glass or as a projected image. *Journal of Environmental Psychology*, *60*, 116–121. https://doi.org/10.1016/j.jenvp.2018.10.009

Taylor, L., & Hochuli, D. F. (2017). Defining greenspace: Multiple uses across multiple disciplines. *Landscape and Urban Planning*, *158*, 25–38. https://doi.org/10.1016/j.landurbplan.2016.09.024

Tennessen, C. M., & Cimprich, B. (1995). Views to nature: Effects on attention. *Journal of Environmental Psychology*, *15*(1), 77–85. https://doi.org/10.1016/0272-4944(95)90016-0

TNO. (2020). *Factsheet Week van de werkstress 2020*. TNO Netherlands. https://www.monitorarbeid.tno.nl/nl-nl/publicaties/factsheet-week-van-de-werkstress-2020/

Twedt, E., Rainey, R. M., & Proffitt, D. R. (2019). Beyond nature: The roles of visual appeal and individual differences in perceived restorative potential. *Journal of Environmental Psychology*, *65*(101322), 1–46. https://doi.org/10.1016/j.jenvp.2019.101322

Ulrich, R. S. (1979). Visual landscapes and psychological well-being. *Landscape Research*, *4*(1), 17–23. https://doi.org/10.1080/01426397908705892

Ulrich, R. S. (1981). Natural Versus Urban Scenes: Some Psychophysiological Effects. *Environment and Behavior*, *13*(5), 523–556.

Ulrich, R. S. (1983). Aesthetic and Affective Response to Natural Environment. In I. Altman & J. F. Wohlwill (Eds.), *Behavior and the Natural Environment* (pp. 85–125). Springer US. https://doi.org/10.1007/978-1-4613-3539-9_4

Ulrich, R. S. (1984). View Through a Window May Influence Recovery from Surgery. *Science*, *224*(4647), 420–421. https://doi.org/10.1126/science.6143402

Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., & Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, *11*(3), 201–230. https://doi.org/10.1016/S0272-4944(05)80184-7

Valtchanov, D., Barton, K. R., & Ellard, C. (2010). Restorative Effects of Virtual Nature Settings. *Cyberpsychology, Behavior, and Social Networking*, *13*(5), 503–512. https://doi.org/10.1089/cyber.2009.0308

van der Molen, H., Kuijer, P., de Groene, G., Sorgdrager, B., Lenderink, A., Maas, J., & Brand, T. (2020). *Beroepsziekten in Cijfers 2020* (ISBN 978 94 91043 21 5; p. 73). Nederlands Centrum voor Beroepsziekten. https://www.beroepsziekten.nl/sites/default/files/documents/beroepsziekten-in-cijfers-2020.pdf

Van Staden, D., Noor Mahomed, F., Govender, S., Lengisi, L., Singh, B., & Aboobaker, O. (2018). Comparing the validity of an online Ishihara colour vision test to the traditional Ishihara handbook in a South African university population. *African Vision and Eye Health*, *77*(1). https://doi.org/10.4102/aveh.v77i1.370

VideoLAN-team. (2016). *VLC media player* (2.2.4 Weatherwax) [Computer software]. VideoLAN.

Wang, R., & Zhao, J. (2019). A good sound in the right place: Exploring the effects of auditory-visual combinations on aesthetic preference. *Urban Forestry & Urban Greening*, *43*(126356), 1–8. https://doi.org/10.1016/j.ufug.2019.05.018

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Watson, D., & Pennebaker, J. W. (1989). Health Complaints, Stress, and Distress: Exploring the Central Role of Negative Affectivity. *Psychological Review*, *96*(2), 234–254.

Watts, G. R., Pheasant, R. J., Horoshenkov, K. V., & Ragonesi, L. (2009). Measurement and Subjective Assessment of Water Generated Sounds. *Acta Acustica United with Acustica*, *95*(6), 1032–1039. https://doi.org/10.3813/AAA.918235

WhatIsMyBrowser.com. (2021, January 10). *WhatIsMyBrowser.com*. WhatIsMyBrowser.Com. https://developers.whatismybrowser.com/useragents/parse/

White, M., Smith, A., Humphryes, K., Pahl, S., Snelling, D., & Depledge, M. (2010). Blue space: The importance of water for preference, affect, and restorativeness ratings of natural and built scenes. *Journal of Environmental Psychology*, *30*(4), 482–493. https://doi.org/10.1016/j.jenvp.2010.04.004

Wilkie, S., & Stavridou, A. (2013). Influence of environmental preference and environment type congruence on judgments of restoration potential. *Urban Forestry & Urban Greening*, *12*(2), 163–170. https://doi.org/10.1016/j.ufug.2013.01.004

Williams, K. J. H., Lee, K. E., Hartig, T., Sargent, L. D., Williams, N. S. G., & Johnson, K. A. (2018). Conceptualising creativity benefits of nature experience: Attention restoration and mind wandering as complementary processes. *Journal of Environmental Psychology*, *59*, 36–45. https://doi.org/10.1016/j.jenvp.2018.08.005

Wilson, E. (1984). *Biophilia* (1st ed.). Harvard University Press.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, *7*(240), 225–240.

Wolkowitz, O. M., Epel, E. S., Reus, V. I., & Mellon, S. H. (2010). Depression gets old fast: Do stress and depression accelerate cell aging? *Depression and Anxiety*, *27*(4), 327–338. https://doi.org/10.1002/da.20686

Wooller, J. J., Rogerson, M., Barton, J., Micklewright, D., & Gladwell, V. (2018). Can Simulated Green Exercise Improve Recovery From Acute Mental Stress? *Frontiers in Psychology*, *9*(2167), 1–10. https://doi.org/10.3389/fpsyg.2018.02167

Wooller, J.-J., Barton, J., Gladwell, V. F., & Micklewright, D. (2016). Occlusion of sight, sound and smell during Green Exercise influences mood, perceived exertion and heart rate. *International Journal of Environmental Health Research*, *26*(3), 267–280. https://doi.org/10.1080/09603123.2015.1109068

Wooster, E. (2008). *While we are waiting: Experiences of waiting for and receiving psychological therapies on the NHS* (pp. 1–16). www.weneedtotalk.org.uk

Zajonc, R. B. (1980). Feeling and Thinking: Preferences Need No Inferences. *American Psychologist*, *35*(2), 151–175.

Zhang, S., Zhao, X., Zeng, Z., & Qiu, X. (2019). The Influence of Audio-Visual Interactions on Psychological Responses of Young People in Urban Green Areas: A Case Study in Two Parks in China. *International Journal of Environmental Research and Public Health*, *16*(1845), 1–14. https://doi.org/10.3390/ijerph16101845

Zhao, J., Xu, W., & Ye, L. (2018). Effects of auditory-visual combinations on perceived restorative potential of urban green space. *Applied Acoustics*, *141*, 169–177. https://doi.org/10.1016/j.apacoust.2018.07.001

Zhao, N., Azaria, A., & Paradiso, J. A. (2017). Mediated Atmospheres: A Multimodal Mediated Work Environment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(2), 1–23. https://doi.org/10.1145/3090096

Zuckerman, M. (1977). Development of a Situation-Specific Trait-State Test for the Prediction and Measurement of Affective Responses. *Journal of Consulting and Clinical Psychology*, *45*(4), 513–523.

8 Appendices

8.1 Original design of study

Originally, this study was designed as a one hour lasting lab study in which a participant was to be stressed and brought back to relaxation with a baseline, and once again stressed and relaxed with a movie that lasted around 10 minutes. The independent variable was the modality the de-stressing movie was presented in (audio only, video only, audiovisually). Heart rate and GSR were to be measured as dependent variables. The hypothesis was that stress levels more rapidly decrease with a mediated natural environment (A or V) then in a control condition, where AV shows an additive restorative effect.

Then the COVID-19 virus struck the Netherlands late February 2020 and all lab-studies at Eindhoven Technical University (TU/e) were cancelled. The movie that was purposely made for this research, edited and nearly finished was no longer usable, since the lab-study that was aimed for was no longer possible. The originally intended movie contained scene changes, which are detrimental for the subjective measure of presence when doing online research. Neither was research allowed in which equipment was sent to a participant using a package or postal service; another approach was necessary. This means that the original research which included objective measures like heart rate and skin conductance measurement was no longer viable and the research design needed to be changed accordingly. The experiment was changed from a lab environment to the home environment of participant, utilizing their own (non-standardized) equipment. Authorware Lab could not be used for running the experiment, since running Authorware on the computer of a participant would mean that Authorware Web Player software needed to be installed.

The change in research environment also meant a change in research paradigm. Online research was decided on quickly, where participants can be stressed and de-stressed in their home-environment for a much shorter time-period than in a lab setting. A short heart rate measurement experiment with a photophletysmography (PPG) based phone app failed to deliver usable results. This short experiment revealed that measuring any objective measure reliably is next to impossible; little control over extraneous variables and thus biases were to be expected. In this research, the only objective measure is the response time in the SART test. From the response time measurement, it can be deduced how long a participant can sustain attention, or the other way around; when participants have a lapse of attention, as well as whether or not participants were doing the SART task seriously instead of rhythmically pressing the space bar.

8.2 Video recordings

Reflecting on the original Kaplan (1995) criteria, an inherently or softly fascinating natural environment should contain enough stimuli to cater for optimal stimulation. A mediated natural environment should contain enough potentially attention grabbing details that can be easily filtered out if they are overstimulating, but very pleasant in giving an engrossing experience. This also couples with the compatibility and complexity criteria Kaplan (1995) states.

The video that was needed for this research should only contain natural elements, no human presence, no pets or domestic animals, no threatening situations, and should not contain no man-made features (Kaplan, 1995). Visual cuts are thus not permissible in the video material, so it is to be in one continuous shot. It should be appealing to watch and inherently fascinating, but then again not attention grabbing to prevent the induction of hard fascination. Therefore, the criteria for selecting video material for this research was that it should be softly fascinating. Waving grass (Gunnarsson et al., 2017), moving clouds and water are preferred (White et al., 2010), but a herd of grazing cows deemed inappropriate.

After reviewing pre-recorded material on several video deposits, it appeared that the sound in the videos was almost never congruent with the images; many nature videos contain music or a voice over to make them more interesting to watch and the video feed usually switches within a minute. Getting stock material was therefore impractical, and it was decided to record material within the HTI group.

Video was recorded on several protected natural reserve locations on the Strabrechtse Heide. Permits to enter the area were requested and granted. A Nikon D3100 14,2 megapixel DSLR camera, capable of 1920x1080 (Full HD resolution) @24fps video capture was used for recording the videos. Gradual grey square filters were used to prevent overexposure. Several lenses were used to capture the image as naturally as possible. The internal audio recording function (mono electret microphone and mono audio input) of the camera was not used, a separate four channel recording using a high quality recorder was made and processed separately. This will be explained in high level in section 2.3.1, and in more detail in the Audio recordings section 8.6.

Video recordings were mostly made at sunrise and sunset, i.e. on the most quiet parts of the day. Animals were present, but not overactive. Explorative tests with a waterproof housing, a slider and time-lapse photography were also undertaken. The videos coming from these tests were not used for this research, for reasons of videos being unnatural in their timing, too salient, too interesting and thus more leaning towards hard fascination. The video that was originally cut and edited lasted 9 minutes and 19 seconds. A second video edit was made, which lasted 8 minutes and 50 seconds. The material overlapped for some scenes. Because of the COVID-19 outbreak and accompanying paradigm shift in which shorter videos were to be used for the online instead of lab research, an analysis was made for consecutive scenes lasting longer than one minute. six scenes (named scene A to scene F — displayed in Figure 14) emerged from the material, with different restoration-wise qualities. Scene B, recorded at water level with the camera in a water housing, was removed from the selection since the video contained no solid ground where the participant would be 'taking a place' in the scene. This negatively affected the ecological validity of the scene, as the video recording was not made at eye level (around 1.55 meters from ground level) (Lindquist et al., 2016; Han, 2017; Zhao et al. (2018)). Scene E contained cows eating grass and was removed from the selection because the cows present in this scene were grabbing too much attention, possibly leading to undesired hard fascination. At this moment, it became clear that more academic work could be done when implementing an additional factor instead of doing an audio/visual/audiovisual manipulation only, and the water level factor was introduced. Scenes were selected that would contain several levels of water presence. From the six 1-minute scenes, four scenes remained that were expanded on with the addition of audio.

A pre-pilot test was used for rating the scenes for auditory, visual and audiovisual qualities and to make a choice between scene A and scene F for the $\text{Water}_{\text{High}}$ condition. See section 2.3.6 for more information. Finally, scene C, D and F were selected for use in the study.

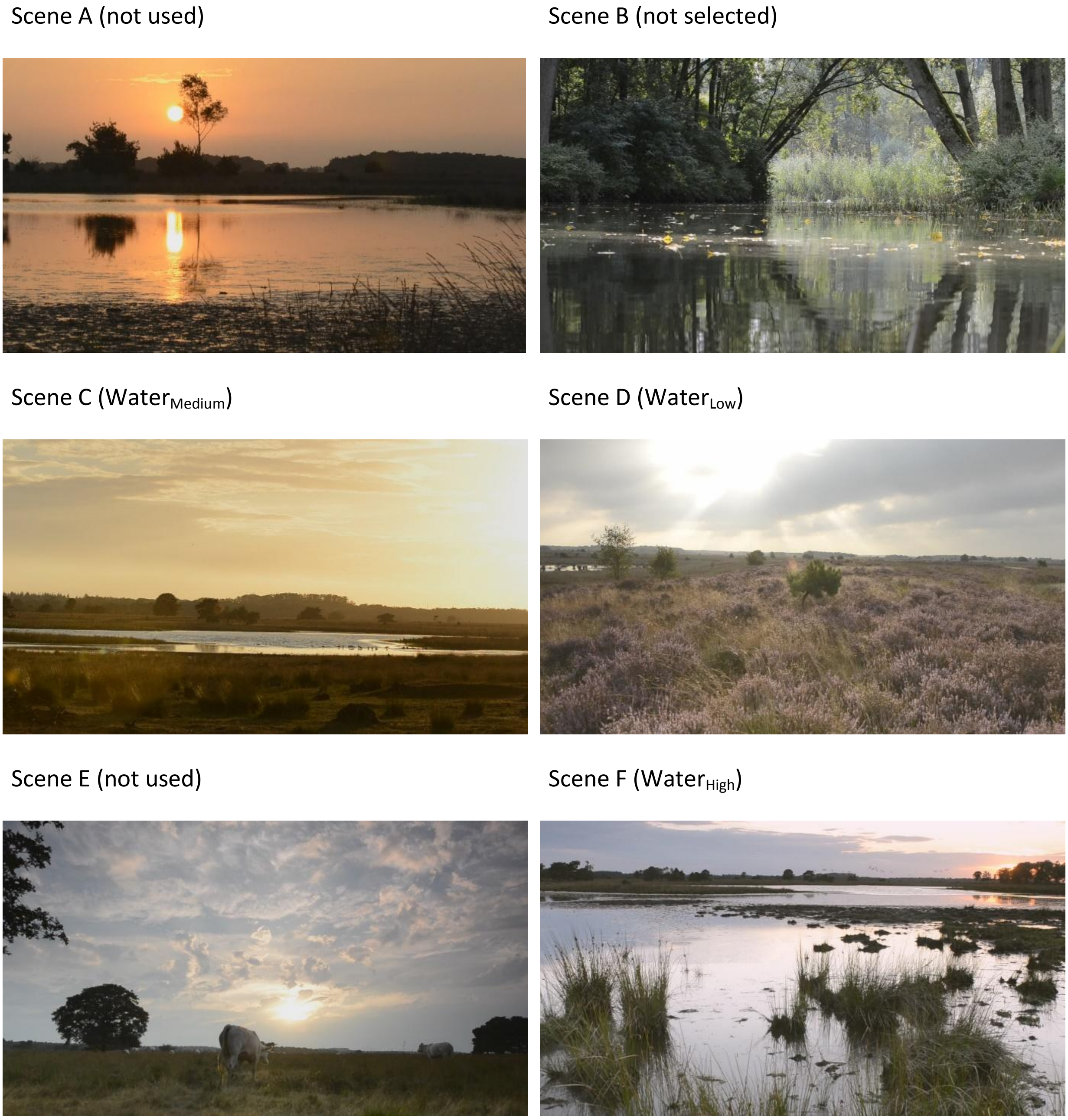

Figure 14: Screen shots of available 1-minute video scenes

Video recordings were luminance matched to prevent brightness bias that could influence affective responses (Lakens et al., 2013). Color matching was also attempted, but a full correction on the scenes was not possible due to un-natural hues of colors for some parts of the videos. For more information on the process of luminance and color matching of the videos, see Appendix 8.3 and 8.5 respectively.

8.3 Video image lightness matching

To prevent biases, all videos are to be as similar as possible, since differences in brightness are known to influence the affective evaluation of pictures (Lakens et al., 2013). The perception of brightness and the consecutive affective judgment of content material in a video depends on several factors. Brightness values are based on characteristics of the stimuli, and do not take additional sources of brightness variations into account, like differences in the brightness settings of computer monitors or environmental lighting conditions (Lakens et al., 2013). The human eye's sensitivity to luminance variations depends on a number of factors, including light level, spatial frequency, and signal content (Bovik, 2005). The quality of the computer screen with respect to brightness, color reproduction and contrast (Kimpe & Marchessoux, 2016). The former are all important metrics that influence affective valuation of the video material on the screen. Not all factors can be corrected because of the online nature of the experiment, but the video image lightness is to be matched.

The videos that were used as stimulus materials were filmed on the same day, but on different locations and with different positions of the sun. During the day the lighting was variable, and so the amount of light on the camera changed. The sun is displayed in different heights above the horizon in the scenes. The $Water_{Low}$ scene was recorded with a higher sun, where the $Water_{High}$ and $Water_{Medium}$ scene were recorded with a lower sun and thus less light. Since the $Water_{Low}$ scene was recorded with a higher sun, several gradual grey square filters were used to prevent saturation of the camera.

Concluding from preliminary measurements, a correction for lightness was necessary. For the preliminary measurements, a snapshot of the mostly stationary video scenes was taken and saved. These saved images were converted to L*a*b* color space using ImageJ freeware (Rasband, 2019). The CIE L*a*b* color space was used since this is a monitor independent color space and has perceptually linear characteristics (Mokrzycki & Tatol, 2011). The L*a*b* color space allows a wider color gamut than the RGB standard used for computer displays and even printed media. The distance between the points defining individual colors is proportional to the perceptual difference between them. The L*a*b* color space description is displayed visually in Figure 15 below. The L component of the L*a*b* color space is defined as Lightness, going from L^*=0 as the darkest black to L^*=100 as the brightest white. The a* and b* components represent green-red and blue-yellow saturation respectively.

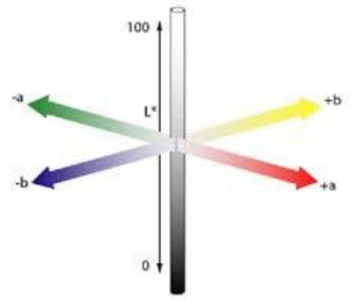

Figure 15: The L*a*b* color space visualized

Calculating the distance between colors is now possible with the formula as described in Equation 3.

$$\Delta E_{Lab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}$$

Equation 3: Calculation of difference in color

As discussed before, no control measures for most of the mentioned lightness perception mechanisms are available because of the online nature of this research. To prevent biases introduced by the stimulus material and thus make more accurate measurements, the mean Lightness of the scenes was adjusted by measuring the lightness of all frames of the scenes and averaging them. The video scene was transformed to 1440 (60 seconds of video material * 24 frames per second) images using ffmpeg (Bellard, 2020). This transformation made it possible to go from Planar 4:2:0 YUV RGB to individual images. After that, the images were loaded piecemealed into ImageJ (Rasband, 2019) to convert from RGB color space to Lab Stack (Image > Type > Lab Stack). The *L** component of the image was then displayed as greyscale. After conversion of all images, an Analyze > Measure operation was done in ImageJ (Rasband, 2019) for the *L** component of L*a*b*. Finally, a measurement was done giving the lightness (*L**) mean and SD for the 1440 images. These values were averaged and are displayed in Table 19.

Average Lightness (*L**)	$Water_{Low}$	$Water_{Medium}$	$Water_{High}$
Mean	59.823	59.887	59.443
SD	24.039	30.947	27.576

Table 19: Lightness (L) mean and SD per used scene

The just noticeable difference (JND) between 2 pictures is 2.3 JND (Mokrzycki & Tatol, 2011). The biggest difference in *L** after adjustment is 59.887 – 59.443 = 0.444, which is well within the JND of 2.3.

Differences between computer monitors are to be expected, since the current thesis employs a between subjects design. Please note that the CIE L*a*b* color space is equipment (and thus also monitor) independent and perceptually linear, and thus lightness is comparable over the different scenes of the stimulus material. When introducing different computer monitors, lightness is not comparable over participants since the color space is no longer calibrated, although the digital scene material is equally light.

8.4 Video scene characterization

This section describes various ways to characterize the video footage with respect to color similarity and brightness using standard measures. The lightness in CIE Lab color space is also described. After inspection, the brightness appeared to be different for the scenes, warranting modification of the brightness and color balance of the stimulus material to make all stimuli as similar as possible. An analysis of the final stimulus materials is discussed in section 8.5.

Measurement of color similarity of the pre-brightness-corrected pictures with Adobe Photoshop (Adobe, 2004).

Table 20: Luminosity measurement per scene — Adobe Photoshop

Scene	R	G	B	Luminosity
C - $Water_{Medium}$	162.79 / 74.13	144.12 / 73.86	144.12 / 73.86	145.52 / 73.26
D - $Water_{Low}$	128.11 / 54.37	116.86 / 59.13	103.95 / 62.74	118.83 / 57.94
F - $Water_{High}$	137.69 / 66.16	131.31 / 62.49	126.20 / 69.14	132.67 / 64.16

* Data is displayed as "Mean / SD"

Measurement of RGB color space of the pre-brightness-corrected pictures with ImageJ (Rasband, 2019)

Table 21: RGB measurement per scene — ImageJ

Scene	R	G	B	(R+G+B)/3	0.299R+0.587G+0.114B
C	162.788 / 74.132	144.116 / 73.858	103.853 / 71.533	136.917 / 72.748	145.109 / 73.439
D	127.875 / 54.905	116.628 / 59.621	103.714 / 63.201	116.056 / 58.961	118.537 / 58.476
F	137.449 / 66.987	131.065 / 63.346	125.951 / 69.930	131.474 / 66.450	132.399 / 65.018

* Data is displayed as "Mean / SD"

The CIE L*a*b* color space was used since it is a device-independent model. The images of the used scenes were converted to CIE L*a*b* (with D65 as white point) using ImageJ (Rasband, 2019) and then analyzed for Lightness.

Table 22: Lightness measurement per scene — ffmpeg and ImageJ

Lightness in L*a*b* color space per scene	**Mean**	**StdDev**	**Min**	**Max**
C	128.747	18.480	91	153
D	117.372	13.611	87	146
F	118.782	16.397	88	159

Scene C is slightly brighter. The minimal noticeable difference is a factor 2.3.

Using Equation 3 the ΔE_{Lab} of each scene was calculated. With Equation 4 the JND between the scenes was rated (Nickerson, 1936).

8.5 Video image color matching

A final variable that could influence the affective perception of the scene is the presence of colors in the stimulus material. During the experiment participants were requested to turn software like f.lux, SunsetScreen or Iris off to prevent biases due to warmer colors or brightness. Lakens (2013) discusses the issue of brightness extensively, with more information on this metric in appendix 8.3. Color and presumably also color temperature can bias affective responses (Zhao et al., 2017). Therefore, the brightness and color temperature was to be kept similar over scenes.

First, a measurement of the colors in RGB color space was made with ImageJ (Rasband, 2019) for a screenshot of the different scenes. The mean values per color are displayed in Table 23. A measurement over all frames was also made. The procedure as described in chapter 8.3 was followed, using ffmpeg (Bellard, 2020) to make separate images from the scene videos and measure the individual RGB components. The weights of the RGB components averaged over all video frames present in the scene videos are displayed in Table 24 below.

	$Water_{Low}$			$Water_{Medium}$			$Water_{High}$		
	R	G	B	R	G	B	R	G	B
Mean	149.806	139.533	127.548	153.847	135.040	95.417	144.613	138.171	133.149
SD	55.050	59.465	62.816	73.994	73.928	71.327	66.905	63.354	69.931
Min	45	37	18	21	12	0	32	27	2
Max	225	255	255	251	223	230	255	246	248

Table 23: RGB values per scene

$Water_{Low}$		$Water_{Medium}$		$Water_{High}$	
Rmean	155,917	Rmean	157,756	Rmean	152,084
Gmean	145,569	Gmean	152,424	Gmean	146,006
Bmean	133,981	Bmean	116,462	Bmean	140,764
Total	145,156	Total	142,214	Total	146,285

Table 24: RGB values and resulting color of used scenes after correction for lightness and color differences

The colors are different over scenes, but are difficult to change so the scenes are perceptually similar. In Vegas 12 (Sony, 2012) a LAB Adjust color editing function is present that can change colors in a video file with a slider for Lightness and one for the a and b channel. The Lightness can be adjusted between 0 and 2.000 (1.000 represents lightness original to the source material), where the a and b channel can be adjusted between -1 and +1. A problem here is that ImageJ (Rasband, 2019) displays values between +-100 or -128 to +127 (CIE L*a*b* officially specifies -128 to +127), i.e. the scaling is different. Moreover, Vegas 12 (Sony, 2012) does not implement measurement of the a and b channels of a video but only gives sliders to adjust a video (or two videos side by side to make them perceptually similar). It is therefore very difficult to make videos that are within JND with respect to color, since Vegas 12 (Sony, 2012) and other software packages are mostly designed for artistic and/or creative video editing, not for low level scientific editing of video material. Adjusting the colors on a video is an iterative process using this type of software. Another problem with editing colors is that grass does not look like grass anymore, the color of the water can even appear different or even unnatural which is to be avoided. So, care should be taken to edit only the source material, and not sequentially editing the processed material since the source material is 8 only bits per color. Several sequential adjustments of the colors can result in substantial quantization errors which can be visible as Mach bands.

$$\Delta E^*_{ab} \approx 2.3\, JND$$

Equation 4: JND calculation

8.6 Audio recordings and processing

Audio was recorded asynchronously, but at the same time as the videos using a Zoom H4n audio recorder. The Zoom H4n was set up in 4CH mode with Mid-Side matrixing turned on. Recording microphones were mounted on top of — and close to — each other on a mic boom in a perpendicular fashion, connected with short XLR cables and fed with 48 volt phantom power from the H4n. These microphones will from hereon be mentioned as the mid-side microphones, or the mid-side audio. The H4n internal XY microphones were also recorded and processed to find the most natural surround sound experience on headphones in the end result. Audio processing on the mid-side audio was done using Steinberg Wavelab 6 (Goutier, 2007) and encompassed extensive use of the audio montage functions, meticulously editing out all non-natural sounds (i.e. mic boom noises, shutter sounds from the camera, car horns, dogs, talking and airplane noises). A sound canvas of several minutes length was created with bird sounds, waving grass and other natural sounds, which could be used in all videos to keep conditions as similar as possible and prevent biases. The experimental conditions are of a between subjects nature, so participants will not hear the other sound conditions. The nature scenes were edited in the same way using the soundscape described earlier and extra sounds that accompany and support events that are present on the screen. The different water sounds were cut and selected from all recordings made (over one hour in total) and cleaned up according to the method described above. From here, the most congruent water sounds were coupled to the scenes, according to the different levels of water presence. In the $\text{Water}_{\text{High}}$ scene, in which water fills the largest part of the screen in comparison to the other scenes, the water sounds were turned up in volume. To add to the realism of water sounds in the $\text{Water}_{\text{Medium}}$ scene, the volume of water sounds was lower, the water sounds were more mono to decrease directionality of the water sounds, and the water sounds were high pass filtered to mimic the damping of high sound frequencies over distance. The numeric characteristics are described in Table 1. The $\text{Water}_{\text{Low}}$ scene did not contain any water sounds. When making the audio montages for the different scenes, the water sounds appeared to come from an acoustically different environment, which was clearly audible when using headphones. Therefore, when going from one piece of a recording to another, sufficient and slowly changing overlap was present so participants would not notice going from one acoustic environment to another. For the four pre-selected one-minute videos, four audio montages for the respective scenes were made that met the criteria for soft fascination and restorative potential. The audio files were then rendered, leaving them for further master processing.

Master processing was purposely done after making a rendered montage to keep all conditions similar among the different scenes. The master editing process was done in this particular order to keep phase relationships between the mid and side channel accurate, also keeping as much bits of resolution available in the recordings as possible and preventing sample rate conversion. The recording was in 48 kHz, WaveLab 6 (Goutier, 2007) was set to 48 kHz and YouTube (Google LLC, 2021) uses AAC-LC compression with 48 kHz sampling rate (Google LLC, 2021) First, the low frequency wind noises were high pass filtered with a shallow roll off filter in Wavelab 6 (Goutier, 2007) to remove low frequency rumble. The produced audio file was saved and loaded into Adobe Premiere 12 (Adobe, 2017) to process the mid-side audio montage to spatial stereo.

For mid-side matrixing the freeware VST GoodHertz Mid-Side matrix processor plugin (Kerr et al., 2014) was used. The gain of the side channels was increased by 3dB to get a natural surround effect giving accurate spatiality when listening to the audio with headphones without grabbing too much attention of the listener. Care was taken not to overdo the spatiality of the audio file, since only a realistic surround experience was needed. When a wow-effect like in the movie Avatar is introduced, participants undergoing the manipulation might be drawn to the spatial effects, instead of calmly listening to the manipulation.

After mid-side processing, the file was saved and returned to Wavelab 6 (Goutier, 2007) to boost the level of the recording with 24dB since the remaining level in the quietest passages was very soft after all loud and windy passages were edited out. Large amounts of headroom were needed to prevent signal clipping and so accommodate for wind noise when recording in the field. A multiple of 6dB was chosen, since amplifying a musical signal +6dB is a simple bit shift operation which does not introduce further quantization errors. The source material was 24 bits, so a 20 bit signal with enough resolution was still present after increasing the gain.

After all audio scenes were edited, processed and amplified, small differences in sound level between the scenes were present. To prevent a loudness bias the level of the audio material was to be matched within the just noticeable difference of 0.25 dB, following (Pierce, 1983). For a more detailed description of the JND properties and loudness matching, please see section 8.7. Attempts to equalize the dBFS loudness level with the Wavelab 6 level meter (Goutier, 2007) and matching the level of the scenes with Dirac 6 (Dirac Research AB, 2021) using the L_{Aeq} parameter were unsuccessful: loudness was not perceived as equal over scenes. Finally, loudness measurement was done with Youlean Loudness Meter 2 (J. Nikolic, 2020) in LUFS (Loudness Units relative to Full Scale). Loudness measurement using LUFS was chosen since LUFS are a perceptually accurate level measurement. Because YouTube (Google LLC, 2021) uses lossy AAC-LC compression that can alter loudness, the level matching was checked by uploading the sound to YouTube (Google LLC, 2021), then re-downloading it and measuring the level of the scene with the Youlean software (J. Nikolic, 2020). A level matching between scenes within 0.1 LUFS was achieved. After checking the loudness levels with headphones on a high quality audio system, no differences in level were perceived. Equipment used to monitor the audio signals during audio editing and master processing included a specialized Metrum Acoustics Non Oversampling DA-converter and Beyer Dynamic DT1350 headphones.

For the participants, headphones were used. Using headphones allows for greater control over the spatiality of the listening experience due to a larger channel separation. Also, reverberation of a room is not present when using headphones. An added side effect is that distracting sounds from the environment of the participant could be filtered out to a large degree if ear buds or closed back headphones were used. For open back type headphones, this effect is obviously not present.

8.7 Stimulus audio loudness matching

To control for the same audio loudness levels over the scenes, measurement and matching of the sounds levels was done with the level meter tool in Wavelab 6 (Goutier, 2007). This gave a peak and an average electrical loudness level for each stimulus. When listening to the levels, these were not perceptually the same. An 'electric' loudness level does not translate directly to an acoustic loudness level since there are more devices in the chain and factors in the sound information that influence the perceived loudness. Loudness is subdivided in electrical and acoustical loudness.

Electrical loudness is given with respect to 0dB FS, where FS stands for Full Scale. 0dB FS is the maximum loudness that can be encoded in PCM digital sound. A negative dB value then indicates that a sound is encoded with an amplitude proportional to that negative number.

The perception of acoustical loudness was documented by Fletcher & Munson (1933) and resulted in equal loudness curves (standardized as ISO 226:2003) which can be used to correct for the ear's frequency response. The perceived loudness of a sound is thus dependent on the pitch of that sound. The perception of loudness depends not only on pitch, but also on dynamics, duration (short term and integrated, or long term loudness), the type of signal (i.e. music, pure tones, noise) and whether or not the sound is stationary. The character of the nature scenes approximate a noise spectrum better than a pure tone spectrum.

In the music industry, louder sound is associated with better sound which led to the loudness war in commercial music. To prevent a loudness bias in the current thesis, the recordings needed to be loudness matched. The Just Noticeable Difference (JND) between 2 sounds is debated on in literature. Acoustically, 2 sounds are different in level when a person can observe a difference in level. Mills (1960) comments that the JND is about 1 dB at 1000 Hz, a little smaller at lower frequencies and 0.5 dB at higher frequencies. Crocker (2007) describes a JND of 0.8 dB at 40 dB SPL for noise sounds. (Houtsma et al., 1980) reports a threshold of 0.5-1 dB for broadband sound levels higher than 20 dB SPL, and 0.5 dB at 80 dB SPL for pure tones. More recently, the JND of stationary pink noise sounds is reported as 0.6 dB, where non-stationary sounds show a JND of 2.6 dB (Rummukainen et al., 2021). From all previous research, different figures for JND's are found. The lowest JND found in literature was 0.25 dB for broadband sounds above 60dB in level (which compares to the sound level of normal conversation) and at frequencies of 1000-4000 Hz (Pierce, 1983). Within this dB difference, the levels of the videos were matched.

Since the electrical signal strength (see Table 25) did not correlate well with the perceived loudness, other methods of loudness measurement were needed. From room acoustics, the L_{Aeq} parameter was researched. This parameter was used for example to measure the perceived loudness of planes around Schiphol airport, and does take the duration of the sound into account with an A-weighing. It measures the energy of the sound and averages it over the integration period. Inconclusive L_{Aeq} measurements were performed with the demo version of Dirac 6 (Dirac Research AB, 2021). Unfortunately, the audio files that arised from level matching using the L_{Aeq} parameter differed substantially when listening to them through headphones. This could be due to the nature of the sounds, or the A-weighing of the sounds. Most of the low frequency content is not taken into account (due to the characteristic of the A-weighing), where natural wind noises have a large amount of low frequency energy. Schlittenlacher et al. (2017) conclude about the L_{Aeq} parameter that "Although the L_{Aeq} correlates highly with the loudness of many sounds, it does not consider the effects of duration on loudness". Inaccurate integration of signal level over time could also be a source for the perceived varying loudness levels of the audio material. Instantaneous loudness and Overall loudness are to be considered separately (Schlittenlacher et al., 2017). A factor that is interesting in this regard is the crest factor, giving an indication of how much power there is in a given time. This also helps in taking the dynamics of a track in consideration.

Scene	Peak loudness	Average loudness
$Water^{Low}$	L: -9.31 dBFS	L: -19.47 dBFS
	R: -9.36 dBFS	R: -19.51 dBFS

$Water_{Medium}$	L: -10.24 dBFS	L: -22.79 dBFS
	R: -13.23 dBFS	R: -23.24 dBFS
$Water_{High}$	L: -14.19 dBFS	L: -24.05 dBFS
	R: -13.77 dBFS	R: -23.46 dBFS

Table 25: Signal strength with respect to 0dBFS

Another way of measuring the perceived loudness of broadband sounds needed to be used and was found in measurement with LUFS (Loudness Unit Full Scale), which are used in the recording industry and for streaming music services. LUFS (also called LKFS) are Loudness Units Full Scale and are described in ITU-R standard BS-1770-4 (2015). LUFS are a way of measuring loudness adopted by the broadcast industry to equalize the volume between for example programs and commercials. The K stands for the K-weighting of the filter used. After filtering, the loudness is assessed, integrated over time. A peak level and a LUFS level indicate the perceived loudness level on the user's side. YouTube (Google LLC, 2021) uses a standard level of -14 LUFS, source material under this threshold will not be amplified (Nikolic, 2020). For the current research, a level of -35 LUFS was observed as shown in Table 26, so loudness was not altered by the video platform.

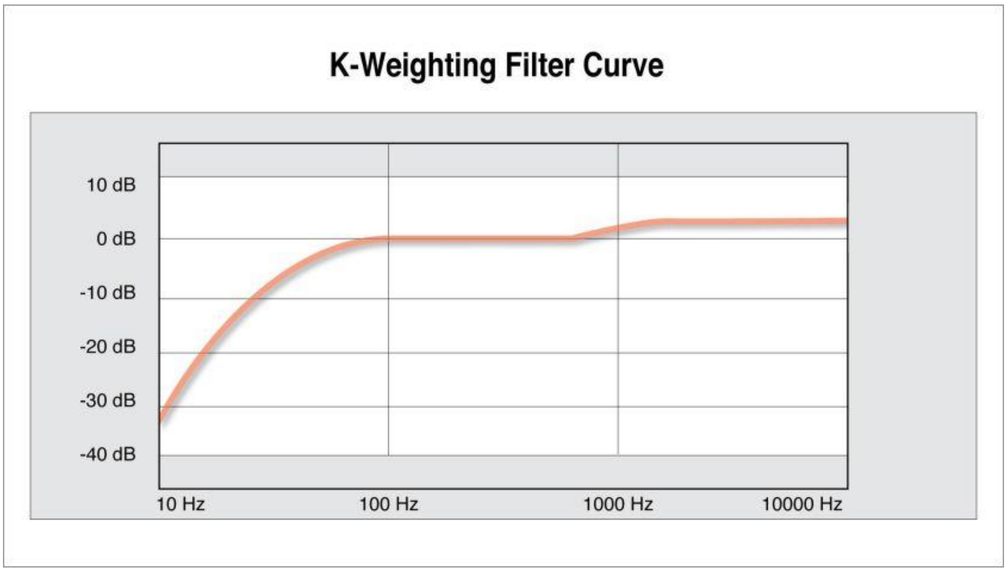

Figure 16: K-weighting filter characteristic — https://sebasthianfalk.wordpress.com/2017/09/29/loudness-standard/ retrieved 20-4-2021

The freeware Youlean Loudness Meter (J. Nikolic, 2020) was employed to measure the perceived loudness of the audio in the scenes. The videos containing the stimulus material were downloaded and then measured for loudness to also take audio processing of the YouTube (Google LLC, 2021) AAC compression algorithm into account. This was an iterative process, adjusting the level of the audio of all scenes until the perceived loudness was within the JND of 0.25dB. After calibration, the level was accurate and adjusted to within 0.1dB, as displayed in Table 26. All videos were uploaded on December 15, 2020.

Scene	Condition	Audio file	Video file	Loudness	YouTube link

$Water_{Low}$	A	Scene D video edit v2 -35 LUFS.wav	White background	-35.5 LUFS	https://youtu.be/YvbjnVfDVBU
$Water_{Low}$	AV	Scene D video edit v2 -35 LUFS.wav	Scene D V 25-10-2020.MP4	-35.5 LUFS	https://youtu.be/jIhUODvsiDQ
$Water_{Medium}$	A	Scene C video edit v2 +24dB +3dB.wav	White background	-35.6 LUFS	https://youtu.be/tYOEfEC48KQ
$Water_{Medium}$	AV	Scene C video edit v2 +24dB +3dB.wav	Scene C AV Voriginal L0.966 a-0.048 b-0.048.MP4	-35.6 LUFS	https://youtu.be/66zL1XBEgNk
$Water_{High}$	A	Scene F video edit -35 LUFS.wav	White background	-35.6 LUFS	https://youtu.be/n91BUBfG8-w
$Water_{High}$	AV	Scene F video edit -35 LUFS.wav	Scene F V 25-10-2020.MP4	-35.6 LUFS	https://youtu.be/LEeLadRhRXY

Table 26: Loudness levels for Audio and AudioVisual conditions

8.8 Hearing test

The objective was to find or make a listening test that could discern whether participants were using only one earpiece of their headphones instead of two, as well as identifying participants with moderate to severe hearing damage in one or both ears. Finally, participants with monaural playback equipment and laptop speaker users also needed to be identified, since that would introduce a bias because of a loss of spatial information in the audio recording. For localized hearing, humans mainly use (among others) two ears which facilitate the perception of timing and volume differences, accompanied by spectral information caused by the head-shadow. This is covered in the head related transfer function (HRTF). Since the audio track for this research is not just stereo (the left channel can contain different information than the right channel), but binaural, working stereo audio playback with (preferably) a high degree of channel separation is needed to be able to mimic the differences in left-right hearing to convince the brain of the participant that he or she is in a natural soundscape environment. Participants were instructed to wear headphones using the text below:

"**You must wear headphones and be in a quiet room!** Otherwise you will fail the screening test and may not get paid!"

The first implementation of the hearing test comprised of the words 'twenty' and 'five, on the right and left earpiece respectively, embedded in noise. Due to channel separation problems induced by (mostly consumer grade) equipment (due to output impedance of headphone amplifiers and a common impedance path in the headphone leads) and the AAC audio compression algorithm YouTube uses (Google LLC, 2021), a participant could hear the word twenty in one earpiece, but also, albeit softly, in the other earpiece. That introduces a problem with the hearing test because the left channel sound bleeds into the right channel. The first test was made with the words twenty and five. After testing with high end equipment, initially no problems were identified. After a preliminary test with YouTube (Google LLC, 2021) audio compression and a broken headphone (only one earpiece was working), a problem with channel separation was apparent. To solve the problem, words with less sibilance were chosen (i.e. the 't' in twenty and 'f' in five contain much acoustic energy, therefore they 'leak' easier to the other channel when listening) to alleviate the channel separation problem. The numbers 'one' and 'nine', when spoken, contain far less sibilance (no 't', 'f' or 's' sounds) then 'twenty' and 'five'. Most consonants, when spoken, contain a consonant and a vowel. A vowel sound was also tried for perceived channel separation, so less of the left channel would 'bleed' to the right channel and the other way around. To wash out any sounds in the opposing channel, white noise was used. The white noise was faded in, after which the number sequence started. Five times the word "One" on the right audio channel, five times the word "Nine" on the left audio channel. Participants were then asked what they heard. The level of the noise was -23.67dB (with a peak of -16.59dB) relative to 0dB full scale. The recording of "one" plus noise was peaking at -6.74dB (mean level of -19.49dB), where the recording of "nine" plus noise peaked at -5.9dB (mean level of -18.87dB), all levels relative to 0dB full scale. The difference between "one" and "nine" is small and not disturbing or distracting for the type of test performed. Participants received the following instructions:

"Put your headphones on if you didn't do that already. Use both left and right earpieces or earbud(s), don't use speakers. Turn your Windows volume down if it is all the way up. Press the 'Play' button. Change the Windows volume until you can hear the sound comfortably. Please don't change the YouTube volume. Press 'Play' again and change volume if necessary. Please pay attention to what you hear in the video. What do you hear?".

In Table 27, the choice options for the hearing test are displayed. Radio buttons were available for the participant to make a selection.

#	Text with choice option
1	I hear noise in both ears, "one" from the right ear and nine" from the left ear
2	I hear noise in both ears, "one" from the left ear and "nine" from the right ear
3	I hear noise in both ears and the numbers "one" and "nine" from both ears
4	I can only hear noise
5	Only one earpiece is working

Table 27: Hearing test choice options

When answer option one was chosen, the participant would automatically continue to the sound level calibration. Option two would lead to instructions to exchange the sides of headphones: "Your headphones appear to be working, but are on the wrong way around. Please change Left and Right and press "Next".".

Option three to five would lead to an error message: " You can close the tab using the cross in the tabs bar or press Ctrl + W.". No further data could be filled in, participants were excluded.

8.9 Headphone level calibration

Setting an equal level for all participants is desirable, since a loudness bias can be present when listening to the natural stimuli and levels differ significantly. Due to the online nature of the research it was not possible to calibrate or set the volume at a participant's home. An approach could be to let participants set the volume level themselves when hearing nature sounds. This could bias the participants because they would hear nature sounds before they would hear the actual stimulus. An advantage though, is that participants will most probably hear what they are intended to hear at a — for them — comfortable level.

Another approach is to make sure the level over participants is equal. For the sound level to be comparable over participants, an effort was done to calibrate the level of the headphones to a stimulus that every participant has access to. Preliminary tests with beeps and claps failed to produce a consistent loudness level. A solution was found on the hearingtest.online website (Pigeon, 2021) in which the fairly consistent loudness of rubbing hands is used to calibrate the sound level of headphones to. Participants were asked to listen to the sound of recorded rubbing hands through headphones, then take off their headphones and rub their hands at 5 cm distance from their nose. If the sound of their own rubbing hands was louder than the recording, the volume of the headphones was to be turned up. Participants were to adjust the volume in the Windows operating system, since the volume of YouTube (Google LLC, 2021) can be re-set due to cookie settings. Doing so, the volume is consistent over the whole experiment. Participants were presented with the following text to set the Windows volume level:

We will now set the level of your headphones again:

- Press the "Play" button. You will hear hands rubbing.
- Remember the sound level and take off your headphones.
- Rub your hands together closely in front of your nose (about 5 cm / 2").
- Adjust the volume of your computer through Windows until the sound level of the video closely matches the sound of your own hands rubbing together.
- Check level matching with headphones back on.
- Repeat the process until levels match.
- Don't adjust the volume after this and press the "Next" button.

Depending on the effort a participant put in, a level matching within 0.5 dB is expected, according to literature in Appendix 8.7.

The sound of rubbing hands was recorded with a Sennheiser MKE42 with MS14P pre-amp and a Scarlett 2i2 2nd generation USB microphone preamp. The level of the rubbing hands recording had a mean level of -29.12dB (peak level of -15.88dB) relative to 0dB FS. The calibration stimulus (rubbing hands) when played through YouTube (Google LLC, 2021) had a loudness of -28.2 LUFS as measured with Youlean Loudness Meter 2 (J. Nikolic, 2020), with a true peak max level of -16.2dB. The loudness difference between the sound level of the stimuli and the sound of rubbing hands was 35.6 LUFS - 28.2 LUFS = 7.4 dB LUFS.

8.10 Stressor — SART procedure

Stress can be induced in experimental settings in various ways, both affectively and cognitively. Stressing participants purely cognitive has the advantage that a more predictable effect on the stress response occurs than in an affective scenario. Cognitive acute stress is therefore easier to reproducibly induce in participants. Also, arousal induction is to be avoided, since it could influence relaxation and thus the measures of fascination and soft fascination (Annerstedt et al., 2013). Chronic stress can not feasibly be induced in an experimental setting, not only due to ethical reasons.

Moreover, chronic stress can reduce the noticeable effect of acute stress (McGonagle & Kessler, 1990). Acute stress can be induced by several means. Protocols like the Trier Social Stress Test (Kirschbaum et al., 1993) or watching a stressful movie like faces of death (Martínez-Soto et al., 2013) stress the affective system. Reverse digit span tasks, the Attention Network Task (Fan et al., 2002) and the Sustained Attention to Response Task (Robertson et al., 1997) are forms of cognitive stressors. From literature research, the SART, ANT and Stroop procedures were promising since they worked in most studies. The SART was among others chosen for its sustained attention component (i.e. directed attention fatigue) easy explanation of the task and the high number of trials per unit of time. In 5 minutes, the SART had 204 trials, ANT 75 trials and Stroop 118 trials. Moreover, Berto (2005), Lee et al. (2015) and Shu & Ma (2019) used the SART procedure with success in their studies.

The Sustained Attention to Response Task (SART) (Robertson et al., 1997) was used to cognitively stress participants in the control and experimental condition and was written in Javascript. A description of the SART task is given in section 2.6.3. The SART stressor was tested and verified for usability in the pilot study with 30 participants (see section 3.1). Response times for pressing the space bar were recorded. Since the SART test was used as a stressor that was implemented before the restorative interventions, unfortunately none of the objective data the SART script recorded could be used to make inferences about how restorative the different manipulations were.

In the original SART procedure, as described by Robertson et al. (1997), 225 single digits (25 of each of the nine digits) were presented visually over a 4.3 min period. Each digit was presented for 250 msec followed by a 900 msec mask. Font sizes of the numbers were 48 point, 72 point, 94 point, 100 point and 120 point, respectively (Symbol font), corresponding to a height varying between 12 mm and 29 mm on a 215 mm x 135 mm screen. The total diameter of the circular mask was 29 mm. The period from digit onset to digit onset was 1150 msec. The number '3' was the target, where participants were not to press the space bar. All other presented numbers required the pressing of the space bar. The measurement of the response time (i.e. time between the presentation of a number and a space bar press) was recorded in an array of numbers. The original SART procedure was followed as closely as possible. For it was not possible to dynamically control for screen size, 5 pre-set font sizes were chosen. The practice run used 18 numbers (with two targets) with two font sizes, where the experimental run uses 225 numbers (with 25 targets) with five different font sizes. The fixation cross was programmed as an svg object to guarantee a similar shape over screens. The control and experimental conditions were both preceded by a practice period consisting of 18 presentations of digits, two of which were targets. Visual feedback on the participants' input was given using the words CORRECT (in green) and INCORRECT (in red), displayed on screen for 500 ms. A 900 ms grace period with a fixation cross was present so participants were not instantly confronted with a number. The visual feedback was not given during the main run with 225 numbers.

Timing accuracy and resolution

To verify the number presentation time of the SART script, a screen capture was made using OBS 25.0.8 software (Bailey, 2020) for screen recording, set for 60 fps. VLC software (VideoLAN-team, 2016) was used for analysis of the video footage and indicated a frame rate of the recording of 60.0024 fps. The presentation of numbers lasted 15 frames, which corresponds to 250 ms. The fixation cross was visible for 54 frames, corresponding to 900 ms. Finally, the feedback was presented for 30 frames, which equals 500 ms. The timing is thus verified to be accurate and follows the original SART procedure.

In the original procedure no timing resolution for the response was given (Robertson et al., 1997). The SART script used in this experiment was purposely written in Javascript, and has a recorded timing resolution of 1 ms. This was the highest measurement resolution obtainable with a browser based script. Research by Bridges et al. (2020) discusses timing accuracy with sub-millisecond resolution, but relay that by addressing studies that average over trials and so have low impact on statistical outcomes. They continue and state that most software packages for online research achieve a minimal precision of 10 ms in all browsers, although LimeSurvey (Schmitz, 2020) was not one of the tested software packages. Furthermore, they investigated sources of latency and concluded that "online methods can be suitable for a wide range of studies, with due thought about the sources of variability that result" (Bridges et al., 2020). The expected deviation of presentation of the number on screen is always +/- 1 frame, which is one divided by the frame rate in seconds. With a refresh rate of 60 Hz, an uncertainty of 16.6 ms arises. The lowest response times on the SART test resulted in a 1 ms response time minimal. On a mean response time of 300 ms (Bridges et al., 2020), a timing resolution for this study of 1 ms shows sufficient resolution; sub-millisecond resolution would not be needed. Absolute timing accuracy was not measured, as this would differ over hardware set ups as used by participants. Over 227 participants, an averaging effect is to be expected, reducing the need for a measurement of absolute timing accuracy.

Participant compliance verification

When analyzing the data and checking for compliance of the participants the number of correct commissions (216 maximum), number of omission errors (27 maximum), commission errors (216 maximum), correct omissions (27 maximum) and mean time for commissions (correct presses) were used to flag low effort responses by participants. When a participant showed constant low response times with many omission errors the participant was considered either early depleted or a non compliant participant.

Technical procedure

Fullscreen was forced during the entirety of the SART procedures, the only way to turn fullscreen off was by pressing F11; the Esc button had no effect. The mouse cursor was also turned off to prevent participants from getting distracted by the mouse pointer. The SART procedure was gracefully integrated into the JavaScript interpreter of the LimeSurvey (Schmitz, 2020) framework. The only visible switchover was the buildup of the used iframe over an empty LimeSurvey (Schmitz, 2020) question page. Technically, the survey was put on hold during the SART procedure, letting LimeSurvey (Schmitz, 2020) wait for input. In the browser the SART procedure was handled in an iframe floating in front of the LimeSurvey (Schmitz, 2020) content. When the SART procedure was finished, the output was handed over by the SART procedure to the LimeSurvey (Schmitz, 2020) question interpreter and stored with the experiment data.

SART script written in JAVAscript

Below is the full code for the SART script with comments to improve readability:

```
    const time = await this.timeResponse(" ", waitTime)  // SAVE RESPONSE TIME
    experiment[k].time = time;                           // STORE RESPONSE TIME IN
ARRAY
```

```
  },                                                    // RESPONSE TIME MEASUREMENT
  pressKey: function (key) {                              // KEY DOWN REGISTRATION
    return new Promise(resolve => {
      const keyTrigger = 'keydown';
      document.addEventListener(keyTrigger, function onKeyPress (event) {
        if ((key === event.key) && (!event.repeat)) {
          document.removeEventListener(keyTrigger, onKeyPress)
          resolve()
        }
      });
    });
  },
  timeResponse: async function (key, timeout) {           // WAIT FOR KEYPRESS
    key = key || " "; // keypress to check
    timeout = timeout || 900; // Timeout in ms

    const start = new Date().getTime();
    await Promise.race([this.sleep(timeout), this.pressKey(key)]);
    const stop = new Date().getTime();
    return stop - start;
  },

 sleep: function (ms) {
    return new Promise(resolve => setTimeout(resolve, ms))
  },
  createExperiment: function (numbers, sizes, repetitions) {
    const experiment = [];

    for (let k = 0; k < numbers.length; k++) {
      for (let l = 0; l < sizes.length; l++) {
        for (let m = 0; m < repetitions; m++ ) {
          experiment.push({
            "number": numbers[k],
            "size": sizes[l],
          })
        }
      }
    }
    return experiment;
  },
  // Fisher-Yates aka Knuth shuffle, see https://stackoverflow.com/questions/2450954/how-
to-randomize-shuffle-a-javascript-array
  shuffle: function (array) {
    var currentIndex = array.length, temporaryValue, randomIndex;

    // While there remain elements to shuffle
    while (0 !== currentIndex) {

      // Pick a remaining element
```

```
        randomIndex = Math.floor(Math.random() * currentIndex);
        currentIndex -= 1;

        // And swap it with the current element
        temporaryValue = array[currentIndex];
        array[currentIndex] = array[randomIndex];
        array[randomIndex] = temporaryValue;
      }
    return array;
    },

      welcomeText: [
        "In this task, a series of numbers will be presented to you. For every number that appears except for the number 3, you are to press the space bar as quickly as you can. That is, if you see any number but the number 3 , press the space bar. If you see the number 3, do not press the space bar or any other key.",
        "Please give equal importance to both accuracy and speed while doing this task.",
        " Press the b key when you are ready to start.",
      ],
      trialText: [
        "We will now do some practice trials to familiarize you with the task.",
        "Remember, press the space bar when you see any number except for the number 3.",
        "Press the b key to start the practice.",
      ],
      experimentText: [
        "We will now start the actual task.",
        " Remember, give equal importance to both accuracy and speed while doing this task.",
        "Press the b key to start the actual task.",
      ]
    }
  },
}
</script>

body {
  background: black;
  cursor: none;                          // TURN OFF MOUSE CURSOR OVER IFRAME
}
```

SART implementation in LimeSurvey

To implement the SART script and store the values generated in user interaction, the following script was used in LimeSurvey:

```
const params = new URLSearchParams(window.location.search)

const id = params.get('Id');

const url = `https://htionline.tue.nl/ae35b623/`;
```

```
        setTimeout(async e => {
      var bc = new BroadcastChannel('preview:compiled');
      bc.onmessage = function(message){
        console.log(message.data);

        document.getElementById('answer184494X3291X27874').value = JSON.stringify({
participant: id, results: message.data });
        document.getElementById('ls-button-submit').click();
      };

        const iframe = document.createElement('iframe');
        iframe.src = url;
      iframe.style.width = '100vw';
        iframe.style.height = '100vh';
        iframe.style.zIndex = 10000000;
        iframe.style.position = 'absolute';
        iframe.style.top = '0';
        iframe.style.left = '0';

        document.body.appendChild(iframe);
      //this.iframe[0].contentWindow.document.body.focus(); // still have to click first, then
press b
      iframe.contentWindow.focus();
      //onload="this.contentWindow.focus()"
      //iframe:focus {
      //      border-color: #ff0000; // { is an unexpected token
      //}
      //iframe.addEventListener("focus", focus, true); // still have to click first, then press b
    }, 100);

        const waitFor = async(target, event) =>
    {
      return Promise.race(event.split('|').map(e => new Promise(r => {
        target.addEventListener(e, d => r(d.detail), { once: true })
      })));
    };
</script>
```

8.11 Experiment and questionnaire software

Software for administering the experiment was needed with all introductory texts, explanations, screening tests, stressor, de-stressor and questionnaires. Following the criteria set for the experiment in Table 28, PsychoPy, Millisecond Inquisit, Lab.js and LimeSurvey were extensively tested. Since the experiment was to be run online, browser compatibility was a very important criterion for software selection and pre-screening with participants. Windows is the most omnipresent operating system, with Chrome as the most popular internet browser. All popular browsers were tested for compatibility with the SART script, which revealed that only Chrome, Edge and Internet Explorer were fully compatible. Other browsers all gave problems with the SART script or video embedding which would result in skipped parts during the experiment, making data for these participants not usable. A choice for LimeSurvey (Schmitz, 2020) was made since a license was available and overall implementation was relatively easy when compared to PsychoPy.

Name	**Allows data export**	**Allows script implementation (Javascript)**	**Allows video play**	**Allows pictures**	**Price**	**Limit to number of participants**
PsychoPy	Yes	SART is built in	Yes	Yes	Dependent on Pavlovia	No
Millisecond Inquisit	Yes	Yes	Yes, from server	Yes, from server	Licence available	No
Lab.js	Yes	Yes	Yes, as embed	Yes	Free	No
LimeSurvey	Yes	Yes	Yes, as embed	Yes	License available	No

Table 28: Selection criteria for software

8.12 Color Blindness questionnaire

Participants were to have accurate color vision, as stated in the eligibility criteria. A participant hindered by color blindness could bias the results, therefore a short color blindness test was needed to filter out these participants before the experiment.

A short Ishihara test with 16 discs containing numbers was proposed and prepared in one picture. Participants were asked how many of the discs contained a number and were given a drop down menu to give the answer. The goal was to prevent participants with severe color blindness to participate in the experiment. A problem with implementing an online Ishihara test is that color accuracy guarantees are limited because computer screens have varying color depth (in budget monitors leading to Mach bands), color gamut and usually lack color calibration. Limited literature was available on the subject of online color blindness testing, although using an online Ishihara test was judged valid for detecting color blindness, but was not valid for discerning the type and severity of color blindness (Van Staden et al., 2018). Unfortunately, this research has limited applicability since no mention was made of various important parameters like screen type, size (viewing distance 75 cm), color gamut, bits of data per color and color calibration. The only mention about the settings was: "The computer monitor for the online test was set against recommended display settings." The test used by Van Staden et al. (2018) was https://www.color-blindness.com/ versus using original Ishihara plates that only test red-green color blindness.

Color calibration is thus a problem, where the original and online Ishihara tests are about the ability to detect color differences. The proposed test would most probably discriminate and eliminate the worst color blind participants. A counter argument was that some people might not know they are colorblind. A choice was made to not implement the Ishihara discs, but to probe color blindness with the question: "Are you color blind?", followed by the answering options "*Yes, No, I'm not sure*" using radio buttons. If a participant answered with "Yes", a text was displayed about the eligibility criteria:

“The answer(s) you gave on the hearing and/or vision test indicated that there is a chance that you don't have normal hearing or eye sight, or a hardware malfunction is present.

In the eligibility criteria it was stated that a participant should have good hearing (no tinnitus) and normal or corrected to normal vision without color blindness. Also, both earpieces of the headphones should be worn and functioning.

You cannot participate in this study.

You can close the tab using the cross in the tabs bar or press Ctrl + W."

8.13 Pre-pilot — stimuli test

The full rubric that was used to measure the proposed video files (translated from Dutch to English).

1 — Audio
Grade between 1-10 for the items below:
Pretty
Naturalness
Interesting
Annoying
Spatiality
Feeling of being there
Amount of birds
Amount of wind noise
Amount of water sounds
Weird things that are not supposed to be there
Relaxing / gives me peace
Congruent (it fits together)
I had no idea what I was listening to
Calmness of the birds
It grabbed my attention
Detail level (enough different things to hear)
Loudness
Open questions:
Does it sound natural to your ears?
Do you miss anything?
What was apparent about the sounds?
Was the spaciousness ok (more or less)?

Table 29: Questionnaire for auditory measures

2 — Image
Grade between 1-10 for the items below:
Pretty
Naturalness
Interesting

Annoying
Spatiality
Feeling of being there
Amount of birds
Amount of wind motion in vegetation
Amount of water
Accessibility of water
Weird things that are not supposed to be there
Relaxing / gives me peace
Congruent (it fits together)
Calmness of the birds
It grabbed my attention
Detail level (enough different things to see)
Colors

Table 30: Questionnaire for visual measures

8.14 Scene choice for $Water_{High}$ condition in pre-pilot

To make a choice between scene A and F for the $Water_{High}$ condition, the ratings obtained in the pre-pilot test are compared on the most comparable items in Table 31 below. The scenes are displayed visually in Figure 14 (section 2.3.5).

n=5 M = mean. Var = variance. SD = standard deviation	High water Scene A — auditory			High water Scene F — auditory			Means comparison	SD comparison
Auditory items	M	Var	SD	M	Var	SD	Choice	Choice
Pretty	6,75	1,69	1,30	6,00	2,00	1,41	F	A
Naturalness	8,20	1,36	1,17	8,00	1,20	1,10	F	F
Interesting	6,60	2,24	1,50	7,20	0,56	0,75	A	F
Spatiality	6,80	0,56	0,75	6,80	0,56	0,75	F	F
Feeling of being there	7,20	0,56	0,75	7,00	0,40	0,63	F	F
Relaxing / gives me peace	6,40	3,04	1,74	5,60	2,64	1,62	F	F
It grabbed my attention	6,20	1,76	1,33	7,00	0,80	0,89	A	F

Scene choice Means		Scene choice SD
A	2	1
F	5	6

Table 31: Pre-pilot scene choice for auditory items

Scene F received higher ratings with a high agreement (low SDs) than scene A for the auditory ratings.

n=5 M = mean. Var = variance. SD = standard deviation	High water Scene A — visual			High water Scene F — visual			Means comparison	SD comparison
Visual items	M	Var	SD	M	Var	SD	Choice	Choice
Pretty	5,75	7,69	2,77	7,00	0,50	0,71	A	F

Naturalness	8,40	0,24	0,49	8,40	1,04	1,02	F	A
Interesting	7,00	0,80	0,89	6,40	2,64	1,62	F	A
Spatiality	6,60	1,04	1,02	7,80	0,16	0,40	A	F
Feeling of being there	6,40	0,24	0,49	7,00	0,40	0,63	A	A
Relaxing / gives me peace	7,80	0,56	0,75	7,40	1,04	1,02	F	A
It grabbed my attention	6,40	0,64	0,80	6,20	0,56	0,75	F	F

Scene choice Means		Scene choice SD
A	3	4
F	4	3

Table 32: Pre-pilot scene choice for visual items

The visual ratings for scene F are higher, while showing slightly higher deviations. When adding the results for auditory and visual ratings, scene F scored better on restorativeness in accord with the ratings given by the 5 pre-pilot test participants. Therefore, scene F will be used for the $Water_{High}$ scene instead of scene A.

8.15 Texts used to inform participants

Participant information form

This document gives you information about the study “The effect of nature on attention”. Before the study begins, it is important that you learn about the procedure followed in this study and that you give your informed consent for voluntary participation. Please read this document carefully. Aim and benefit of the study The aim of this study is to measure human performance under different conditions. This information is used to better determine the effect of stress responses on performance. This study is performed by Roelof Lochmans, a student under the supervision of Dr. Kynthia Chamilothori and Prof. Wijnand IJsselsteijn of the Human-Technology Interaction (HTI) group of the Eindhoven University of Technology. Eligibility criteria You are between the age of 18 and 45 and are proficient in the English language. For the experiment you use a desktop computer or laptop (no tablet or phone) running the Windows operating system with one of the following browsers: Chrome, Edge or Opera. You turn off apps or programs that influence the color of your screen; for example f.lux, Iris or SunsetScreen. You have normal or corrected-to-normal vision, are not colorblind, and have no hearing loss or hearing deficiencies (i.e. hyperacusis, ringing or tinnitus). Lastly, you have stereo headphones that you will be using for the duration of the experiment. Procedure In this study you will receive a short hearing test, a short test for color blindness and perform 2 cognitive tasks, each followed by a questionnaire. You are asked to wear headphones during the study. Risks The study does not involve any risks or detrimental side effects. However, as you will be asked to perform cognitive tasks that demand your attention, you might feel fatigued after doing so. Duration The instructions, measurements and debriefing will take approximately 20 minutes of your time. Participants You were selected from the Academic Prolific database, accessible for researchers and participants at https://www.prolific.co/. Voluntary Your participation is completely voluntary. In the consent form, you can indicate whether or not you agree to also use your data for scientific purposes. You can refuse to participate without giving any reasons and you can stop your participation at any time during the study. You can also withdraw your permission to use your data immediately after completing the study. None of this will have any negative consequences for you whatsoever. Compensation You will be paid £2.1 through Prolific following the completion of the experiment. Confidentiality, use and storage of data All research conducted at the Human-Technology Interaction group adheres to the Code of Ethics of the NIP (Nederlands Instituut voor Psychologen – Dutch Institute for Psychologists), and this study has been approved by the Ethical Review Board of the HTI department of the TU/e. In this study personal data (for example your age, gender, and if applicable your participant database ID) and experimental data (your response times and responses to questionnaires) will be recorded, analyzed, and stored. The goal of collecting, analyzing, and storing this data is to answer the research question and publish the results in the scientific literature. To protect your privacy, all data that can be used to personally identify you will be stored on an encrypted server of the Human Technology Interaction group for at least 10 years that is only accessible by selected HTI staff members. The coded data collected in this study will (to the best of our knowledge and ability) not contain information that can identify you. No information that can be used to personally identify you will be shared with others. No video or audio recordings are made that could identify you. We will not share personal information about you with anyone outside of the research team. Only the researchers will know your identity and responses and we will store that information in an encrypted and password protected database. Sharing of data The data collected in this study might also be of relevance for future research projects within the Human Technology Interaction group. The aim of those studies might be unrelated to the goals of this study. The collected data can therefore also be made available to HTI researchers in an online data repository with restricted access. In the next page, you can indicate whether or not you agree with the use of your data for future research within the Human Technology Interaction group. You are not obliged to let us use and share your data. If you are not willing to share your data in this way, you can still participate in this study. Then your data will only be used in any scientific article resulting from this study but will not be shared with other researchers. Further information If you want more information about this study, the study design, or the results, you can contact Roelof Lochmans (contact email: r.h.lochmans@student.tue.nl).If you have any complaints about this study, please contact the supervisor, Dr. Kynthia Chamilothori, at humantechnologyinteraction.tue@gmail.com. You can report irregularities related to scientific integrity to confidential advisors of the TU/e.

Informed consent form for participants

Informed consent form Study Title: "The effect of nature on attention"

- I have read and understood the information of the corresponding information form for participants.
- I have been given the opportunity to ask questions. My questions are sufficiently answered, and I had sufficient time to decide whether I participate.
- I know that my participation is completely voluntary. I know that I can refuse to participate and that I can stop my participation at any time during the study, without giving any reasons.
- I know that I can withdraw permission to use my data up to 24 hours after the data have been recorded.
- I agree to voluntarily participate in this study carried out by the Human Technology Interaction research group of Eindhoven University of Technology.
- I know that no information that can be used to personally identify me or my responses in this study will be shared with anyone outside of the research team.

Certificate of consent
Press Next if you agree to the statements above to voluntarily participate in this study and provide consent to participate in this study. Additionally, if you agree that your data may also be used for future research, please check the option below.

Certificate of consent
By clicking the NEXT button you agree that you want to voluntarily participate in this study and provide consent to do so.

Experiment introduction

Find a calm and quiet place where you can sit behind a desk or table and will not be disturbed or distracted (also by outside events) for at least 20 minutes. Use a cabled internet connection if possible, or sit close to your wifi access point. Important for the research is that you turn your phone OFF. Great place and posture for research:

Figure 17: Preparation - seating position

The picture in Figure 17 was displayed adjacent to the text. On the next page, the following instructional text was shown:

Please close all programs except your browser. Press F11 to put the browser in full screen mode. You will be reminded how to turn off the full screen mode at the end of the study. Please note that in this study it is not possible to go back to previous pages. When you press the "Next button", your answers to questions are definitive.

Debriefing

You have reached the end of the experiment. Thank you for participating! Your response is very valuable to us: this study aims to investigate how experiences from nature contribute to stress reduction. To keep the study valid for other participants as well, please do not share details about the experiment itself! You can turn your phone back ON again. You can now press F11 to turn off fullscreen mode. By pressing "Next" you save all your responses. You will be redirected to Prolific automatically. * You can also choose to not save your answers and leave the experiment. You can do this by clicking the grey "Exit and clear survey" text in the upper right corner.

8.16 Browser identification with User agent string

To identify the operating system and browser the participant was using, the user agent string was saved in LimeSurvey (Schmitz, 2020) using the code below:

```
<script>
  window.onload = function() {
    document.getElementById('question24019').style.display = "none";   // HIDE QUESTION
    document.getElementById('answer347974X2584X24019').value = navigator.userAgent;
// STORE THE VALUE "navigator.userAgent" IN SURVEY ID 347974, GROUP ID 2584, QUESTION ID
24019
   };
</script>
```

The user agent string was then decoded using the website https://developers.whatismybrowser.com/useragents/parse/#parse-useragent (WhatIsMyBrowser.com, 2021). Data was stored alongside the original user agent string information. This data was then analyzed for wrong operating systems, hardware and browsers. Participants not using the correct device or software could not see the SART task and skipped through the embedded videos because they were blocked by the browser or got stuck during the research. These participants did not have complete data and were excluded from the analysis.

8.17 Random video play for pilot

For delivery of the restorative stimulus containing natural elements, YouTube (Google LLC, 2021) was chosen as platform. Since this research contains 9 experimental conditions, 9 videos were made. To prevent biases during the pilot phase, the videos were randomly assigned to the participants using a short piece of code written in JavaScript and embedded in LimeSurvey (Schmitz, 2020), which was then called by the browser of the participant. The video that was played during the experiment was saved as a video ID and the index in the random generation list: {"id":"h97oNNsJsUg","index":8}. Storing both the video ID and the index was done for checking and recoding purposes.

The code for random video play is displayed and explained below:

```
<script>

        setTimeout(() => {

    // add YouTube video link codes (not the full links!) as needed, the code plays one randomly
selected video. The used video code is stored with the participant ID.

    const vidIds = [
     // AudioVisual (CDF)
     'qXMN0klqO98',
     'oLvyNobbECY',
     'y09wyo3vkDA',
     // Audio (CDF)
     'Jpk_UYG336U',
     '7JgFqpqZjqE',
     'qc_tOnC2Abw',
     // Video (CDF)
     'qd143j8vUEA',
     '2AUVVqo1zyE',
     'byuCrZA5K9Q',
    ];

    const i = Math.floor(Math.random() * vidIds.length);          // RANDOM SELECTION
    const id = vidIds[i];

    document.getElementById('answer236682X537X5885').value = JSON.stringify({ id, index: i });
                                                        // STORE ID AND INDEX IN LIMESURVEY
```

```
    const iframe = document.createElement('iframe');            // PLAY VIDEO IN IFRAME
    iframe.allow = 'autoplay; encrypted-media;';
    iframe.src = `https://www.youtube.com/embed/${ id
}?rel=0&amp;controls=0&amp;showinfo=0;autoplay=1`;              // URL GENERATION
    iframe.style.width = '100vw';                               // STYLE OF IFRAME
    iframe.style.height = '100vh';
    iframe.style.zIndex = 10000000;                             // PLACE VIDEO IFRAME ON TOP
    iframe.style.position = 'absolute';
    iframe.style.top = '0';
    iframe.style.left = '0';
    iframe.style.cursor = 'none';                               // TURN OFF MOUSE CURSOR
    iframe.style.border = 'none';                               // REMOVE WHITE BORDER

    document.body.appendChild(iframe);

    document.documentElement.requestFullscreen();               // PUT IN FULLSCREEN

    setTimeout(() => {

        document.exitFullscreen();

      document.getElementById('ls-button-submit').click();

    }, 62000);          // AFTER 62000ms THE "Next" BUTTON IS PRESSED AUTOMATICALLY
  }, 200);
</script>
```

8.18 Fascination questionnaire

For measuring fascination several measurement instruments were evaluated. See Table 33 for an overview of reviewed scales. The Perceived Restorativeness Scale (PRS) (Hartig et al., 1997) was most prominently present in the reviewed literature — and thus best comparable with other literature — as well as being most applicable and delivering the most positive results. Another advantage of using the PRS is that it is a validated instrument with a high reported internal consistency. The PRS consists of 26 items, divided into subscales that measure Being away, Fascination, Coherence and Compatibility from ART theory. The PRS development goal was to develop a valid, reliable measurement tool that could represent the constructs of Being away, Fascination, Extent (consisting of Scope and Coherence) and Compatibility (Hartig et al., 1997). The statements were chosen so they are not referring to natural, built, indoor or outdoor environments. The fascination subscale consists of five items and employs a seven point Likert scale (0 = not at all, 6 = completely). The Fascination scale of the PRS used in this study is displayed in Table 3. Item four from study one was not included, because in their study two it was replaced with item five (Hartig et al., 1997, p. 182).

Questionnaire	**PRS**	**STAI**	**PANAS**	**ZIPERS**	**POMS**
Name of questionnaire	Perceived Restorativeness Scale	State Trait Anxiety Inventory	Positive And Negative Affect Scales	Zuckerman Inventory of PErsonal	Profile Of Mood States

				ReactionS	
Reference	Hartig et al. (1996)	Spielberger (1983)	Watson et al. (1988)	Zuckerman (1977)	McNair et al. (1971)
Used in reviewed literature:	Berto (2005)	Diette et al. (2003)	Berman et al. (2008)	Ulrich et al., (1991)	Barton & Pretty (2010)
	Pasini et al. (2014)	Annerstedt et al. (2013)	Bratman et al. (2015)	Valtchanov et al. (2010)	Ma & Shu (2018)
	Jahncke et al. (2015)	Bratman et al. (2015)	Dopko et al. (2019)		Wooller et al. (2018)
	Lee et al. (2015)		Zhang et al. (2019)		Zhang et al. (2019)
	Sona et al. (2019)				
	Shu & Ma (2019)				

Table 33: Comparison of questionnaires in reviewed literature

Additionally, presence and immersion were also shortly investigated; these measures indicate how engaged participants are with the stimulus. Presence is described as the perceptual illusion of non-mediation (Lombard & Ditton, 1997). Immersion is "a psychological state characterized by perceiving one-self to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences" (Witmer & Singer, 1998). De Kort et al. (2006) did research on the effect of presence on restoration with inconclusive results. Adding a measurement for presence and immersion could be an addition to the body of research because it could possibly answer or exclude part of the question "why" a natural setting is restorative. Because of the online nature of the research, the lack of objective measures to compare the presence measurements with and the limited time that the attention of a participant is available it was chosen not to measure presence and immersion.

8.19 Demographics questionnaire

The full set of demographics questions is displayed in Table 34 below:

#	Question	Answer format	Answer options
1	What is your gender? / With which gender do you associate yourself the most?	Radio buttons	Male Female Prefer not to say Other
2	What is your age?	Fill in field	Numbers
3	Which country are you from?	Drop down menu	List of countries
4	Which country are you	Drop down	List of countries

	currently living in?	menu	
5	When relaxing, which option do you prefer?	Radio buttons	Hearing sounds, for example music Looking at pictures, for example pictures No preference between the two options
6	What computer type were you using?	Radio buttons	Laptop Fixed PC setup Tablet Phone Other
7	What is the (diagonal) size of your computer screen?	Radio buttons	Between 10 and 15 inches / between 25 and 38 cm Between 15 and 19 inches / between 38 and 48 cm Between 19 and 25 inches / between 48 and 63,5 cm I don't know Other (Size or type number)
8	What type of headphones were you using?	Radio buttons	Earbuds Over ear headphones Speakers Other (fill in Brand and Type)

Table 34: Demographics questionnaire with answering options

Questions one to six and question eight are reported in section 2.1. The distribution of data on question seven (diagonal screen size) is reported below, in Table 35.

Category	Frequency
Between 10 and 15 inches / between 25 and 38 cm	55
Between 15 and 19 inches / between 38 and 48 cm	69
Between 19 and 25 inches / between 48 and 63,5 cm	64
I don't know	30
Other:	9
- 27 inch	6
- 32 inch	2
- 49 inch	1
- 1920 x 1080	1

Table 35: Distribution of screen size

Scene	Frequency	Male	Female	Other
Video - $Water_{Low}$	24	12	12	0
Video - $Water_{Medium}$	25	12	12	1
Video - $Water_{High}$	24	11	12	1
Audio - $Water_{Low}$	25	13	12	0
Audio - $Water_{Medium}$	26	13	13	0
Audio - $Water_{High}$	26	12	12	2
AudioVisual - $Water_{Low}$	26	13	13	0
AudioVisual - $Water_{Medium}$	24	11	13	0
AudioVisual - $Water_{High}$	27	13	14	0

Table 36: Distribution of participants over conditions — including gender