AALBORG UNIVERSITY COPENHAGEN

Semester: Medialogy M.Sc. 4th Semester - Spring 2021

Title: **CraveInVR**: Investigating immersive VR technologies to simulate craving-inducing social situations targeting use in CET for individuals with AUD

Aalborg University Copenhagen

A.C. Meyers Vænge 15

DK-2450 Copenhagen SV

Semester Coordinator: Stefania Serafin

Secretary: Lisbeth Nykjær

Project Period: February 1st – May 28 th 2021

Semester Theme: Master's Thesis

Supervisors:

Rolf Nordahl rn@create.aau.dk
Ali Adjorlu adj@create.aau.dk

Members:

Lasse Embøl lsaren16@student.aau.dk
Carl Hutters chutte16@student.aau.dk
Andreas Junker ajunke16@student.aau.dk
Daniel Reipur dreipu16@student.aau.dk

Abstract:
The study presents an investigation of the potential of using immersive VR (IVR) experiences in the field of cue exposure therapy for individuals with alcohol use disorder. Current tools of eliciting craving for therapists to teach coping skills do not fully consider the complexity of alcohol cues, with emphasis on the effect of social presence and social pressure, and do not properly represent high-risk situations experienced by individuals in reality, leading to transfer gaps and missing learning generalizations. Therefore, tools should approach real-life situations as much as possible, leading to the experiments presented in this study. A within-subjects experiment (n=25, m=16, f=9, SD=38.3) utilized a novel presence-matching design, exploring which elements of a 3D-IVR experience of a virtual bar environment contributed to participants' subjective feelings of presence by letting them upgrade certain parameters of the environment. Results showed a strong preference for upgrading parameters of, in order, soundscape richness, virtual agent animations, virtual agent reactions, and geometric realism. However, subjective intervals chosen for parameter fidelities in the implementation could skew results. Another within-subjects experiment (n=26, m=15, f=10, SD=30.83) compared a virtual bar environment in a 3D-IVR experience with a similar 360-IVR experience, comparing reported feelings of alcohol craving (F(2,75)=13.7, p=.995), as well as feelings of presence through illusions of place (PI) (F(2,75)=8.38, p=.017) and plausibility (Psi) (F(2,75)=19.1, p=.014), triangulated with physiological measures and a qualitative follow-up assessment. A neutral virtual scene of a forest was used to obtain baselines between every virtual bar condition. While results showed a statistically significant difference in feelings of PI and Psi with a slight preference for the 360-IVR experience, differences in reported feelings of alcohol craving were statistically insignificant. Interestingly, physiological data revealed greater levels of affect in the 3D-IVR experience. Regardless of conditions, both comparisons of conditions with the neutral forest scene was statistically significant for craving (F(2,75)=13.7, p=.001), suggesting a potential for an IVR-experience of a virtual bar to be able to induce feelings of craving in individuals.

Copyright © 2021. This report, appended, and/or related material may not be partly or completely published or copied, or be used for commercial purposes, without prior written approval from the authors.

CraveInVR

Investigating immersive VR technologies to simulate craving-inducing social situations targeting use in CET for individuals with AUD

Lasse Embøl, Carl Hutters, Andreas Junker, Daniel Reipur *
ANDALACA, MED10, Aalborg University Copenhagen

Rolf Nordahl, Ali Adjorlu †
Supervisors, ME-Lab, Aalborg University Copenhagen

The virtual bar environment modelled after Tagensborg Bodega in Copenhagen.

Abstract— The study presents an investigation of the potential of using immersive VR (IVR) experiences in the field of cue exposure therapy for individuals with alcohol use disorder. Current tools of eliciting craving for therapists to teach coping skills do not fully consider the complexity of alcohol cues, with emphasis on the effect of social presence and social pressure, and do not properly represent high-risk situations experienced by individuals in reality, leading to transfer gaps and missing learning generalizations. Therefore, tools should approach real-life situations as much as possible, leading to the experiments presented in this study. A within-subjects experiment (n=25, m=16, f=9, SD=38.3) utilized a novel presence-matching design, exploring which elements of a 3D-IVR experience of a virtual bar environment contributed to participants' subjective feelings of presence by letting them upgrade certain parameters of the environment. Results showed a strong preference for upgrading parameters of, in order, soundscape richness, virtual agent animations, virtual agent reactions, and geometric realism. However, subjective intervals chosen for parameter fidelities in the implementation could skew results. Another within-subjects experiment (n=26, m=15, f=10, SD=30.83) compared a virtual bar environment in a 3D-IVR experience with a similar 360-IVR experience, comparing reported feelings of alcohol craving ($F(2,75)=13.7$, $p=.995$), as well as feelings of presence through illusions of place (PI) ($F(2,75)=8.38$, $p=.017$) and plausibility (Psi) ($F(2,75)=19.1$, $p=.014$), triangulated with physiological measures and a qualitative follow-up assessment. A neutral virtual scene of a forest was used to obtain baselines between every virtual bar condition. While results showed a statistically significant difference in feelings of PI and Psi with a slight preference for the 360-IVR experience, differences in reported feelings of alcohol craving were statistically insignificant. Interestingly, physiological data revealed greater levels of affect in the 3D-IVR experience. Regardless of conditions, both comparisons of conditions with the neutral forest scene was statistically significant for craving ($F(2,75)=13.7$, $p=.001$), suggesting a potential for an IVR-experience of a virtual bar to be able to induce feelings of craving in individuals.

1 Introduction

Alcohol use disorder (AUD) is a mental disorder defined by behavioral and physical symptoms, such as alcohol withdrawal, higher tolerance of consumption, and an intense desire to consume alcohol, defined as *alcohol craving* [1]. AUD is prevalent throughout the world and severely affects people in almost all age groups. It has significant human and economic costs, e.g., through law enforcement and healthcare, loss of productivity, and other direct and indirect consequences. These include harm to others, development of disabilities, as well as premature loss of life [2]. AUD is the world's third largest risk factor for disease and disability, contributing to, e.g., violent behavior, child neglect and abuse, mental health deterioration, and workplace absenteeism [3].

Recent studies reveal the COVID-19 pandemic and the resulting lockdowns to have enabled increased alcohol consumption through elevated levels of social isolation and uncertainty about job security, psychologically affecting the population [4–6]. As a result, the world has seen a general increase in alcohol relapse, as well as an increase in patients diagnosed with AUD [5, 6]. Finally, these individuals continue to consume alcohol despite their knowledge of its contribution to physical, psychological, social, or interpersonal problems [1]. To

*e-mail:{ajunke16, chutte16, dreipu16, lsaren16}@student.aau.dk
†e-mail:{rn,adj}@create.aau.dk

learn to cope with alcohol craving, patients must acknowledge cognitive problems directly related to the onset of alcohol consumption and relapse [7, 8]. Craving is more likely to be felt in so-called high-risk situations, where alcohol has been consumed previously, based on classical conditioning [9, 10]. The presence of others, providing a social context and social pressure in particular, also play a key role in inducing craving in AUD patients [11]. In addition, high-risk situations for AUD patients usually involve social interactions. Craving can be felt through exposure to alcohol cues, both during consumption, in a withdrawal phase, and during periods of abstinence, risking individuals to relapse into greater alcohol consumption and subsequent abuse [1]. Two main treatment approaches of pharmacological and psychotherapeutical assessment exist. This study will refrain from researching pharmacological assessment methods, instead focusing on psychotherapeutical methods.

A widely used psychotherapeutical treatment method focusing on behavioral assessment is *cognitive behavior therapy* (CBT). CBT challenges maladaptive thoughts, changes behavior, and improves emotional regulation through techniques developing coping strategies in patients [12]. CBT has seen success in AUD treatment as it teaches and helps patients to better cope with future high-risk situations [13].

A practice used in combination with CBT is *cue-exposure treatment* (CET) [14]. CET utilizes the cue reactivity and undesired response that has been conditioned in the patient. In a substance use disorder context, CET exposes individuals to high-risk situations and craving-inducing cues while prohibiting consumption of the substance. Through repeated exposure, the experienced craving level in patients will, in theory, decrease [14]. Typically, after cue exposure, a session with a therapist assesses the feelings, thoughts, and emotions of the patient to apply skills training and other techniques to seek cognitive and behavioral change [12]. Methods of inducing craving in patients include visual cues (tools such as images or videos) or having patients imagine or recall and retell high-risk situations [15]. Some studies even use physical cues as tools allowing patients to touch and smell alcohol, and in modified versions of CET, where the treatment goal is moderate drinking, consume it [10, 14, 16]. CET has originally been used and has seen large effect sizes in the treatment of phobia and anxiety [10]. As such, studies have also been conducted on the efficacy of CET for AUD. There are, however, issues regarding the measurable evidence of the effectiveness of CET. A meta-review by Mellentin et al. concluded there are inconsistencies between studies in formats, number of sessions, types of cue stimuli, and more [14]. Seen in the meta-review, combining CET with coping skills training has shown to be an effective treatment method by itself, but that makes it difficult to determine the effectiveness of CET alone. However, such studies presented in the meta-review do still suggest CET for AUDs to be more effective than no treatment, and indicate that CET is at least as effective as other psychosocial treatments [10]. Additionally, CET for AUD has shown to be more effective than CET for other substance use disorders, making it more obvious to pursue [14]. However, while CET for anxieties has consistently demonstrated large effect sizes, CET for substance use disorders (including AUD) has to date shown less benefit [10].

Head-mounted display-based immersive virtual reality (IVR) is seeing increased use in the healthcare sector. The technology can be used to address, mitigate, and spread awareness on difficult problem areas of mood disorders, posttraumatic stress disorder, addiction, autism, and more [17–19]. Recently, IVR has also been explored in CET contexts, used as a tool to simulate high-risk situations to induce craving in patients, approaching actual *in vivo* exposure [10, 14]. Compared with traditional *ex vivo* tools (i.e., using images, video, imagination), IVR is highly immersive and can facilitate higher feelings of presence in a given setting. Results of comparing IVR tools to traditional cue-inducing tools in substance abuse generally reveal IVR experiences to be more effective in inducing craving [20, 21].

Most of the studies cited in this paper involving IVR as a tool in CET for AUD focus primarily on whether VEs are at all capable of inducing craving or not. However, many do not fully address the technological subaspects of the virtual environment (VE) either, and, amongst others, what role they play in affecting the perceived realism of the VE in general including social virtual agents. Additionally, comparisons between VR technologies are lacking regarding how they contribute to potential increases in craving. In that realm, two main parallel directions of IVR exist; presenting virtual environments in the medium either by the use of 360°-recorded video or by utilizing 3D-rendering [22, 23]. Two main differences between these concern the ability to act individually (i.e., the degree to which a user can move one's virtual self around), and to interact (i.e., use one's senses, hands, or body) to navigate and produce changes to the environment that subsequently reacts to that change. In environments presented by 360°-recorded video, users take part in an individual experience devoid of valid effectual sensorimotor actions, and can only rotate towards actions or events passively around their own axes (3DoF movement, the degrees of freedom available for pitch, roll, and yaw rotation) [24, 25]. In tracked 3D-rendered environments, 6DoF is available, adding translation to the user's locomotion abilities together with valid effectual sensorymotor actions [25]. On the other hand, 360°videos allow for capturing complex scenarios and maintain immersion through photorealism, which 3D-rendered VEs yet does not utilize in the same fashion [23, 26].

Combining the promise of IVR tools for therapy while looking at the previously mentioned social context that play a key role in inducing craving, social situations and interaction are also very complex elements to simulate in 3D-rendered VEs, whereas they are easier to portray in 360°video. To induce craving in an IVR-CET intervention as much as possible, the environment should feature such situations as it highly affects the believability of the VE, focusing on the 3D virtual agents or actors in the 360°video and their interaction coherence and complexity. Some studies have previously experimented with the inclusion or exclusion of virtual agents in general, to measure social pressure in IVR and how it affects craving [7, 27]. Other studies have then specifically tweaked VE fidelities and investigated how changes affected participants' sense of feeling the environment as being real [28–31].

To our knowledge, however, little research has been done on whether and how feelings of craving are affected by greater fidelities of social context, interaction, and pressure of virtual agents. Moreover, little research has been done on how the presentation and fidelity of the VE facilitates feelings of presence in the context of IVR-CET for AUD, also regarding the comparison of the effect between 3D-rendered and 360°-recorded VEs. As interventions seek to close the transfer gap between ex vivo clinical exposure and in vivo exposure, closing the gap would mean the learned skills and processes for coping with cravings should more easily translate to real life situations. As in vivo exposure undoubtedly elicits higher cravings, the ex vivo *therapy tools* should seek to be as immersive as possible, assuming to create greater feelings of presence, plausibility, and believability in patients to approach craving levels felt during in vivo exposure. The hypothesis states that an IVR experience can help close this CET transfer gap, and this paper seeks to determine the best technological approach in developing an IVR tool for AUD CET with a focus on social contexts.

Wile 360-video can excel in realism aspects, 3D VEs can excel in interactivity and locomotion aspects. These aspects are what this paper seeks to compare in the quest of eliciting craving, detailed in Section 4. Firstly, due to social contexts being crucial in simulated high-risk situations for CET, the VE must feel plausible to the user. Hence, the first experiment (*Experiment A*) focuses on discovering which elements of a 3D-rendered VE contributes the most to experienced plausibility in that environment. Secondly, also with the goal of simulating high-risk situations involving social aspects, the second experiment (*Experiment B*) compares 360°-video generated VEs with 3D-generated VEs with a focus on inducing alcohol craving in participants.

2 Background

This section presents research on related concepts, theories, methods, and tools brought forth in the introduction. It will elaborate upon the concept of AUD, alcohol-related cues, and craving, and how it is utilized in therapy and treatment. Furthermore, it will explain the concept and importance of context and social interaction in treatment methods, before presenting research of applicable immersive technologies that can include these aspects.

2.1 Alcohol Use Disorder

According to the *Diagnostic and Statistical Manual of Mental Disorders (DSM–5)*, alcohol use disorder (AUD) is defined as "*a problematic pattern of alcohol use leading to clinically significant impairment or distress*" [1]. Individuals diagnosed with AUD have symptoms such as frequent intake of alcohol in larger amounts over a prolonged period, resulting in failure to fulfill obligations at work, school, or home, or absenteeism from social or recreational activities. Frequent intake of alcohol in larger amounts leads to increased tolerance, in turn increasing the alcohol consumption amount needed to achieve intoxication or desired effects. Other symptoms cover continued use of alcohol and unsuccessful efforts to abstain from or cut down on consumption, despite knowledge of its contribution to physical (e.g., blackouts or liver disease), psychological (e.g., depression), social, or interpersonal problems (e.g., violence or child abuse). They often feel a strong desire to consume alcohol, known as *alcohol craving* [1]. Finally, individuals with AUD often spend much time in situations related to alcohol, either obtaining, consuming, or recovering from it. The severity of AUD is specified as either mild, moderate, or severe, depending on the number of symptoms the individual possess [1].

2.1.1 Why do AUD individuals drink?

Individuals with AUD have *conditioned* themselves to develop *habitual actions* and experience *craving* when exposed to certain *alcohol-related cues* [14, 32].

The process of associative learning between two otherwise unrelated stimuli is called *classical conditioning*, which over time forms new habitual actions in individuals [14]. In an alcohol-related example, classical conditioning can be explained by the alcohol representing the unconditioned stimulus. The known effects of consumption elicit the unconditioned response of craving. When a previously neutral stimulus is then repeatedly paired with the unconditioned stimulus, the neutral stimulus becomes associated with the unconditioned response over time. The pairing has made the previously neutral stimulus become a conditioned stimulus, an alcohol cue, capable of eliciting conditioned responses of craving [14].

Prolonged consumption of alcohol often results in increased consumption due to increased tolerance. That is, the perceived positive psychological and physical effects of alcohol diminishes. As it manifests, the AUD individual must conversely increase their consumption to obtain the same positive effect. This vicious cycle might lead the individual towards addiction [14]. In severe cases of addiction, whenever individuals try to break out of the cycle, they can physically respond with withdrawal symptoms such as anxiety, nausea, vomiting, and racing heart, and even tremors, cramps, and hallucinations leading to cardiac arrest [1].

Furthermore, continued alcohol consumption also leads to attentional biases towards alcohol cues [11, 33]. Attentional bias refers to how selective attention is directed towards cues that elicit strong emotional responses, and hereby also how it affects the individual's perception [34, 35]. Attentional bias has been linked to cases of information processing bias and behavior that leads to the development and maintenance of addiction, and may also prohibit an individual's ability to consider alternatives in situations where subjects of attentional bias are present [34, 36]. As such, an individual with AUD will direct their attention to alcohol-related cues if present, potentially distracting them from other relevant elements in the current situation. Additional factors contributing to continued alcohol consumption in individuals with AUD stem from interpersonal problems (e.g., breakdown of relationships and family relations [37]) and stigmatization [38]. Individuals with AUD apply existing stereotypes, prejudice, and discrimination to themselves. This has been associated with depressive symptoms, low self-esteem, and low self-efficacy [38]. Furthermore, individuals with AUD often suffer from comorbidities such as bipolar disorder, schizophrenia, antisocial personality disorder, anxiety, and depressive disorders [1, 39]. As such, individuals with AUD face great difficulty in coping with stressful or negative situations, responding to the situation by consuming alcohol [1, 40]. Alcohol may also be consumed to alleviate the unwanted effects of other substances (e.g., cocaine, heroin, or hypnotics), or to substitute them if they are not available [1, 41].

In attempts at quitting alcohol use, transgressions often occur. An initial transgression is defined as a lapse, the consumption of alcohol after trying to quit or after a treatment course. A lapse may lead to multiple transgressions, which eventually equate to *relapse* if the amount of alcohol consumption returns to what is similar to the previous level of consumption. As such, relapse may be defined as a process rather than a distinct event [42]. Individuals trying to remit from AUD often experience relapse [8, 43].

In addition to these AUD issues, patient lapses and relapses can be explained by several internal factors and events happening in the life of the patient, exhibiting in order [8, 42, 44]:

- An imbalanced or addictive lifestyle, or noncalculated irrational exposure to high-risk situations
- Ineffectiveness in coping responses in such situations
- Decreased self-efficacy, and distorted positive outcome expectancy for the effects of alcohol
- Lapses (initial use of alcohol)
- Abstinence violation and transgressions, leading to relapse

Studies have investigated whether craving is an important relapse determinant, showing contradicting results [8, 41, 42, 45]. Yet, craving is often used as a treatment outcome measure because of its role as a response to exposure [9, 44, 46]. Hence, knowing what causes craving may aid in improving components focusing on craving. In addition, *craving* was added to the list of AUD symptoms in the latest edition of the DSM (DSM-5), reiterating the relevance for further understanding craving [1]. Apart from the uncertainties revolving alcohol craving as a relapse determinant, some influences and triggers of craving have been established.

Craving influences individuals by interfering with their mental state through cognitive overload, affecting coping skills in high-risk situations [14]. Through repeated exposure and failed attempts at coping, self-efficacy decreases over time, in turn increasing the influence of craving. Craving also distorts the perceived outcome of alcohol consumption. These factors make it more likely for the individual to relapse [14].

2.1.2 What triggers Craving?

Alcohol craving is a cue reaction triggered and influenced by reactivity to alcohol stimuli and related cues [9, 44, 46]. Dependent individuals are seen to be more reactive to these stimuli and cues, triggering increased craving and negative affect in behavior [40].

These cues can either be proximal, contextual, or complex [47]: *Proximal cues* are the type of cues closest to actual stimuli or consumption, as they are presumably ubiquitous [47]. In alcohol contexts, they can be visual or auditory, such as seeing or opening a bottle, pouring the beverage into a glass, or watching the same unfold in a commercial on TV. They can also be olfactory, gustatory, or tactile, i.e., smelling, tasting, or holding a bottle of alcohol. These are the most common cues used in traditional cue reactivity studies [47]. Some of the cues are seen to elicit stronger responses, such as olfactory or gustatory cues [48].

Contextual cues Unlike proximal cues, contextual (or distal) cues are not directly linked to alcohol consumption [47]. Still, they are regularly present during consumption. Due to previous paired associations of these settings with proximal cues (i.e., conditioning), contextual cues are not dependent on the presence of proximal cues, and can as such cause cue reactivity on their own. Key factors are environments, situations, or settings in which consumption has occurred previously, especially in bars or at parties [11]. Presence of others is common in such settings, and craving is also highly affected by social contexts [7, 11].

Complex cues The combination of both proximal and contextual cues, i.e., being in an environment or a situation linked to alcohol consumption, while alcohol is also in proximity, results in a complex cue [49]. Such cues should approach real-life use situations [47]. Proximal or contextual cues can be subject to either instant or delayed gratification, but proximal cues see individuals succumbing to their feelings of craving sooner, consuming alcohol. For contextual cues, however, feelings of craving can persist to allow gratification when the situation or context allows for consumption if it previously did not [47]. Social situations by themselves are, however, unlikely to evoke strong direct positive valence and craving. Conversely, isolated proximal cues by themselves in traditional clinical trials *are* capable of eliciting feelings of craving [50]. Alcohol-related contexts, even without proximal cues, seem to be an important factor in evoking craving. As such, the study found that only showing images of alcohol, or presenting alcohol for consumption, are neither necessary nor sufficient to evoke craving, suggesting that pairing only simple stimuli with alcohol is unlikely to represent a context as a whole [50]. Other studies found a combination of alcohol cues and contextual settings to induce greater craving, also when compared with neutral settings [20, 27]. Generally, studies that included virtual agents to invoke social pressure elicited greater responses of craving than was elicited presenting alcohol cues alone, although with a greater effect in social drinkers [27, 51, 52]. Compared with clinical settings, contexts in real life are characterized by higher environmental complexity. Naturally, complex contexts are challenging to simulate in ex vivo therapy when also presenting social situations, capable of inducing greater craving or diminishing avoidance skills, than presenting simple alcohol cues alone [7, 53]. Moreover, difficulties persist in separating social craving from context craving [54].

2.2 AUD Treatment

AUD is particularly difficult to treat as it affects both the individual and the people surrounding them. Often it is difficult to admit to the problem and recognize the ailment, and it may be a difficult task for individuals to seek treatment [55]. Many AUD treatment methods and approaches have been developed with roots in both medication or therapy, such as cognitive behavioral therapy (CBT) [56–58], motivational interviewing [59], mindfulness [60, 61], as well as alternative methods, such as acupuncture [62]. Motivational interviewing is a technique that motivates patients by exploring and resolving their ambivalence as a key to change their behavior. It is usually carried out during medical consultations [63]. The goals of deploying a mindfulness approach to AUD treatment help patients gain an increased awareness of triggers related to their condition, thereby actively thinking about them and not acting upon the urges they evoke [64]. In AUD treatment, it is often seen that strategies involving a number of treatment methods are employed.

2.2.1 Cognitive Behavior Therapy

Many adapted methods for treating individuals with AUD originate in CBT [65], a system of psychotherapy based on a theory of personality and psychopathology that emphasizes cognitive processing as being the primary benefactor in the development of psychological distress [56]. Theory is built upon the connection between thoughts, emotions, and behavior and in some cases a fourth dimension of autonomous or physical reactions [44]. Several studies have demonstrated the efficacy of CBT treatment when used in the treatment of AUD and a wide range of other disorders [66–70]. The purpose of CBT is to rectify defective information processing by altering dysfunctional thoughts and assumptions that maintain self-damaging behaviors, thoughts, and emotions [56]. Therapy is based on a collaborative process between the patient and therapist, where behavioral and verbal techniques are used to examine the patient's own beliefs. Here, illogical assumptions are disputed, and the patient is equipped with coping skills and experiences that allow for a more adaptive cognitive processing [12]. Early in treatment, the focus is on evaluating the patient's assumptions related to their ailment, including distortions in logic and problematic behaviors as a result hereof [56]. The purpose is to change the patient's systematic bias in the thinking process. More concisely, beliefs are posed as a hypothesis. Then, they are logically examined and tested through behavioral experiments with the therapist. Change in behavior occurs when the patient's assumptions are modified in a way that fits more closely with the situation, where their false assumptions occur [57].

Successful CBT may only be carried out if a number of prerequisites are met [56]. First, the treatment sessions must contain a "framework" that is comprehensible to the patient and compatible with their personal belief systems. This framework must also be able to arouse affective engagement and be applicable in problematic everyday situations, such as when faced with craving in SUD contexts. Finally, it is important that established frameworks are deployed, tested, and trained in patients' everyday situations. In CBT, it is fundamental that changes in the condition of a patient may only happen if they are engaged in real problematic situations and when experiencing an *affective arousal* related to the their condition [56]. These states of affective arousal are commonly referred to as *hot cognitions*. The importance of engagement may be exemplified by contrasting depressed and phobic individuals. Here, the thoughts of a depressed individual are particularly pervasive and always present and available for examination by the therapist as opposed to patients suffering from phobia or panic disorder which are generally in little distress in contact with the therapist and may show little anxiety anywhere outside of the situation with the phobic stimulus. To facilitate changes among phobic and panic patients, anxiety must be induced either through in vivo exposure, imaginary techniques, or using provocative techniques such as hyperventilation [56]. As such, cognitions which are anxiety-related can become highly salient and accessible, opening themselves up to testing and subsequent modification. The focus on just evoking thoughts and behavior related to affect is not enough to solve the problems of the individual. Through their lives, individuals might experience cathartic episodes, providing immediate relief. However, if the underlying issues have not been addressed and proper coping mechanisms have not been developed, the individual will receive no benefit from the relief, and a similar episode might happen again [56]. In CBT, the misconstrued emotional and cognitive thoughts and behavior in a particular situation can be experienced and recognized simultaneously, helping the patient to address, disapprove, and change the dysfunctional or deviating cognition [56].

Techniques used in alcohol-related CBT often focus on either *coping* with or *reducing* craving [71]. Coping skills training (CST) tries to arm the patient with skills they can utilize in high-risk situations where craving is felt. CST can be specific to certain situations or involve general social skills to cope when one's thought process might be affected in the presence of others. Some approaches to CST treatment can include training of relapse prevention, training of social or communication skills, training of urge-specific or refusal skills, as well as mood management training [71]. CST can be more effective than comparative treatments in outcome improvements (e.g., relapse severity and frequency) of individuals with AUD [71].

A frequently researched technique aiming at *reducing* craving is *cue exposure treatment* (CET).

2.2.2 Cue Exposure Treatment

Cue exposure treatment has its roots in exposure therapy, traditionally used to directly manifest certain undesired conditioned responses in patients by exposing them to, e.g., phobia- or anxiety-related stimuli [72]. The idea of exposure therapy is for the patient to habituate through repeated exposures to stimuli, ideally conditioning the patient to reduce or prevent their undesired conditioned response, and aim towards complete reversal or extinction. Psychologically, habituation explains the decrease of a conditioned response to a repeated stimulus [14]. The exposure can help concretize the conditioned response in a specific context and be handled in combination with a specific coping strategy for that response [71]. In CET however, the stimuli are not presented directly, instead exposing the patient to proximal, contextual, or complex cues that evoke similar responses of craving [10, 49].

CET has been used as a technique in CBT to treat, e.g., phobias, OCD, PTSD, and anxiety disorders [73, 74]. It has seen limited effects in use with substance use disorders of, e.g., nicotine or alcohol [11, 14, 71, 75–78]. However, for alcoholism, CET is found to be more effective than CET for other substance use disorders [10, 14]. The potential reduction of craving in patients entails cues interfering less with patients' mental state, in turn facilitating mental surplus, enabling patients to utilize coping skills in high-risk situations [14]. Presence of alcohol cues has been indicated to trigger cognitive distortions that disrupt and impede attention, and negatively influence the ability of the AUD individual to utilize learned coping skills; this even applies to AUD individuals who are adept in using coping skills. CET could therefore be used for reducing alcohol cues' disruptive effects on the ability of individuals with AUD to use coping skills to decrease reactions through habituation or even gradually eradicate the response (i.e, extinction). Few studies have evaluated the effectiveness of CET on AUD. CET alone can result in favorable outcomes, but it seems more beneficial to utilize the potential of practicing coping skills at the same time, having CST and CET complement each other [71]. AUD individuals' coping skills are less effective compared with the coping skills of individuals with no reported AUD. AUD individuals are being trained using CST, learning coping strategies to combat high-risk situations. It secondly teaches them general social skills to improve sober relationships and challenge the positive outcome expectancies of alcohol consumption. Both of these aspects contribute to increased self-efficacy, creating a positive 'loop' of utilization and improvement of coping skills and social skills entailing increased self-efficacy [71].

In CET, the effects of classical conditioning are utilized through continued, repeated exposure to the cues of a conditioned stimulus (e.g., going to a bar that sells alcohol, a simulated high-risk situation), provoking a conditioned response of craving. Meanwhile, the patient is hindered gratification from the unconditioned stimulus of alcohol. This weakens the conditioned response of craving in the absence of such unconditioned stimuli, through the process of habituation [79]. Ideally, over time, the response of craving will lose its reinforcing properties as well as turn the conditioned stimulus of the bar back to be a neutral stimulus. The idea of reducing and eventually extinguishing craving in patients is not in itself sufficient for a successful treatment. With successful cue exposure, hot cognition situations arise, potentially paving the way for a therapist to more effectively employ the strategies found in traditional CBT [12]. Both exteroceptive stimuli (sight, smell, and look of the alcohol or the situation, context, or location in which it is presented), interoceptive stimuli (physical states, emotions, and thoughts), as well as withdrawal symptoms can turn into conditioned stimuli [44].

To evoke feelings of craving in patients during CET, different types of tools are used to convey both proximal, contextual, and complex cues [49]. Pictures and videos can be used to show relevant stimuli such as alcoholic beverages or situations involving alcohol consumption [15, 80]. Others utilize the patient's imagination by having them imagine or recall and retell high-risk situations [15]. Sometimes, role-playing is used, e.g., where the therapist invites the patient to drink alcohol [77]. Some methods involve alcohol itself, and in cases with a treatment goal of moderate drinking, patients have been prompted to consume it [10, 14, 16]. To avoid overwhelming the patient, it is recommended that craving is gradually induced by exposing the patient to increasingly intense alcohol cues [44]. An example could be presenting a proximal cue, e.g., a bottle of beer, to the patient, asking them to describe its qualities (its color, temperature, shape, etc.). Next, the patient is asked to open the bottle, describing the sound of opening it, smelling it, and describing how it feels to hold in their hands. Then the patient pours the beer into a glass, being asked to listen to the foam rising. Finally, the patient is asked to bring the glass up very close to them, while they pay attention to their own reactions and how they can increase their craving even further [44].

In CET, rather than attempting to avoid cue exposure entirely during periods of treatment, it is acknowledged that such cues are impossible to avoid in real life, and, therefore, treatment includes exposure to cues together with learning strategies for coping with the exposure and possible undesirable reactions [71, 75]. CET can then be used where direct exposure to the stimulus is not possible or feasible, if the stimulus can trigger a response that is too intense, or if the context in which the response is elicited is made up of several cue stimuli instead of one single stimulus, all contributing to the undesired conditioned response.

CET gained interest as a technique supporting CBT [10, 72, 77], as the methods of using approaches of logic alone to treat patients in traditional CBT have been debatable in terms of effectiveness [81, 82]. General CBT might target personality, anxiety, and depression disorders better when in a clinical setting, as they are mostly omnipresent in the patients' thoughts regardless of situation and context [1]. For disorders such as phobias or substance abuse, the context and situation play a significant role in whether patients experience and are affected by stimuli and cues, and might be more difficult to conduct in a clinical setting. Aside from the concept of matching thoughts to beliefs and actions in CBT, there is skepticism of how much is open to introspection and imagination, and how the semantic and declarative nature of CBT contrast the emotional and implicit nature of the patient and their actions; as such, the therapy focus must be on changing the behavior and *way of being* more directly, rather than logically challenging beliefs [81]. Here, CET has been thought to fill the gap, as strategies in CBT could be supported by the more direct, stimulus-based conditioning approaches seen in CET.

2.2.3 Is CET effective?

The success of CET can, in general, be defined as the individual's success in abstinence from alcohol or a reduction in craving when facing high-risk situations, and studies have found CET comparable to CBT and behavioral self-control training [10]. However, Kavanagh et al. found no greater effect when combining CET and CBT compared to CBT alone, and they found CET alone resulted in more alcohol consumption than CBT patients [83]. A problem in CET concerns the renewal effect. The renewal effect happens when a conditioned stimulus is paired with an unconditioned stimulus in one context, to then be extinguished in a different context. There, a return to the original context where the pairing happened will activate the renewal effect. However, indications of generalization of extinction learning in AUD patients have been shown through extinction in one context, reducing cue reactivity in another context [14, 84, 85]. Another problem, referred to as *spontaneous recovery*, happens when extinguished responses reemerge upon re-exposure to stimuli following a period after extinction learning [14, 86]. Responses to conditioned stimuli can also re-emerge after post-extinction exposure to the conditioned stimulus - this is known as *reinstatement*. Summarized, CET may fail to extinguish the most salient conditioned cues, because it does not specifically address renewal effects, spontaneous recovery, reinstatement, or cue selection [87].

CET studies have also seen inconsistencies in format, relating both to the number of therapy sessions held and the types of cues and stimuli presented. While subjective reports show that imaginary methods used in clinical CET have a positive effect on treatment, there is still a lack of tools that can perform objective measurements needed for proper assessment of such treatment [88, 89]. However, craving episodes can vary in both intensity, latency, frequency, and salience, making the study results incompatible [88]. The result is that no consistent evidence for the effectiveness of CET for AUD currently exist [10, 14]. Another important concern in evaluating CET relates to the imaginary exposure methods themselves. They are used because CET cannot portray various high-risk situations directly. However, this is also difficult because there are limitations in the person's mental ability to imagine and recall such high-risk situations [71]. Although doubt exists on whether treatment manuals can effectively transfer skills and knowledge into actual treatment and therapy, they are increasingly being used as sources for describing, researching, teaching, and practicing evidence-based psychotherapy [90, 91]. Some treatment manuals for AUD exist, also integrating coping skills training (CST), mainly based on the work by Monti and Rohsenow et. al [72, 92]. However, meta-analyses show that not all studies systematically follow the same procedures, methods, or techniques, as well as not all researchers or therapists agree on or fully know and understand the evidence-based principles and guidelines outlined in the manuals [14]. Due to the structure of the manual, it may not be at all possible to follow it rigorously in the first place. As such, differences in performed studies, e.g., theoretical approaches, sample size, and measurement methods, as well as study and follow-up duration, affect their replicability. As with the above, missing standardized procedures as well as current issues with measurability complicate work towards replicability in CBT and CET studies, as well as psychotherapy in general, and must be addressed [93–96]. Accordingly, it is important to investigate the use of tools and practices that can potentially solve the problems of efficacy through measurability and replicability, while still possibly closing the transfer gap between imaginary and real settings of high-risk situations and subsequent cue reactivity and response leading to craving and alcohol relapse [14, 97].

Patients can experience relapse and craving even when learning relapse prevention. They cite a lack of competence in controlling craving while directly exposed to cues in real life, as opposed to the techniques and skills learned in clinical settings [44]. The cause of inconsistency might also stem from how the different interventions present the complex nature of contexts and cues, as the complexity makes common high-risk situations generally difficult to produce in clinical contexts (ex vivo) [98]. Interventions teaching pre- and postcoping strategies for situational relapse have usually defaulted to utilizing imaginary exposure, i.e., through visual imagery or storytelling [44]. However, such methods fail to represent the full spectrum of situations and exposure to cues in real life, also potentially introducing patient anxiety in the process. As such, it is difficult to transfer ex vivo exposure to later real-world encounters (in vivo), which has been described as a *transfer gap* [97, 99].

As traditional exposure methods mainly use substance-related cues detached from social, environmental, and otherwise in vivo exposure, with ineffective results, research has investigated the use of immersive virtual reality (IVR) as an intervention tool to enhance or substitute traditional ex vivo cue exposure experiences [11, 14, 54, 88, 100]. To be able to effectively utilize IVR as a substitute for traditional cueing methods, it is necessary to understand the possibilities the technology offers as well as the many concepts related to its successful implementation. Therefore, the following section will introduce the concept of IVR and grant a brief overview of the technology and its capabilities.

2.3 Immersive Virtual Reality

Immersive virtual reality is a term that covers technology that simulates interactive three-dimensional environments using computer technology. Modern IVR systems immerse users in these environments by simulating human modalities [101]. Most commonly, the focus is on simulating visuals and audio as these modalities dominate humans' perception of their immediate environment [102, 103]. This approach has resulted in the development of immersive virtual reality (IVR) technologies, a consequential one of which is the head-mounted display (HMD); these oppose non-immersive technologies which are often pooled with IVR technologies in meta-analyses, affecting several outcomes [25, 104]. As such, the immersive variant used in the experiments of this study is referred to as IVR in this paper. The HMD displays a VE while simultaneously being tracked in 6DoF in both location (forward/backwards, left/right, and up/down movements) and rotation (in roll, pitch, and yaw dimensions). This facilitates a direct mapping between real life and simulated movements of the user.

Most commercially available HMDs come with built-in headphones providing audio in the VE, as well as two additional controllers (or other technologies) for hand tracking. These controllers support vibrations for simulating haptics. In this way, it is most common for an IVR system to facilitate the simulation of audio, visuals, and haptics, while tracking the user's head and hands. The system may also include additional devices to add immersion capabilities. Examples are additional 6DoF trackers for legs, arms, hip, and feet, haptic feedback devices, eye tracking, facial tracking, or devices such as *Leap Motion* [105], letting users track hand movements without the need for controllers.

IVR has seen increased use in design, training, education, and therapy [97]. Here, the immersive qualities of IVR can help induce feelings of presence in users, giving them a sense of 'being there' in the virtual world, and act as if it was real life [106, 107]. As alcohol craving is seen to be highly affected by social contexts and social pressure, creating virtual environments that facilitate immersion and presence should focus on the design of virtual agents to let the user experience virtual social situations and interaction. The goal is to achieve a perceptual illusion of social presence in IVR that feels believable to the user [108, 109].

2.3.1 Essential Concepts of Virtual Reality Experiences

To be able to effectively immerse users in VREs and thereby induce feelings of presence, it is necessary to understand the dynamics of such experiences as well as the illusions and effects the user may experience [106]. Perhaps the most fundamental of these is the concept of presence, a central factor contributing to the effectiveness of a VRE [110].

Immersion and Presence: Slater proposed the degree of immersion can be objectively measured and assessed as characteristics of the IVR medium and describe to which extent the medium can elicit sensorimotor contingencies, delivering an *inclusive*, *extensive*, *surrounding*, and *vivid* illusion of a VE to the user [25, 111]. Here, the inclusive illusion is measured by how much it excludes the real world. HMDs are more inclusive than a standard desktop monitor while blocking out external stimuli, and when the audio is provided through headphones, it may block outside real-life sound sources, while also being spatialized in VR to be able to trick the user into hearing audio actually coming from a certain direction. The extensive illusion describes how many of the sensory modalities, such as audition, vision, somatosensation, olfaction, or even gustation, are utilized in the VRE, and how accurately they are simulated. The surrounding illusion describes to which degree the IVR medium is panoramic rather than narrow. Lastly, the vivid illusion describes the resolution and fidelity, mostly concerned with visuals in terms of pixel density and refresh rate of the head-mounted display [111].

While immersion is related to the technical aspects of IVR and its equipment, the concept of presence relates to the tendency of users responding realistically to situations and events within a VE [25]. Presence was originally elucidated as the (psychological) feeling of 'being there', which in the case of IVR media refers to being in the VE, however, the deconstruction of presence postulated by Slater in [25] is adopted in this paper. Here, presence is separated into two independent components - the *place illusion*, the feeling that you are physically placed in the environment and the *plausibility illusion*, the feeling that events in that environment are really happening, such as the behavior of virtual agents or events that unfold dynamically in reaction to the user [25, 30].

When users experience high feelings of presence in IVR media, the VE becomes more engaging than the real world, feeling like a place they visit, instead of just a representation hereof. To achieve high feelings of presence, behavior in the VE should be consistent with similar events in the real world. As an example, the user should feel their virtual self-representation is actually themselves and the movement of the virtual body feels like their own movement. Presence is both subjective and objective [111]. The subjectivity relates to the user's own discernment of whether they are actually 'there' in the VE, if they think of it as 'place-like'. The objectivity relates to the observable behavior of a user in the VE - if they behave similarly to how they would behave in the real world when exposed to the same situation or event. The subjectivity is considered linked to the degree of immersion, while objectivity is considered linked to more fundamental aspects of immersion. The impact of the immersion aspects mentioned earlier is mediated through two filters [111]. First, the application context, as, e.g., an application concerned with audio should have high quality audio, whereas visuals are regarded as less important, and second, the perceptual requirements of the individual. Preferences in information through various senses tend to vary between individuals and the application needs to accommodate differently for each individual to successfully construct their internal world model. For example, some users might focus on audio more than visuals and vice versa [106]. Furthermore, the experienced scenario in a VRE is important [111]. A VE should have the story, interactions, dynamics, and events be independent of the real world in which the user can act and where there will be an autonomous response. The scenario of a VRE plays a large role in creating the interaction between the user and the VE and hence helps with removing the user from the real world. The better the VRE is at presenting an alternate self-contained world, the greater the feeling of presence as well [111].

Embodiment: With IVR comes the possibility to enhance the experience through a first-person perspective. The user can be self-represented by and control an avatar, adopting the properties of the virtual body as if it was their own body [112]. The effects of the avatar's visibility (e.g., controller, hand, or full-body representation) and anthropomorphic quality (human-like or not) have been found to be highly complex and task-dependent [113, 114]. By adopting the properties of the avatar, the resulting illusion elicited in users is defined as a sense of embodiment [115, 116]. It includes a combination of three intertwined subcomponents:

- The *sense of ownership* refers to the temporary illusion of owning and being in the virtual body, an outcome of brain processes that integrate different sensory cues into the unified perception of self-identity. The cues can be both visuospatial, visuotactile, visuomotor, or visuoproprioceptive, but they must all be synchronous to the movement of the user's actual body, and they must adhere to physical laws.

- The *sense of agency* refers to the perception that the user is an agent of the limbs and properties of the virtual body, being able to move and control them as if they were their own. These virtual limb movements and feedback must also be synchronous with the user's real limb movements, or a sensory mismatch will occur.

- The *sense of self-location* refers to the perception of actually being in the physical space and volume where the virtual body is located. It adds to, but is not the same as, the feeling of place illusion that covers the sense of oneself actually being inside a VE with or without a virtual body.

Together, these components can help the user achieve the feeling of self-presence, affecting the user's actual senses, mental model, and emotional and psychological states and traits to align with the experiences of the virtual body inside the VE [29]. However, the virtual body does not need to match the properties of the user's real body. Studies have revealed what is referred to as the *Proteus Effect*, a phenomenon in which the behavior of the user is affected by the characteristics of their presented avatar. In line with self-perception theory, it is believed to be due to individuals tending to adopt the traits and conform their behavior to what they think others expect of their embodied avatar [112, 117]. As an example, individuals assigned to taller avatars conformed to being more confident and acting more aggressively than were individuals given shorter avatars [118].

Locomotion in Virtual Environments: VR locomotion refers to the concept of users moving their viewpoint in space in a VE. Different IVR locomotion techniques have been developed [119], and can be categorized as either discrete or continuous locomotion. Discrete locomotion is when the user is instantaneously moved from one place to another, e.g, by teleporting, using point-and-click methods. Continuous locomotion can be achieved through, e.g., controller input, accelerating the user in a certain direction, real-walking, or gestures such as walking-in place or arm-swinging [120]. Real-walking is advantageous compared with other techniques in terms of eliciting increased feelings of presence [121], improved spatial knowledge, and reduced cybersickness in VEs due to its multimodal nature [122]. On the other hand, one of the shared advantages of the other locomotion techniques is when the VE is larger than the user's available physical space. Here, real-walking limits the IVR interaction space to the physical interaction space of the user, while other techniques allow unrestricted interaction space movement [120]. Redirected walking aims at combining the advantages of real-walking with unlimited virtual interaction space by redirecting users to unnoticeably walk a different physical path than their virtual path [122].

Cybersickness: Cybersickness is a prevalent phenomenon in IVR, resulting in several symptoms including nausea, disorientation, headaches, sweating, and eye strain. While there are some relations between other motion-related phenomena such as simulator sickness and motion sickness [123], there can still be differences in symptom intensity and context. Common theory proposes that cybersickness is caused by sensory mismatch in multiple channels of perception such as the proprioceptive, visual, and vestibular senses [123]. In support, several studies suggest that reducing sensory mismatch will help reduce cybersickness [124–126]. Nevertheless, cybersickness has been negatively correlated with presence, making it an essential factor to consider in IVR development [127]. The quality of the HMD and any additional equipment may decrease or increase the amount of cybersickness in individuals. Here, the addition of immersive features such as stereoscopy, or increased field-of-view of the display, may increase the risk of cybersickness, albeit with an increase in believability due to the increased likelihood of experiencing feelings of vection, the illusion of self-motion [127]. Furthermore, the design of the experience may influence experienced cybersickness. Here, it is important to design interactions, camera motion, and locomotion that feel natural and intuitive to reduce cybersickness [127].

2.3.2 Social Situations in IVR

When designing social situations in IVR, they can be studied to understand how to incorporate the concepts of complex situations when implementing these in VR. Social action theory investigates how social behavior relates to cause and effect, as well as to how the perceived meanings of a context (one's world view) shape social actions and situations [128, 129].

Analyzing the social action theory, humans can then generally perform either intrinsic or extrinsic actions that acknowledge or affect others [129]. Intrinsic actions do mostly not consider consequences when trying to achieve a goal. Extrinsic actions take into account the consequences of their actions, as well as social norms and expectations. Either way, social actions are always taken into account by understanding and assigning meaning to the actions, attitudes, and behaviors of other people [129]. As such, social action theory can help design interventions that feature social interaction. The role and behavior of agents toward the user must factor in the meaning and understanding the individual has of the social situation, affecting their internal values, which in turn affect their social actions. As such, the theory can also be used when choosing methods of measurement of the user in IVR, and what data to monitor in relation to the social encounter.

The social action theory, however, does not include the interaction itself happening after an action has been performed. Here, the theory of goal attainment explains how human perception or judgement can shape the actions of an individual [130]. It proposes the action is understood by the recipient, who then provides a *reaction*. The exchange of action and reaction are what makes an *interaction*, and continued exchanges constitutes goal-focused *transactions*. In all steps, *feedback* (or in the social action theory, *meaning* or *understanding*) is created to fuel a new perception or judgment in the individual, a looping process. Being aware that feedback is also shared and received through every step of the process, the design of the role of social situations in IVR must factor in how social relations can change with the individual depending on the interaction and transaction [131].

Virtual Agent Believability: Virtual agents may help improve the plausibility and dynamicity of a VRE if they are created with intelligent behavior, personal agendas, and in a manner that is coherent with the world they are presented in [109, 132]. Furthermore, a perceived lack of believability in virtual agents may lead to a loss of presence [133].

The difficulty is that creating a place illusion [30] is trivial compared to creating a plausibility illusion, due to the innate ability of humans accepting the presented environment as being real, contrasting with the approach employed when evaluating the plausibility of the events in that environment. Specifically, when looking at virtual agents, it is common to refer to their contribution to the plausibility illusion as being the *believability* of agents in the environment [109].

Commonly, the notion of agent believability is defined as agents that seem to be life-like, whose actions make sense, and who allow for the suspension of disbelief [109]. Other definitions include the agents' ability to convey their personality [134], and some define it by the extent to which users believe they are interacting with a sentient being [135].

Research found that virtual agents demonstrating self-awareness, awareness of their environment, and their own interaction possibilities are perceived as being significantly more believable, further arguing to their point that agent environment and self-awareness are essential components of virtual agent believability [109, 136]. This awareness is demonstrated through the dynamic and contextualized behavior of agents. Alongside the previously mentioned aspects, part of the perceived realism of agents is tied to their appearance and behavior. Here, uniqueness and feasibility is important, although artificial humanoid beings are particularly prone to suffer from the *uncanny valley* phenomenon. The concept of the uncanny valley explains the relationship between humanlikeness and believability that is nonlinear. At a point where the agents become more human-like, their believability starts to drop rather than increase [137]. When creating complex behavior for virtual agents, it is important to consider human behavior being complex and varied, especially regarding social interactions, which involves many indirect behaviors [109]. Starting from direct behavior, verbal interaction is most often done through scripted dialog, with a finite and predetermined amount of responses. Recent advances in voice recognition fidelity and availability have made it possible to make verbal interaction much more believable with the use of machine learning to determine appropriate responses depending on the situation [138]. These advances will expand upon the possibilities of verbal communication with the user. Verbal interaction can be complemented with nonverbal interaction, such as body language, eye gaze, and facial expressions. Virtual agents aligning their gaze towards the user in a lifelike manner is important to synchronize conversation between the two involved participants [109]. Realistic gaze behavior has shown significant improvement in communication between humans and virtual agents [139], including the realistic synchronization of torso and head posture during gaze shifts [140]. Gestures allow for more lively interaction between the virtual agent and the user and may increase expressivity [109]. When implementing gestures for virtual agents, it is especially important for the upper limbs of the agent to be timed correctly to the flow of the conversation [141]. Lastly, it is important to mention facial expressions as they allow for virtual agents to express their emotions, demonstrating their personality or current mental state [141].

As such, implementing complex and dynamic virtual agents can be used to create social situations in IVR that support CET experiences where complex cues are sought. They can also help in terms of evoking, e.g., social pressure or social anxiety, if needed, for training of other skills such as coping [142].

2.3.3 3D-VR versus 360°-VR

In general, an IVR environment can either be designed as an entirely 3D-generated VE, or be represented in the HMD through a stitched-together live action recorded 360°video. With a focus on social context, as well as immersion and feasibility, 360°video might be advantageous over entirely virtual environments in aspects of narrative and technical immersion [26]. Especially the visual and social complexity and coherence of the surrounding VE, actors, or narrative can be portrayed more accurately, facilitating higher illusions of plausibility [143, 144]. Although, when there are limited modalities and no interaction, the lack of dynamicity, autonomy, and active engagement might hinder higher levels of perceived believability, and hence, immersion [131, 144]. Specifically, 360°video VEs are limited in terms of facilitating valid sensorimotor actions and effectual actions [25], negatively influencing both the place and plausibility illusion. The user can act only as an observer or a passive participant with 3DoF motion, compared to what is possible in 3D-generated VEs with 6DoF motion [26, 145]. Responses can vary between experiences, e.g., favoring 360°environments according to participants' psychological response, and, conversely, favoring 3D-generated environments according to their physiological response [146]. Some research shows no significant differences, however, in positive affect, heart rate, or feelings of presence between 360°-VR and 3D-VR environments [144]. Although, the VRE in the mentioned experiment was relaxing in nature and did not seek to heighten anxiety or stress levels in the participants, but to reduce them instead. As such, further research must determine if the same holds true for arousing VREs that elicit situational anxiety or stress, of which the participant needs to apply coping skills to manage.

2.4 Using IVR for CET

To better close the transfer gap and generalize the presumed reduction in craving happening ex vivo, the context in which cues are exposed should approach an in vivo context. IVR as a treatment tool in craving reduction has the capability of utilizing more human senses than has traditional ex vivo treatment tools, enabling immersive embodied experiences that mimic high-risk situations comprising complex cues. Such an experience can make patients feel present in the environment. Here, CET using IVR can utilize virtual agents to facilitate dynamic and natural interactions in complex social situations. As such, background research has defined the aspects important to facilitate these interactions and how to create a plausible social experience. Additionally, IVR-CET can be performed in a controlled environment that more easily enables gradual exposure, while taking into account any privacy-focused and ethical issues that could arise when compared to treatment taking place in vivo.

The overarching hypotheses are that effective CET should reduce craving in individuals. Subsequently, it should also minimize cognitive load and reduce impairments by applying appropriate coping skills, avoiding distorted alcohol consumption outcome expectancies as well as decreased self-efficacy that would lead to relapse. The IVR tool itself can then be considered effective when it can convincingly mimic high-risk social situations and induce patient craving. To begin designing interventions for IVR-CET, appropriate assessment methods will be analyzed, before a related work section will present and analyze similar research studies and their designs and intervention methods, which in turn will aid in designing the IVR tool and intervention of this study.

2.5 Assessment of IVR and IVR-CET

This section outlines the different assessment methods of VREs in general, as well as IVR-CET experiences. Two main measures of presence and craving will be highlighted. The background section recommended the measurement of craving to assess the impact of the IVR-CET experience on individuals. Additionally, in determining and comparing elements of the implementation of either 3D- or 360-generated environments, there is also a need for evaluating presence.

Measuring presence is deemed important since it relates to the ability of a VE to approach real life. Presence is considered a factor in enabling, or perhaps generalizing, elicitations of feelings of craving in a VE [10]. As previously mentioned, presence can be subdivided into illusions of place and plausibility. Additionally, on the factors leading to presence, and as immersion mainly is restricted to system characteristics facilitating illusions of place, Skarbez argued for introducing a term called *coherence* [147]. Coherence would explain how participants experience illusions of plausibility through characteristics of the specific VE deemed *expected*, *reasonable*, or *believable* (see arguments by Rovira [148] and Slater [25] for further elaboration), but at the same time does not assume any goal of simulating or replicating reality.

Assessment of the 3D-VE should determine which environment and scenario parameters are important to facilitate presence regarding these three subdivisions of place, plausibility, and coherence. Where the immersion may facilitate illusions of *place*, facilitating illusions of *plausibility* will happen in immersive environments where certain contexts "feel as real life" in the VRE. Results of interventions evaluating the degree to which a 3D-VE can facilitate these illusions can then be used to design a competent experience for IVR-CET. However, there are concerns of using presence as a measurement, e.g., to show the relation between presence and effectiveness, as well as to explain what has been known as the 'book problem' of explaining why a (technically) non-immersive medium such as a book can still immerse participants fully, just by using their imagination [149]. One useful situation is in terms of pain or stress distraction, coping, or avoidance, where the higher the presence felt in a VE, the more distracted the user is from their real environment.

2.5.1 Presence

Below, four presence assessment methods are listed, which are each used in different contexts [29]. Interventions in VREs are highly subject to a break in presence (BIP), since many require the participant to mentally, or even physically, meta-'step out' of the VE context to assess their current feelings, emotions, etc. [147, 150]. A BIP can result in adverse physiological effects, disorientation, and loss of sense of control, because of the violation of expectations. As such, it might be desirable to have the measurement done inside the VE, or even done subconsciously, when the measured variables allow for it.

Self-reporting includes questionnaires, ratings, and scales [29, 97, 150]. This is currently the most popular method of assessing presence. It can be used to address both place and plausibility illusions of presence in addition to technical immersion. Presence can either be measured by the degree of experienced presence, or as the lack thereof (a presence conflict). However, despite its quantitative nature, self-reporting can introduce biases that can be difficult to address [28]. Researcher bias would consist of choosing or creating specific questionnaire items based on theory and hypotheses, instead of having participants define their feelings of presence. Subjective bias, however, would exist when the participant is tasked to meta-evaluate themselves concerning their own behavior or activity, either after they have experienced something already, or during the experience, which leads to a BIP. However, when asking participants to report when they feel a BIP, instead of reporting when they feel present, the reporting itself should not introduce additional breaks. Regardless, self-reporting methods can be easy to apply if used *outside* the VE but can conversely require a substantial implementation to apply *inside* the VE. Some game engine frameworks and assets have been devised on how to present self-reporting methods in VEs, e.g., the VRTK toolkit and VRate Unity3D [151, 152], both integrating subjective assessment questionnaires for ease of use. Although, except for, e.g., a single-item VAS presented as sliders in the VE, most questionnaires relating to the experience are presented post-experience [150]. A problem of gathering accurate measures of current or changing feelings of presence occurs if these are no longer felt outside the VE, rendering the VAS less reliable. However, research suggests only presenting one questionnaire post-experience, as scores from multiple questionnaires are seen to be highly correlated [29]. For measuring direct presence (including illusions of place), use of the System Usability Scale (SUS) questionnaire or the Bouchard et al. single-item measure has been suggested. Regarding focused assessments of the plausibility illusion, research suggests the Reality Judgment Presence Questionnaire (RJPQ) by Baños et al. to be relevant [147, 153].

Behavioral assessment includes different types of measures that relate to what is done by participants inside a VE, compared to what would be done by participants in the real world, or their behavioral responses to what is observed in the VE [29, 97]. This method tries to objectively capture the natural reactions of participants, making them generally non-intrusive, avoiding a BIP. Although most of the time, an action related to the user needs to happen in the VE to trigger a behavioral response, e.g., a ball must be thrown at them for them to catch it, something needs to fly or fall towards them to elicit an avoidance gesture, or an object is within reach for users to grab, etc. It should, however, be possible to observe behaviors without such events, although they are perhaps more sporadic, random, and not generalizable. Additionally, behavioral responses can also be social in nature, replying or responding to conditioned interactions or cues [29] by, e.g., speech, posture, or gestures. This angle can be utilized when virtual agents are present, gauging if the user perceives the agents to be real and act according to social norms when the agents behave towards the user in certain positive or negative ways. In general, behavioral assessment is very context-dependent, making them hard to generalize.

Psychophysiological assessment includes quantitative measurements of the autonomic physiological output of human activity through psychological affect (recording biosignals, e.g., ECG, EEG, EMG, EDA, and more) [29, 46, 97]. This can in practice only be utilized when there are virtual events or elements that can spark such psychophysiological responses, i.e., confrontations with something stressful or arousing [154], limiting use. Nonetheless, if users elicit a sense of presence by responding to the environment as if it was real, psychophysiological responses should be able to record exactly that. As such, BIP either caused by the environment or the researcher of the experiment can be measured [29]. EDA measurement can be used to detect and quantify general physiological arousal as well as more specific emotional arousal, but does, however, not detect emotional valence; only indicating the levels of arousal [155]. Additionally, care must be taken to ensure no restriction of movement or introduction of inconvenience will occur by wearing the equipment needed to perform the measurement. If such precautions are taken, the results gathered by physiological readings are both objective, sensitive, and reliable because it can be measured during the VRE continuously over a longer time [154]. This can be compared with, e.g., discretely measuring responses post- (or pre-) experience, as seen in other methods. The result can be reliable, but only if the equipment has been attached to the user properly, and it can measure the data reliably. Measuring changes in heart rate (HR) and heart rate variability (HRV) (using electrical ECG) is seen to be the most usable, since it is more consistent, sensitive, and reliable than other physiological recording methods, when ectopic beats and outliers have been filtered [156]. Less HRV indicates an increase in a stress response to either an perceived event or a BIP [157]. Changes in skin temperature come in second, depending on exposure duration, with changes in skin conductance (from readings of EDA through, e.g., optical PPG sensors) coming in third [154]. Heart rate or heart rate variability measurements have not been successful in measuring presence, however, but can measure anxiety levels when exposed to emotional virtual experiences, which can then correlate with feelings of presence or immersion [144]. There can, on the other hand, be confounding effects of individual characteristics [147].

Psychophysical assessment includes measuring the relations between external (physical) stimuli and the responding internal (psychic) sensation [28]. While not directly measuring presence per se (as other methods do), it measures changes in the felt presence, made by changes to the coherence or immersion of the virtual experience. Here, a technique that can be called "presence matching" has been devised from works by Slater, Skarbez, Bergström, and López et al. has been utilized [28, 30, 31, 149, 153]. After presenting a full-configured (high fidelity) VRE to participants at first, different lesser-configured (low fidelity) versions of the VE are presented, impacting place or plausibility illusions depending on if either or both immersion and coherence aspects are targeted. Afterward, participants must manipulate certain criteria until the experience matches the presence felt at the beginning. One study did not ask participants to match their previously felt presence but instead asked which parameters could make the experience better [31]. The matching method is seen as a form of *metamerism*, from the 'color matching' theory this method draws upon. Elaborated, metamerism covers the phenomenon of seeing two different colors to be similar under one light, but not under another light. A point of equilibrium can then be established for feelings of presence, where participants would respond to the VE in the same way they would in the real world, indicating that no further improvements to the VE are needed. The order of transitions, i.e., which criteria are manipulated first, can then be objectively measured and averaged over several trials and participants. This generates equivalence classes of agreement defining which criteria of a VRE, of infinite combinations of hardware and software elements, have the highest priority in eliciting presence. These can be presented in empirical probability functions [30]. The method can not, however, produce a single value for representing levels of felt presence as other assessment methods can. Conversely, single values of representing presence are missing much of the context and complexity of such a feeling, perhaps rendering single values irrelevant. Individual presence matches can be achieved in different configurations with different manipulated criteria and can be recorded and utilized as such. That is difficult, if not impossible, to record using, e.g., questionnaires, where similar responses could mean individually different things [30]. On that note, the presented parameters in the lesser-configured experience should be perceptually distinctive to match the perception of visual and behavioral realism for any participant. Although infinite combinations and distinctions together with the perceptually *never fully realistic* VREs can make results difficult to replicate in research or apply to practical IVR applications [153].

2.5.2 Craving

To measure subjective alcohol craving related to exposure to VE cues and context, studies by Ghita et al. [52, 158–160]. have identified, or used in studies themselves, visual analog scales during IVR exposure to measure subjective momentary craving and craving-related anxiety over time *during* IVR exposure (pioneered by Bordnick et al. [20].), the *Multidimensional Alcohol Craving Scale* (MACS-/MACS-VR) to measure craving intensity *before and after* IVR exposure [161], pre-post visual attention tasks (VAT) to explore gaze patterns related to attentional bias [162], as well as the *State-Trait Anxiety Inventory* (STAI), assessing the individual's level of anxiety at a particular moment (state) and in general (trait) [159]. Anxiety in this context can be seen as more related to stress or distress, with the individual exhibiting psychophysiological and behavioral responses to alcohol exposure, perhaps seeking to remove oneself from the cues that can contribute to a relapse. Objective physiological measurements of biosignals, as detailed in Section 2.5.1, can be linked to subjective feelings of craving by exposing individuals to alcohol cues and recording stress or anxiety-related arousal responses [10, 46, 78, 163, 164]. Examples of these can be increases in heart rate, skin conductance, skin temperature, as well as blood pressure and salivation. The responses can be increasing or decreasing from baseline levels depending on the measure, context, and resulting affect, e.g., when measuring heart rate versus heart rate variability, or when measuring decreases in skin temperature but increases in skin conductance. Additionally, heart rate can be affected by both stressful and 'confusing' situations, whereas measures of EDA mostly only respond to stressful situations [147]. Across literature, however, different theories currently exist on why these physiological responses occur when exposed to alcohol cues, not having reached a consensus as of yet [165].

2.5.3 Cybersickness

While cybersickness (CS) and presence are seen to be negatively correlated [127], measuring CS and sensory mismatch is still relevant in IVR applications because of shared sensorimotor processes. Measurements of CS can be done by using either self-reporting, behavioral, or psychophysiological methods. The most popular and convenient method is to measure CS via a subjective self-report questionnaire presented after the experience [127]. Questionnaires can capture a wider range of symptoms experienced with CS [123], but the subjective aspect can make subsequent analyses and comparison difficult. In addition, the delay between experiencing CS and filling out a questionnaire post-experiment can affect results. However, single-question CS prompts presented during IVR exposure can mitigate these issues, if such prompts are balanced in intervals and frequency to not experience a great BIP. Some examples of questionnaires are the general Simulator Sickness Questionnaire (SSQ), a shortened form of the SSQ called SSC (taking two questions from each SSQ component), and two questionnaires more specific to IVR called the Cybersickness Questionnaire (CSQ) and the Virtual Reality Sickness Questionnaire (VRSQ) which reuses the oculomotor and disorientation parts of the SSQ while discarding the nausea parts [123, 127, 166, 167]. The omission of nausea items (in, e.g., the VRSQ), while deemed irrelevant for VREs, can be complemented with psychophysiological assessment, if inclusion is necessary. Behaviorally, the user can respond to CS by objectively monitoring decreases in task competence or if the user terminates the experience early [127], but these can also be complex to analyze in isolation, as other factors may affect the elicited behavior. Physiologically, a user's increased respiration rate, electrical activity in the stomach, HR, or EDA can point to induced CS, among others [123, 127]. Such methods should, however, provide an objective, but complex measure of CS. As much as one wants to avoid a BIP, the added measurement hardware can feel intrusive to the user [123], suggesting to find a balance in frequency and intervals to mitigate a BIP. However, experiments of introducing combinations of psychophysiological assessment and short questionnaire prompts during a VRE have been conducted, successfully removing delays of questionnaire measurement resulting in enhanced reliability of results [127]. Interestingly, when measuring for feelings of presence and/or CS, many factors of the VRE can overlap in measurement, especially when monitoring stress responses [127]. I.e., the stress response to a VRE where the user balances on a plank high above the ground, a physiological stress response can be elicited by the user, indicating feelings of presence and/or CS. The relation between CS and presence can therefore be complex and even intertwined when affected by the immersive factors of vection, navigation control, and hardware displays [127].

Summarized, as all assessment methods have benefits and drawbacks, triangulations using multiple methods are desired to interpret and correlate results. If the results correlate across triangulated measures, there is a strong indication of causality where the dependent variable is changed by a manipulation, subsequently proved by statistical significance. If the results do not correlate, the causality can not be proved sufficiently.

The section on presence measures can again highlight some benefits and drawbacks of the methods chosen to obtain data, especially relating to the BIP-inducing self-report scales during IVR exposure (also compared to behavioral or psychophysiological assessment), or to the pre-post questionnaires not measuring momentary craving or anxiety.

3 Related Work

3.1 Related Work on CET and IVR-CET for AUD

This section presents related studies and work using CET and IVR-CET for individuals with AUD. This includes proposed implementations, assessment methods, and results. First, presentations of other related studies testing the efficacy of CET and IVR-CET for treatments of AUD are performed, and second, studies are presented that examine the difference in effects of alcohol cues between individuals with varying degrees of alcohol consumption.

3.1.1 Efficacy of CET and IVR-CET

Lee et al. [168] then did an experiment with IVR-CET on eight individuals with AUD for eight sessions. Cues and contexts were selected based on a survey, resulting in the creation of an oriental pub and a western bar. Measurements of alcohol craving, urge, and related thoughts were established through the Penn Alcohol Craving Scale (PACS), the Alcohol Urge Questionnaire (AUQ), and the Obsessive Compulsive Drinking Scale (OCDS). A repeated-measures analysis of variance indicated no significant difference between pre-treatment and post-treatment scores, although, when excluding people who had abstained from alcohol for more than a year, the AUQ results were significant.

Bordnick et al. [20] also constructed five VEs to assess IVR-CET. One of them was a control environment, created as a neutral cue environment devoid of alcohol cues with a present scent of vanilla. The remaining four environments were all alcohol-related environments and consisted of a typical hotel bar environment, a kitchen environment being prepared for a party, a typical party environment, and a home office environment with two people drinking and arguing, trying to draw the user into the disagreement. Scents of pizza, coffee, and different alcoholic beverages were present in the alcohol craving inducing environments. 40 individuals with AUD were then exposed to the environments and assessed for craving and level of presence. High levels of presence were reported by participants, indicating the environments to be presumably realistic and compelling. The craving measured in the home office environment was also significantly higher than in the neutral environment. However, the bar, kitchen, and party environments all resulted in significantly higher feelings of craving than in the home office environment. Their meta-analysis revealed that cue reactivity effect sizes were much smaller for alcohol than for other substances, but the authors found effect sizes thrice as large. They propose the complexity of cue presentation (VR versus traditional cue presentation methods) may account for these differences. Regarding the home office environment resulting in significantly lower craving than the other alcohol craving inducing environments, the authors speculate that it might be due to comparative incongruence.

Related, Cho et al. [7] investigated the influence of social pressure on alcohol craving. To do so, they developed eight conditions; two non-alcohol-related settings (a street and an office) and two alcohol-related settings (a bar and a restaurant) with and without virtual agents providing social pressure. For conditions without virtual agents, the alcohol-related settings induced significantly more craving than did the non-alcohol-related settings. However, for conditions *with* virtual agents providing social pressure, there was no significant difference between alcohol and non-alcohol related settings. They conclude that to induce alcohol craving, using a virtual agent is more important than using alcohol. They speculate that it is due to social pressure situations producing more stress or negative emotions.

While these studies show mixed results of IVR-CET, it is important to note that by the writing of this paper, they are dated by more than a decade. This is important when considering research into and developments of CET but even more important regarding VR as a rapidly developing immersive technology.

Therefore, more recently, a study done by Ghiţă et al. [158] compared the efficacy of IVR-CET with CBT for individuals with AUD. The IVR-CET consisted of exposure to alcohol-related cues (22 different alcoholic beverages) and contexts (4 environments: restaurant, bar, pub, and home), while the CBT consisted of classical standardized therapy for the treatment of addiction. In both sessions, the participants completed several assessments of alcohol craving such as AUDIT, MACS, MACRS-VR, STAI, and VAS. Their results found that both sessions lowered all scores of craving and anxiety in the participants, but IVR-CET seems to be slightly better. They therefore propose IVR-based CET as a complement to existing treatment methods for AUD individuals. Ghiţă et al. [160] later demonstrated that six IVR-CET sessions over a cause of five weeks effectively reduced anxiety, craving, and attentional bias towards alcohol cues. However, only one individual, diagnosed with severe AUD, took part in the study.

Hernández-Serrano et al. [33] also did a study comparing classical standardized treatment with classical standardized treatment complemented by IVR-CET. Changes in alcohol craving were measured with the MACS and their results found a significantly greater improvement in alcohol craving levels when complementing classical treatment with IVR-CET than using classical treatment alone. They propose that including IVR-CET in classical standardized programs may benefit in treating individuals with AUD, especially among individuals with an intense craving for alcohol.

Both of these findings therefore show a significant increase in the efficacy of IVR-CET in their results, especially when included in existing treatment methods. This is assumed to be due to technological advancements pushing the capabilities of VR further.

3.1.2 Cue Reactivity Effect on Individuals With Varying Degrees of Alcohol Consumption

Lee et al. [27] followed up on the study done by Cho et al. in the previous section using the same IVR system to carry out a similar experiment, with the hypothesis of AUD individuals demonstrating different patterns of social pressure-induced craving according to the presence of alcohol cues than would a control group comprising generally healthy individuals. Fourteen AUD individuals and fourteen age-matched healthy individuals took part in the experiment. In a two-by-two factorial design, four conditions of the study were established, comprising IVR environments with or without alcohol cues and with or without social pressure. For both groups of participants, social pressure significantly induced craving in the environment without alcohol cues. However, in the environment with alcohol cues, social pressure did not result in additional increases in craving for the AUD individuals group, but did so for the healthy group. The authors refer to the ceiling effect as a possible explanation for these findings, an effect relevant to craving assessment. Here, AUD individuals who already rated their craving level as very high in the alcohol cue environment without social pressure, might have hindered the registration of potential increases in craving caused by subsequent social pressure. Additionally, it is possible that cognitive and neuro-chemical reactions to alcohol cues could have interfered with the AUD individuals' ability to assess their own internal state. Limitations of the single-item visual analog scale in terms of assessing cue reactivity should therefore be considered. Cravings induced by direct alcohol cues or by social pressure might be related to different craving components. Nevertheless, the findings of the study indicate that AUD individuals are vulnerable to environments with alcohol cues more than are individuals with no AUD.

Following the study done by Bordnick et al. in the above section, Ryan et al. [51] investigated differences in craving levels between binge drinkers and non-binge drinkers using the same IVR environments. The experiment exposed both groups to all IVR environments, and the results saw binge drinkers reporting significantly higher alcohol cravings and thoughts of alcohol than non-binge drinkers, albeit not in all IVR environments. There was no significant difference in the neutral environments. In line with differences between individuals' drinking habits and their feelings of craving, Deok-Yong Kim and Jang-Han Lee [53] found that, contrary to light social drinkers, heavy social drinkers also took longer to move away from alcohol-related situations compared with non-alcohol-related situations in an approach–avoidance task.

Simon et al. [169] also conducted a similar study immersing heavy drinkers and occasional drinkers in a virtual bar with alcoholic beverages. Heavy drinkers reported significantly higher levels of craving than occasional drinkers after exposure. They also discovered the alcohol craving after exposure was significantly related to the level of perceived ecological validity of the VE and the perceived ecological validity more strongly increased craving in heavy drinkers. Simon et al. therefore state that perceived ecological validity is an important experimental parameter to study craving and virtual reality can be a useful tool for studying alcohol addiction as well as treating it.

3.1.3 Summary

There is a clear tendency for IVR-CET to produce significantly more craving than classical standardized treatment, especially when looking at more recent studies and when including IVR-CET in existing treatment methods. As part of IVR-CET, environments have been shown to be an important aspect in inducing an adequate amount of alcohol craving, with environments such as bars and parties resulting in higher cravings than e.g., a home office environment. There is also a tendency for heavy drinkers or individuals diagnosed with AUD reporting significantly higher alcohol craving than occasional drinkers or healthy individuals. It was found that social pressure significantly increased induced craving in an environment without proximal alcohol cues. However, in an environment *with* proximal alcohol cues, additional social pressure did not increase craving for AUD individuals, but did so for healthy individuals.

4 Methods

With IVR technology showing great applicability potential in CET for individuals with AUD, two experiments were carried out in this study, both investigating how to best utilize the technology.

First, performing CET in high-risk situations involving social contexts is crucial, as such contexts often constitute a central element in inducing craving. Here, social pressure challenges and interferes with AUD individuals' ability to apply learned coping- and refusal skills. However, to simulate a high-risk situation involving social aspects, the VE must feel plausible to the user. For these reasons, the first experiment (*Experiment A*) focused on discovering which specific elements of a 3D-rendered VE contributed most to experienced plausibility in that environment.

Secondly, also with the goal of simulating high-risk situations involving social aspects, it is of interest to investigate which of the two

Table 1: **Changeable experimental parameters of plausibility.**

Parameter		Level and Description
Agent animations (**pAA**)	**0**	no animation
	1	basic lipsync and body movement
	2	+ advanced lipsync, eye and body movement
Agent reactions (**pAR**)	**0**	no reaction to user action
	1	reaction to user action)
Soundscape richness (**pSR**)	**0**	no sound
	1	music and talking
	2	+ ambient bar sounds, outside sounds
Geometric realism (**pGR**)	**0**	low level of detail
	1	medium level of detail
	2	high level of detail

general methods for constructing VEs (360°-video and 3D rendering) would potentially be most fitting in the context of IVR-based CET for AUD individuals. Thus, the second experiment (*Experiment B*) compared 360°-video generated VEs against 3D-generated VEs with a primary focus on inducing alcohol craving in participants.

Sections 4.1 and 4.2 present the two experiments, followed by section 4.3, detailing the design of the virtual bar used in the experiments.

4.1 Experiment A: Plausibility

4.1.1 Experiment Design

Experiment A used a within-subjects design where participants experienced a virtual bar in five trials of randomly chosen parameters of environment configurations, letting them change aspects of the environment to match what felt real to them. The experiment design drew inspiration from the presence matching methods used in similar studies [28, 30, 31, 149, 153]. The general premise was to split a VE into four different parameters adjustable by their level of fidelity - this is explained in detail in the next section. Here, participants were first exposed to the VE with all parameters set to the highest level of fidelity. Then, all parameters were set to a starting configuration featuring the lowest possible level of environment fidelity of each parameter, and participants were then tasked with adjusting the parameters until their perception of the current environment matched their perception of the first environment in terms of plausibility. In total, the participants completed five trials of the parameter matching task, with up to seven possible parameter changes for the participant to perform in a single trial. There were no minimums for the number of changes, i.e., participants could declare a match at the starting configuration. Based on the *plausibility* and *virtual agent believability* research presented in Section 2.3, it was hypothesized that parameters related to virtual agent believability (pAA and pAR - see below) would be deemed more important than the others in terms of facilitating plausibility.

Parameters: In this study, the different adjustable aspects of the virtual bar are referred to as *parameters*. The parameters cover fidelity levels of different aspects of the VE, focusing on audio, visuals, and agent believability. There were four different parameters comprising *Agent animations* (pAA), *Agent reactions* (pAR), *Soundscape richness* (pSR), and *Geometric realism* (pGR). These are described by the vector *S = <pAA, pAR, pSR, pGR>*. Table 1 provides an overview of the different parameters. The following describes the levels of each parameter, ranging from the lowest to the highest level of fidelity:

Agent animations (pAA):
[Inanimate, level 0]: In their idle state, the virtual agents are not animated and hence are completely motionless.
[Basic animations, level 1]: In their idle state, the virtual agents have basic binary lip synchronization when talking, meaning their mouth movement is limited to two states of being either fully open or completely closed. Their body moves realistically in the context (sitting, drinking, talking, etc.).
[Advanced animations, level 2]: In their idle state, a more advanced lip synchronization based on visemes is used for the virtual agents as well as eye blinking.

Agent reactions (pAR):
[Non-reacting, level 0]: The virtual agents do not react to any user actions.
[Reacting, level 1]: Reactions from virtual agents are enabled - see section 4.3 for a detailed description of these reactions. Note that enabling reactions overrides both *pAA* and *pSR*, meaning virtual agent animations and sounds will play regardless of the current level of *pAA* and *pSR* if a reaction is triggered.

Soundscape richness (pSR):
[Silence, level 0]: No sounds are present.
[Essential sounds, level 1]: Music is playing through the speakers inside the bar and virtual agents are talking.
[Rich soundscape, level 2]: In addition to music and talking virtual agents, sounds typically related to a bar are present (e.g, glasses clinking, beer being poured, and noise from electronic appliances). City sounds and passing cars from outside the bar are also present.

Geometric realism (pGR):
[Low, level 0]: Objects and virtual agents in the bar are presented using textures 1/8th the resolution of the original size (1024x1024). Agents have their skin weight count reduced to one per bone, and the environment is only illuminated with ambient lighting. Furthermore, the image resolution is downscaled by 50 percent.

[Medium, level 1]: Objects and virtual agents in the bar are presented using textures half the resolution of the original size (1024x1024). Agents have their skin weight count set to two per bone, and the environment is illuminated with ambient lighting and four point lights. The image is no longer downscaled.
[High, level 2]: Objects and virtual agents in the bar are presented using the original textures (1024x1024). Agents have their skin weight count set to unlimited, and the environment is illuminated using baked light maps from 21 scenes placed around the virtual environment. The image is no longer downscaled.
Fig. 5 shows the lowest and highest fidelity geometric realism.

4.1.2 Participants

25 participants (n_M=16 (66%), n_F=9 (33%)) were sampled from the university campus as well as through social media. They were split into age groups of 18-25 (n=19), 26-35 (n=4), and 56+ (n=2), SD=38.30, with no participants in the age group 36-55. The experiment was approved by the ethics board of the university, participants did not know about the experiment procedure and outcome beforehand, and there was no compensation for participation. In total, there were around 25% with no experience, around 10% with little experience, and 65% participants with much gaming experience. For IVR experience, around 45% had no prior experience, around 25% with little experience, and 30% with much experience. In the age groups, the 56+ had no experience in both gaming and VR. The 26-25 had all much experience in 3D-gaming, with their experience in IVR being evenly spread out. The 18-25 had four with no experience (n_M=1, n_F=3, 21%), two with little experience (n_M=1, n_F=1, 11%), and 13 participants with much experience in gaming (n_M=10, n_F=3, 69%).

4.1.3 Apparatus

The experiment was conducted by two researchers in the Multisensory Experience Laboratory at Aalborg University in Copenhagen. One researcher (INSTRUCTOR) was in charge of instructing and guiding the participants throughout the experiment, and the other researcher (OPERATOR) controlled the computer running the VRE. See Fig. 1 for an illustration of the three main areas of the experiment.

A table and a chair were located in the center of the room, matching the height of the seat and table presented in the virtual bar. The participant was seated with their back facing the *Main PC* running the VRE, on which researchers could monitor participants while inside the virtual bar. Before being exposed to the VRE, participants signed a written consent form and answered a pre-experience questionnaire, filling out demographic information and prior experience with both video games in 3D-generated worlds and VR. The written consent form and questionnaire are shown in Appendix A and Appendix B, respectively.

During the briefing, participants were shown a video demonstrating the VE with all parameters set to the lowest fidelity level, after which each parameter is upgraded once in the following order: *pAA*, *pAR*, *pSR*, *pGR*. Both screen-recordings and a simultaneous live action video of one of the researchers performing the upgrades were shown to demonstrate how the participants were supposed to do the upgrades themselves.

As for the VR setup, the experience was presented using an HTC Vive Pro IVR set and peripherals (A Vive Pro HMD with a display resolution of 2880x1600/615PPI, a refresh rate of 90Hz, and a 110°FOV, as well as accompanying Vive controllers and base stations for tracking). Instead of using the built-in headphones of the HMD, the Sennheiser HD 600 external over-ear headphones were used to achieve a higher fidelity audio experience.

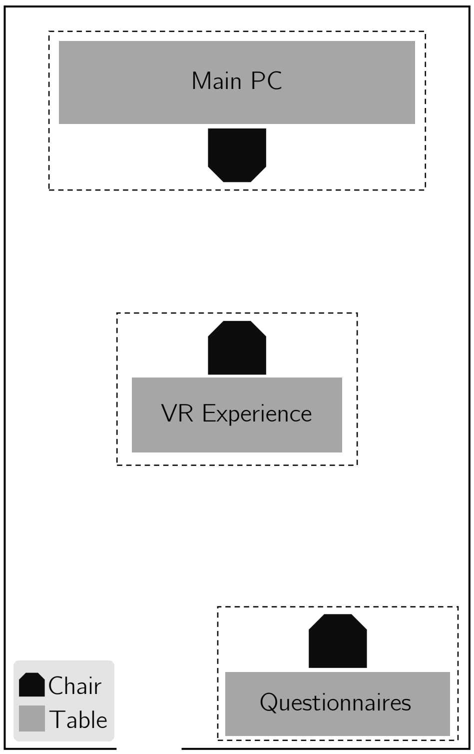


Figure 1: *An illustration of the setup used in both Experiment A and Experiment B comprising the Main PC area in which OPERATOR stayed throughout the experiment, and the Questionnaires area for participants to fill out the consent form and questionnaire before they entered the VR Experience.*

A custom user interface (UI) was developed for progressing through the experiment. The UI would display the trial numbers and current configurations, allowing the operator to transition between conditions by fading in and out visuals and audio seamlessly. The UI can be seen in Appendix G.

Measures:

- Cybersickness was assessed post-VR by presenting the 9-item VRSQ to the participants, assessing oculomotor symptoms and disorientation. The questionnaire was presented post-experiment to not cognitively overload participants or introduce new BIPs during the experiment, as the IVR tasks could already feel complicated. To minimize the bias caused by the delay between the VRE and subsequent cybersickness assessment, however, the questionnaire was presented immediately after the experience.

- Additionally, users could comment on which VRE elements they noticed the most, as well as which elements they found to be most important to include in such a VRE.

4.1.4 Procedure

The procedure for the experiment is presented step-by-step in the list below. An extended version of the procedure, including both a Danish and an English version, can be seen in Appendix C.

- Step 1: The participant enters the laboratory and is greeted by both INSTRUCTOR and OPERATOR. INSTRUCTOR explains the role of each researcher and then explains the goal of the study.

- Step 2: The participant is asked to sign the consent form and answer the pre-experiment questionnaire on the laptop in the *Questionnaires* area in Fig. 1.

- Step 3: After answering the questionnaire, INSTRUCTOR introduces the four adjustable parameters by showing the demonstration video to the participant on the laptop.

- Step 4: INSTRUCTOR guides the participant to the table in the *VR Experience* area in Fig. 1, and the participant is told the parameters can be upgraded by the participant saying the name of the desired parameter, and providing a brief explanation as to why they wish to upgrade that specific parameter.

- Step 5: The participant wears the HMD and headphones and is immersed inside the virtual bar with all parameters set to the highest level of fidelity. Here, INSTRUCTOR guides the participant on how to use the controllers and HMD for interacting with objects and looking around.

- Step 6: Upon familiarization with the IVR environment, INSTRUCTOR explains the matching task to the participant.
  - First, all parameters are assigned the highest level of fidelity, i.e., the parameter configuration $<2,1,2,2>$. This is referred to as the reference environment. Here, participants are asked to spend two minutes paying attention to and remembering how real they feel the VE is, as they will later be tasked to match that same sense of reality.
  - Then the parameters are reconfigured to either of the following five starting configurations: $<0,0,0,0>$, $<1,0,0,0>$, $<0,1,0,0>$, $<0,0,1,0>$, or $<0,0,0,1>$.
  - The participant is then tasked with increasing the level of the parameter they feel would contribute the most to match or reach the same sense of reality as in the reference environment. Every time they increase the level of a parameter, INSTRUCTOR will ask if they feel the current environment matches the feeling of realness they had in the reference environment. If not, then the participant is asked to again increase the level of the parameter they feel will contribute most to reaching a match. This continues until they eventually reach the highest level, where they cannot progress any further. They are free to increase the level of the same parameter multiple times in a row, but can only change one parameter at a time in a transition.
  - OPERATOR will change the environment according to the answer of the participant, fading out the current configuration, selecting the chosen transition, and fading in the new configuration. If the participant forgets which parameters there are to choose from during the experiment, a grabbable clipboard sitting on the table in front of them in the virtual bar environment will list the four parameters. Note that if the participant wants to perform a certain transition, but INSTRUCTOR is strongly convinced that the participant is choosing the wrong parameter for the desired transition, INSTRUCTOR will gently suggest the correct parameter for the participant to choose. E.g, if the participant says "I wish to upgrade visuals because the people are not moving.", then INSTRUCTOR could say "You might be referring to animations, then. Do you think so too?".

- To counterbalance the experiment, this matching task is repeated for a total of five trials, one for each starting configuration. Moreover, the order of starting configurations is randomized for each participant. These measures mitigate the order and sequence effects and strengthen the internal validity of the experiment.

- Step 7: Once the five trials have been completed, the participant takes off the technical equipment and is asked to complete the cybersickness questionnaire on the laptop in the *Questionnaires* area in Fig. 1.

- Step 8: After the participant has completed the questionnaire, INSTRUCTOR and OPERATOR receive and respond to any eventual questions or comments from the participant, and then thank the participant for taking part in the experiment. Afterwards, the operator cleaned the table, questionnaire laptop, and the used equipment with disinfection wipes, and additionally used the *CleanBox CX1* UVC light decontamination device to fully disinfect the HMD, all in accordance with COVID-19 regulations enforced by the university, and proposed through research [170].

To avoid participants increasing parameter levels at random until they reached the highest level for each, two steps of the procedure were explained differently to them. First, instead of telling the parameters in the reference environment were assigned the highest level of fidelity, they were told that all parameters were assigned a random level of fidelity. Second, instead of telling about the five possible starting configurations, participants were told the parameters for each trial were configured randomly. These changes were made to counter the possible statistical dependence that could be found between trials if every trial started from the same configuration, enticing participants to repeat the same transitions from the previous trial.

4.2 Experiment B: Craving in 3D-VR versus 360-VR

4.2.1 Experiment Design

Experiment B used a within-subjects design, where participants experienced four short independent scenarios in a IVR-bar, while researchers assessed their experienced craving, PI, Psi, and CS. Each scenario was set up in both a fully 3D-generated environment and in a 360-video recording of the actual bar, resulting in a total of eight scenarios, presented to the participants in random order. Hence, the two bar conditions of the experiment are the 3D-generated environment and the 360-video recording, which will henceforth be referred to as the *3D* condition and the *360* condition, respectively. An additional *neutral* condition was designed to be presented between each of the two bar conditions with the intent of using it to establish a baseline for the four assessments as well as to allow participants to familiarize themselves with being in an IVR and using VR controls before being exposed to the bar. To compare the bar conditions against each other, the experience in both conditions followed the same design, including environment, narrative, and soundscape. Proximal, contextual, and complex alcohol cues were placed in the environment as objects presented on the table in front of the participant or the objects found around the bar, as well as the social context of having friends and bar patrons around the participant. Each of the eight scenarios was designed to induce one of the basic emotions of anger, disgust, anxiety, and happiness, and with an inclusive narrative where the participant was placed amongst three other individuals. The scenarios were designed based on review recommendations for evoking certain emotional states [171] in participants, and in collaboration with a clinician from Rigshospitalet in Copenhagen specializing in AUD treatment using CET.

A description of the scenarios can be seen in Table 2.

The immediate purposes of using these scenarios were twofold; (1) to promote the highest amount of craving by creating a stressful social situation with many alcohol cues in the vicinity, facilitating an emotional social reaction, (2) to increase the external validity of the experiment results.

Table 2: **An overview of scenes of different emotional situations.**

	Scenes
Happiness	Friends praise the participant as a great friend, expressing appreciation and giving compliments.
Anxiety	Friends expect the participant to give a great speech after listening to the speech of the other friend.
Anger	Outrage from a friend, caused by a lack of social awareness from the other friend.
Disgust	A friend vomiting at the table, as a result of an elaborate description of a gross bathroom.

To avoid physiological baseline readings being biased in terms of being too sensitive to positive physiological changes, these were obtained using a *vanilla* baseline technique [172], where participants were transitioned to a *neutral* VE at the beginning of the VR part of the experiment, but also between each scenario and at the end of the VR part of the experiment. Participants sat on the chair at a table where they reported to which extent they felt craving, presence (through either a place or plausibility illusion-related question), as well as cybersickness. The neutral environment was presented between 3D and 360 bar conditions, designed to be devoid of any alcohol-related cues, albeit still controlling one's virtual hands to establish proper vanilla baseline readings.

4.2.2 Participants

26 participants (15 males, 10 females, one non-binary, in the age groups 18-25 (n=14), 26-35 (n=11), (36-45) n=1, SD=30.83, were sampled from a combination of social media posts and students at Aalborg University Copenhagen. 12 had tried VR never, once, or twice, nine had a few times (4-10), and five had many times (+10). All were alcohol consumers, with six consuming less than one standard unit per week in the last six months, 17 consuming between 1-6, and 3 consuming 7 or more.

4.2.3 Apparatus

Experiment B was carried out in the Multisensory Experience Laboratory at Aalborg University Copenhagen by two researchers. One researcher (INSTRUCTOR) was in charge of instructing and guiding the participants throughout the experiment, as well as monitoring the recording of data, while the other researcher (OPERATOR) controlled the PC running the VRE, observing participant behavior in each scenario once a physiological baseline had been obtained. A table and a chair were located in the center of the room, matching the height of the seat and table presented in the virtual bar. The participant was seated with their back facing the Main PC running the VRE. Before being exposed to the VRE, participants signed a written consent form and answered a pre-experience questionnaire, filling out demographic information and their experience with VR, as well as drinking habits. The questionnaire is shown in Appendix F.

For the VR setup, the HTC Vive Pro IVR set was used, comprising the Vive Pro HMD with a display resolution of 2880x1600/615PPI, a refresh rate of 90Hz, and a 110°FOV, as well as accompanying Vive controllers and base stations for tracking. Instead of using the built-in headphones of the HMD, the Urbanears Pampas external over-ear headphones were used to achieve a higher fidelity audio experience.

A custom UI was developed allowing OPERATOR to switch to the next scene after the physiological measures reached baseline. The UI can be seen in Appendix G.

Measures:

- Momentary craving, feelings of plausibility and place illusion, as well as cybersickness were assessed with a single-item VAS presented in VR in front of the participant after each scene exposure. An example of the VAS UI is seen in Fig. 7.

  The *craving VAS* was formulated as follows:
  "On a scale from 0 (not at all) to 100 (more than ever), how much do you want to drink alcohol right now?"

The plausibility and place illusion assessment questions were formulated based on the two instructions given to participants in a study by Slater et. al. [28], where results show that participants instructed to *"Pay attention to [their] feeling that [they] are in that room that [they] can see"* deem aspects of a VE that relate to *place illusion* more important than aspects relating to *plausibility illusion*. On the other hand, participants instructed to *"Pay attention to how real this feels"* deem the aspects of a VE that relate to *plausibility illusion* more important than the aspects relating to place illusion. The *plausibility VAS* was therefore formulated as follows:
"On a scale from 0 to 100, how real did it [the scene the participant has just experienced before the question] feel".
And the *place illusion VAS* was therefore formulated as follows:
"On a scale from 0 to 100, how much did you feel that you were in the environment that you just saw?"

The *cybersickness VAS* was formulated as follows:
"On a scale from 0 (not at all) to 100 (more than ever), how much discomfort (fatigue/disorientation) do you feel right now?"

The single-item CS VAS was chosen due to its ability to be presented in VR between bar conditions, instead of presenting participants with a lengthy questionnaire following the end of VR exposure. This allowed for gathering specific CS responses in both conditions separately, as well as in the moment, not just assessing the experience as a whole.

- Through the use of the CE medical-grade Empatica E4 wristband [173], physiological signals were obtained [174]. The wristband technology was chosen because of its versatility, portability, and unobtrusiveness, even though it lacks in stability and accuracy compared to more established pieces of technology utilizing other sensors or measurement locations and techniques [175]. Baselines were declared obtained when readings across measures were steady over a period of ten seconds.
  - Electrodermal activity (EDA) was sampled exosomatically with DC and constant current at 4 Hz with a 1-digit 900 picosiemens resolution, to reveal changes in skin conductance response (SCR) [174, 176, 177]. Computed SCR data from EDA readings should show short *phasic activations* of responses of arousal to short-term events occurring in the presence of discrete environmental stimuli and cues.
  - Photoplethysmography (PPG) sensor data was sampled at 64 Hz to obtain the blood volume pulse (BVP), as well as the interbeat interval (IBI) at a 1/64 second resolution, where non-prototypical beats were discarded and IBI sequences were not smoothed using the built-in algorithm, to reveal changes in heart rate and heart rate variability [174]. Gaps in the IBI data when BVP signal was unreliable were identifiable when samples $T = T2 - T1 \neq IBI(T2)$. Data is used to calculate heart rate (HR) and heart rate variability (HRV), of which the former should increase, and the latter should decrease when stress or arousal increases in the participant.
  - Optical skin temperature readings were sampled at 4 Hz and a $0.02°C$ resolution to reveal temperature changes [174], which should see a slight increase.
  - Additional readings from a three-axis accelerometer sensor were sampled at 32 Hz and an 8-bit resolution, used to discard extreme data points caused by, to the experiment, unrelated movement or responses by the participant, distorting readings [174, 176].

Distinctions between baseline readings and individual trial readings were marked by the built-in event marking function [174]. Data was monitored by INSTRUCTOR for accuracy during the experiment by the Bluetooth streaming function displaying readings on a connected smartphone screen. Especially, subsequent analyses looked closely at the data gathered from the wristband when the participants were under potential emotional or cognitive stress, as measures of these can be less accurate [176].

4.2.4 Procedure

The procedure for the experiment is presented step-by-step in the list below. An extended version of the procedure, including both a Danish and an English version, can be seen in Appendix E.

- Step 1: The participant enters the laboratory and is greeted by both INSTRUCTOR and OPERATOR. INSTRUCTOR explains the role of each researcher.
- Step 2: The participant is asked to sign the consent form and answer the pre-experiment questionnaire on the laptop in the *Questionnaires* area in Fig. 1.
- Step 3: After answering the questionnaire, INSTRUCTOR answers potential questions and then guides the participant to the table, and assists the participant in equipping the Empatica E4 on their non-dominant hand, explaining the importance of not moving that hand/arm too much during the experiment. They are also informed the physical *Event Marker* button on the E4 wristband will be pressed by INSTRUCTOR after each scene.
- Step 4: The participant wears the IVR-HMD and headphones, after which the experience begins. The participant is first placed in the neutral condition, where INSTRUCTOR monitors the physiological signals. Once a baseline reading is obtained, INSTRUCTOR signals OPERATOR, who will then progress the VRE to the *VAS* scene, where the participant reports their craving level, their PI or PSI level (chosen randomly for each participant), and finally their experienced cybersickness. From here, the participant experiences one of the eight *bar* scenes, after which they answer the three questions in the *VAS* scene again. This cycle is repeated 8 times until the participant has experienced all *bar* scenes - however, after experiencing the final *bar* scene, the participant does not return to the *neutral* scene, but instead takes off the VR equipment and the Empatica E4 wristband.
- Step 5: Finally, the participant is asked the following question, which is recorded by a microphone: "As you probably noticed, some of the bar scenes were constructed by video recordings and the other scenes were constructed entirely by virtual objects. Do you prefer one version over the other, and if so, why?"

4.3 Virtual Environments Design and Implementation

This section outlines the general design of the virtual environments, as well as the changes made to tailor the virtual environments to the two experiments of the study.

4.3.1 3D-generated Environments

The 3D-generated VEs were developed in *Unity (v. 2020.2.1)* [178] with supplementary primary plugins: *OpenVR (v. 1.11)* SDK for VR support [179], Google Resonance SDK (v. 1.2.1) to implement spatial audio, occlusion, and reverb utilizing SADIE HRTFs [180], as well as *Glycon3D*, *LipSync Pro (v. 1.531)* and *Humelo LAVSTAR (v.1.0)* for capturing and recording custom virtual agent motion capture data and generate mouth movement [181–183]. A flowchart illustrating the entire implementation pipeline of the three aspects of the VE categorized as *virtual agents*, *geometry*, and *audio*, as well as the tools and plugins used are shown in Fig. 2:

The final design of the bar and the neutral environment is shown in Fig. 3.

The virtual bar was designed to look and feel as realistic as possible to approach a true-to-life experience. To achieve this, the real bar *Tagensborg Bodega* in *Nørrebro, Copenhagen* was used as a reference for modelling the virtual bar and the objects within it. On-location photographs, measurements, and audio recordings of the bar were used as references for implementation. The view of the street from inside the

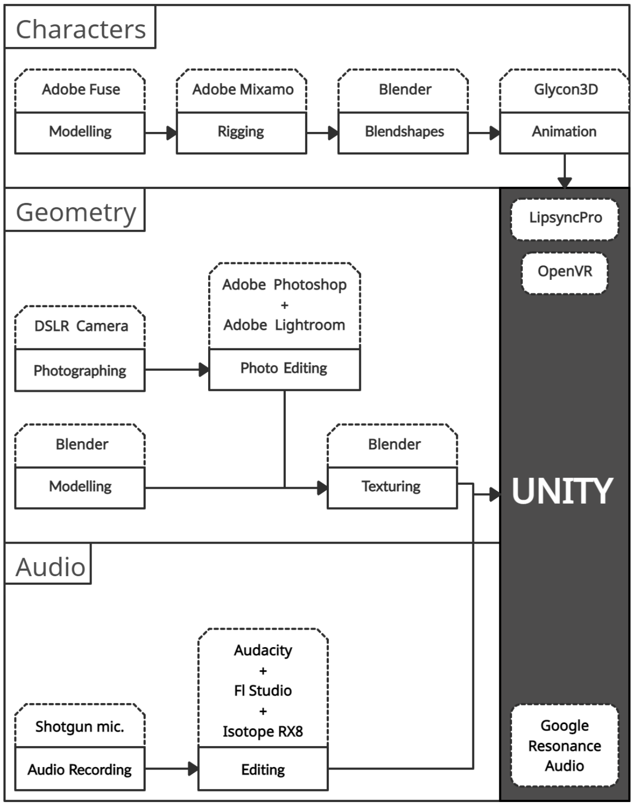


Figure 2: Flow chart illustrating the workflow in implementing the three aspects of the VEs (virtual agents, geometry, and audio) into Unity. Each aspect is split into its own section with arrows indicating the direction of the workflow. The dashed lined boxes indicate tools, and the solid boxes below indicate what the tools have been used for. Plugins used in Unity is shown with a dashed lined box inside the 'Unity' section.

bar was created using a 360-degree stitched photo from Google Street View wrapped around the building. The bar is populated with animated virtual agents and a realistic soundscape. The virtual agents in the experience are animated and sit at tables around the bar while drinking, talking, and watching television. The soundscape includes sounds such as music from hanging speakers and patrons conversing with each other at the bar, as well as sounds from appliances and other hardware in the bar, beer bottles being opened, and outside traffic. Objects in the immediate vicinity such as bottles, ashtrays, and drink menus can be grabbed and rotated, prompting a collision sound when dropped. The bottles pour liquid if they are tipped beyond the horizontal threshold.

The neutral environment was designed to not induce craving in participants, and was therefore representing a forest containing no alcohol-related cues, with birds singing and including an otherwise calm soundscape.

4.3.2 360-video Environment

For the 360-video condition, manuscripts and film shooting plans were written for the production of four two-minute scenarios, each inducing one of the basic emotions of anger, disgust, anxiety, or happiness, covering the whole valence spectrum. The scenarios featured gradual increases of emotion throughout the scenario, ending with an emotional climax. Actors were recruited, and after signing a talent release form, they together performed each written scenario at the real bar in Copenhagen, Tagensborg Bodega in Nørrebro. They were instructed to address or gaze at the camera, like it was the user sitting in its position. The camera matched the point-of-view of an individual sitting at the

Figure 3: The virtual bar from the front and back, as well as the neutral environment.

roundtable in the bar. For the production phase, the independent recordings of the 360°recording device were mapped and stitched together, and in Unity, the video was mapped to a sphere around the user to facilitate the 360-video effect. The device features directional microphones, so the audio would also be spatialized.

4.4 Experiment A Design Changes

From the general design of the 3D-generated bar described above, in Section 4.3.1, some changes were made to accommodate the design of Experiment A.

The user sits at one of the tables with the wall to their back, looking out at the bar and the main seating area with patrons. Looking to the right, the user can see the back half of the room, seating more patrons.

Five virtual agents can react to the user if the user gazes at them or at an object in the bar, triggering a scripted, animated, and voiced response by using raycasts and a predefined physics layer where the interactable objects are stored.

An overview of the reaction events and how these can be triggered by the user can be seen in Fig. 4.

Trigger	Reaction Event
Looking at the TV on the wall	Patron watching soccer on the TV looks at user and asks about the game
Looking at the fan in the ceiling	Patron watching soccer on the TV looks at user and comments on the fan
Looking at the patrons at the big round table	Patron 1 notices user and nods
Looking at the patrons at the big round table	Patron 2 notices user and asks if they need anything
Looking at the patrons at the big round table	Patron 3 notices user, but looks away again
Looking at the slot machines across the room	Friend comments on the slot machines
Looking at any of the paintings on the wall	Friend comments on the paintings on the wall
Tilting the beer more than 45 degrees or knocking it over entirely	Friend is surprised and tells the user that he might have had enough
Grabbing the drink menu	Friend asks about the drink menu

Figure 4: Table of reactions possible for the user to trigger where the left column shows what actions trigger the associated reaction in the right column.

As explained in Section 4.1, for Experiment A, fidelity levels of the chosen parameters were adjustable, which was of course also accommodated for in the design of Experiment A.

4.5 Experiment B Design Changes

In Experiment B, the user would experience two different VEs, the bar environment (in both 3D and 360 conditions) as well as a neutral environment. The virtual 3D-bar environment in Experiment B was as much as possible an exact replica of the 360-video bar environment. To accurately replicate these scenarios, dialogue from the 360-video recordings were sampled and cut from the actors to be individually placed and spatialized at each location of their three virtual agent counterparts in the 3D environment. A comparison of the two environments can be seen in Fig. 6. The scenarios portrayed in the bar scenes were as described earlier in Section 4.2.

A: The lowest level of fidelity, with decreased virtual camera resolution and texture resolution.

B: The highest level of fidelity, with highest virtual camera resolution of the environment and textures.

Figure 5: Environment changes for Experiment B showing the lowest (A) and highest (B) fidelity level of the geometric realism parameter.

Two deliberate differences were however made to the 3D-generated bar; (1) the user had tracked virtual hands and (2) proximal cue objects in the form of bottles of beer and ashtrays, matching those found in the real bar were interactable and could as such be grabbed and manipulated by the user.

The user was seated in the middle of the virtual bar at a large round table, with three virtual agents acting as the user's friends. The bar was otherwise populated with proximal and contextual alcohol cues and virtual agents in the background to match the 360-video scenarios, also adding to the social pressure context.

Finally, the VAS for reporting experienced craving, PI or Psi, and CS was represented in VR, the design of which is seen in Fig. 7. The value was incremented or decremented by pressing either the right or the left arrow, respectively. To submit the current value, the 'OK' button could be pressed.

5 Results

This section first presents results from Experiment A followed by results from Experiment B.

5.1 Experiment A Results

indicat

5.1.1 Overview

Below, we outline the measures and methods used to process data from the experiment, before showing the accompanying results. The section reports on the following measures:

- Step-wise probability distribution after each transition

The 360-video environment.

The 3D-generated virtual bar.

Figure 6: A visual comparison between the two conditions of the 3D-generated and 360-video environments used in Experiment B.

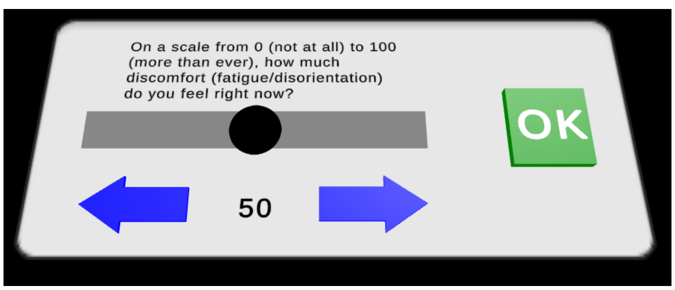


Figure 7: One of the three VAS as presented during the experience (cybersickness). The buttons are diegetic (physically in the virtual world) and provide visual and haptic feedback.

- Configurations declared as a match as well as their related probabilities
- Probability of having reached and passed a configuration that is otherwise declared a match in an earlier trial, or by another participant
- Marginal probabilities of how much each of the levels of the four parameters contributed to the declared match configurations
- Finally, cybersickness results as well as subjective results from the questionnaire.

From a total of 125 trials with 451 observed transitions (n=25), on average, participants made 3.61 ± 1.11 of up to seven possible transitions over five trials before declaring a match, depending on the starting configuration. Determined by an unpaired t-test for the two gender groups and one-way ANOVA for the independent three-sample groups of age, gaming, and VR experience, there are no statistically significant differences in transition averages in the related groups of:

- Gender (Male N=16, M=3.51, SD=0.70, Female N=9, M=3.78, SD=1.41) (t(23)=0.64, p=0.5278)
- Age (F(2,122)=1.83, p=0.16)
- Gaming experience (F(2,122)=1.02, p=0.36)
- IVR experience (F(2,122)=2.87, p=0.06)

Interestingly, there were only one trial ending with a match in the full configuration $< 2,1,2,2 >$, even starting from configuration $< 0,0,0,0 >$, making all seven transitions. No trials starting from one of the other starting configurations *<pAA+1 | pAR+1 | pSR+1 | pGR+1>* ended in the full configuration, making only five transitions at most, or six transitions starting from configuration $< 0,0,0,0 >$.

5.1.2 Transition Probabilities

The 54 possible configurations ($3_{pAA} \times 2_{pAR} \times 3_{pSR} \times 3_{pGR}$) a participant could experience are denoted by S. The set of all possible transitions between configurations is thus a subset of $S \times S$. Each transition consists of a configuration the participant was in at time t and a configuration transitioned to at time $t+1$ written in the form $[pAA_t, pAR_t, pSR_t, pGR_t] \to [pAA_{t+1}, pAR_{t+1}, pSR_{t+1}, pGR_{t+1}]$. From this, a Markov transition matrix P is generated, comprising probabilities p_{ij} that a participant in configuration $i \in S$ would step-wise perform a transition k to arrive at configuration $j \in S$, not taking previously reached configurations into account. Being a 54×54 matrix, P is clearly a very sparse matrix as it contains only 74 nonzero entries. It should be noted the total number of possible configurations to reach is not 54×54 but 135, given the restrictions of allowed transitions described in Section 4. Given P, the probability distribution over the configurations of the k^{th} transition can be computed. That is, after transition k, the probabilities of that transition resulting in various configurations can be computed. Table 3 shows the highest probabilities of transition $k = 1,2,\ldots,6,7$ resulting in configuration $< pAA, pAR, pSR, pGR >$ - only probabilities greater than 10% are included for ease of reading. Transition k_7 (configuration $< 2,1,2,2 >$) is absorbing in nature because it is the final possible configuration, as transitions are only allowed in one forward direction, and it is therefore omitted from the table.

Exemplified, it can be seen from the data that transition k_1 (with any starting configuration) most likely results in configuration $< 0,0,1,0 >$ with a probability of 34%, followed by the second most likely configuration being $< 1,0,0,0 >$ with a probability of 24%, and so on. Likewise, the transition k_2 will most likely result in configuration $< 1,0,1,0 >$ with a probability of 37%, followed by the second most likely configuration being $< 0,1,1,0 >$ with a probability of 18%, and so on.

5.1.3 Match Probabilities

Fig. 8 shows three aspects of match configurations; (1) in blue bars, it shows the percentage of total match configurations a given configuration makes up, and (2) in orange bars, the probability distribution over matches being declared when reaching a certain configuration $< pAA, pAR, pSR, pGR >$ - that is, for a participant experiencing a configuration $< pAA, pAR, pSR, pGR >$, the table in Fig. 8 show the probability of the participant declaring that configuration a match - and (3) in gray bars, the total number of times a configuration was reached by participants. Configuration $< 2,1,2,2 >$ is absorbing in nature as it is the final possible configuration.

5.1.4 Parameter-specific Probabilities

The probabilities of individual parameters and their fidelity levels contained in a match configuration are also computed and shown in Table 4. Exemplified, the probability that a match configuration contains the parameter *pAA* at fidelity level 0 is 6%, and the probability that a match configuration contains the parameter *pAA* at fidelity level 1 is 76%. The results show configurations at fidelity level 1 to contain the vast majority of probable match configurations, with many participants not reaching fidelity level 2. However, almost all participants upgraded all parameter levels from 0 to 1, with the exception of 25% not upgrading the *pGR* parameter. Table 5 shows an extended view of match

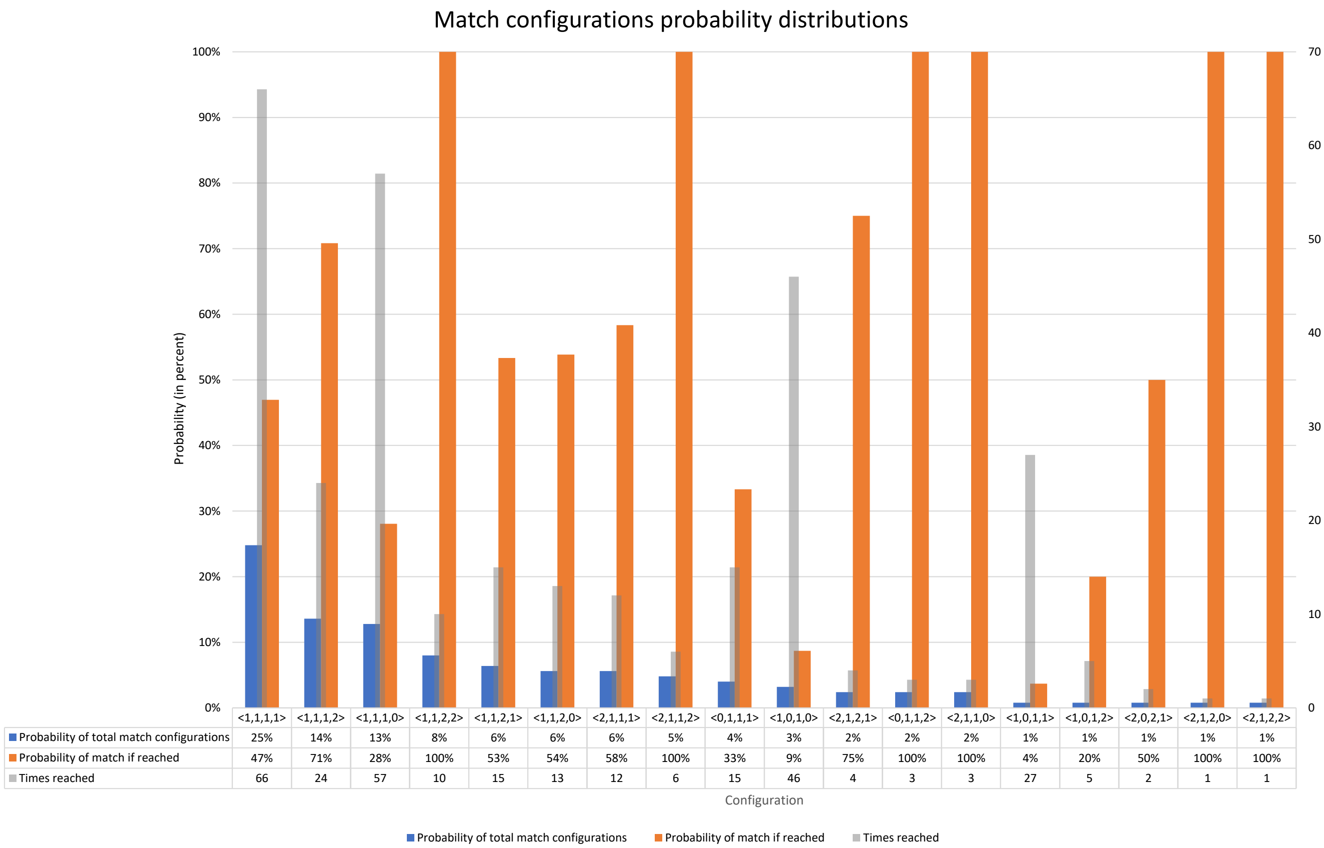


Figure 8: The blue bars represent how much configurations constitute (in percent) of the total number of match configurations. The orange bars represent (in percent) the times a configuration was declared a match if this configuration was reached. The gray bars indicate (as a count) the total number of times a given configuration was reached.

configuration probabilities divided into the different noted groups of demographics of gender, gaming experience, and VR experience.

5.1.5 Cybersickness

Results of the VRSQ are presented to reveal potential problems with and bias due to cybersickness. The total average across all participants is 0.09, and the total VRSQ scores for each demographic are presented in Table 6.

5.1.6 Post-experiment questionnaire

In the post-experiment questionnaire, participants could comment on elements of the VE they noticed the most, and which elements they felt most important to include. Through keyword coding and analysis, Fig. 9 shows the results of these questions related to the parameters (pAA, pAR, pSR, pGR) of the experiment. The soundscape richness parameter pSR was revealed to be most important and also most noticeable by participants, with the agent animations parameter pAA coming in second. Animation reactions pAR and geometric realism pGR are in third and fourth place, switching order between the questions. Some responses could not be explicitly categorized in the four parameters, i.e., "atmosphere", "interaction", and "hands", and have instead been added separately. Appendix D lists the raw data with gender color-coding.

5.2 Experiment B Results

Triangulation of the results is done for validation across multiple assessment methods, as well as to analyze different dimensions of the same output.

5.2.1 VAS Results

To compare the VAS results, Fig. 10 shows box plots of VAS scores of each condition for both craving, PI, Psi, and CS.

To test for significant differences between conditions, a one-way analysis of variance (ANOVA) was used to compare the effect of the 360, 3D, and neutral conditions on the dependent variables of alcohol craving, PI, Psi, and CS. The assumption of sphericity was assessed using *Mauchly's test* and *John, Nagao, and Sugiura's test*, and the assumption of normality was evaluated with the *Shapiro-Wilk test*, all with a significance level of .05.

Most of the samples did not meet the assumption of being normally distributed, however, *Mauchly's test* evaluated both craving (p=.24), PI (p=.32), and Psi (p=.73) to meet the assumption of sphericity, whilst *John, Nagao, and Sugiura's test* evaluated all to violate the sphericity assumption. Taking into account the relatively large sample sizes, and parametric tests' general robustness against non-normality, a one-way repeated measures ANOVA was carried out on the craving, PI, and Psi variables. *Huynh–Feldt correction* was used due to the disagreement in results between *Mauchly's* and *John, Nagao, and Sugiura's test*. For multiple-comparison post-hoc analysis, *Tukey's* test was used. Since the CS variable violated both the assumption of normality and sphericity in both tests, the non-parametric *Friedman test* was used. Here, for multiple-comparison post-hoc analysis, the *Nemenyi test* was used.

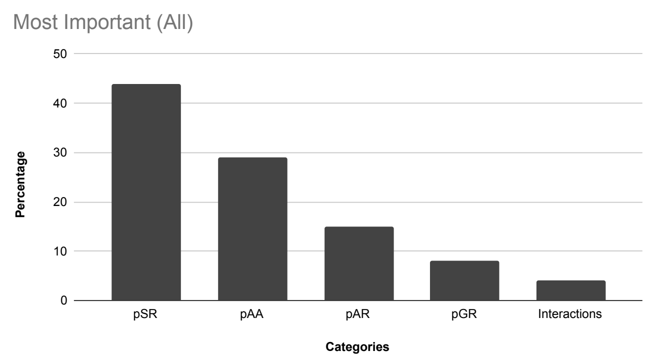
For both ANOVAs, effect sizes are calculated with the *Partial* η^2 formula. Here, η^2 values of 0.01, 0.06, and 0.14 indicate small, medium, and large effects, respectively. Finally, respective means are also present from the table. Results from both ANOVAs and the *Friedman test* as well as the multiple comparisons from the ANOVA are shown in Fig. 11.

To discover potential tendencies in VAS scores of craving, PI, Psi, and CS as a result of the duration spent in VR for participants, Fig. 12 shows plotted VAS scores with respect to the number of scenes participants have experienced.

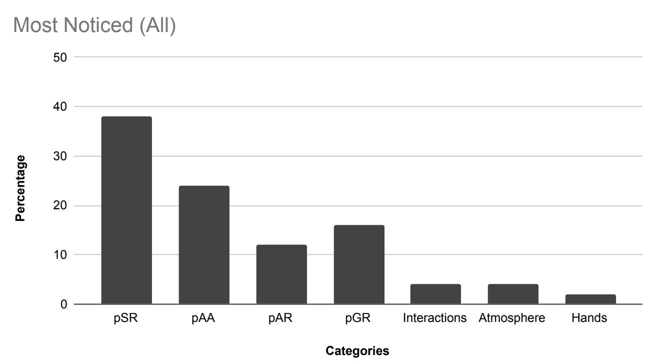
Finally, Fig. 13 shows VAS measurements and their relation to each

Table 3: **Probability distributions for each transition (only probabilities greater than 10% are included for ease of reading). The probability distribution over the first transition is split into two rows, the top one representing probabilities with the starting configuration $< 0,0,0,0 >$, and the bottom one with any starting configuration.**

Transition (k)	**Config (<pAA, pAR, pSR, pGR>)**	**Probability (percent)**
1 starting config <0,0,0,0>	<0,0,1,0>	50%
	<1,0,0,0>	25%
	<0,0,0,1>	16%
1 any starting config	<0,0,1,0>	34%
	<1,0,0,0>	24%
	<0,1,0,0>	22%
	<0,0,0,1>	20%
2	<1,0,1,0>	37%
	<0,1,1,0>	18%
	<1,1,0,0>	15%
	<1,0,0,1>	14%
	<0,0,1,1>	11%
3	<1,1,1,0>	47%
	<1,0,1,1>	22%
	<0,1,1,1>	12%
4	<1,1,1,1>	67%
	<1,1,2,0>	13%
5	<1,1,1,2>	44%
	<1,1,2,1>	28%
	<2,1,1,1>	22%
6	<1,1,2,2>	50%
	<2,1,1,2>	30%
	<2,1,2,1>	20%



A



B

Figure 9: **A:** parameters deemed most important to include in the VE for participants to feel presence. **B:** parameters noticed the most by participants in the experiment.

Table 4: **Probabilities (in percent) of match configurations containing parameters at given fidelity levels.**

	pAA	pAR	pSR	pGR
Fidelity level 0	6%	6%	0%	25%
Fidelity level 1	76%	94%	75%	45%
Fidelity level 2	18%	-	25%	30%

Table 5: **Probabilities (in percent) of match configurations containing parameters at given fidelity levels for genders, prior experience with video games in 3D generated worlds, and prior experience with VR.**

PARA-METER	ALL POOLED	GENDER FEMALE	GENDER MALE	GAMING NONE	GAMING MUCH	VR NONE	VR MUCH
pAA 0	6%	29%	4%	17%	4%	9%	5%
pAA 1	76%	47%	81%	70%	76%	73%	75%
pAA 2	18%	24%	15%	13%	20%	18%	20%
pAR 0	6%	7%	8%	13%	2%	9%	0%
pAR 1	94%	93%	93%	87%	98%	91%	100%
pSR 0	0%	0%	0%	0%	0%	0%	0%
pSR 1	75%	73%	73%	97%	68%	82%	80%
pSR 2	25%	27%	28%	3%	32%	18%	20%
pGR 0	25%	24%	28%	17%	26%	36%	25%
pGR 1	45%	56%	49%	37%	48%	35%	38%
pGR 2	30%	20%	23%	47%	26%	29%	38%

Table 6: **Table presenting the total VRSQ score for each group categorized by demographics as well as prior experience with 3D video games and VR.**

Gender	
F	M
.09	.10

Age Group		
18-25	26-35	56+
.10	.05	.10

3D Experience			VR Experience		
0-3	4-10	10+	0-3	4-10	10+
.12	.05	.89	.87	.10	0.12

other, evaluated using a two-tailed Pearson's correlation coefficient.

5.2.2 Physiological data

The physiological responses were recorded to assess whether or not participants produced a change in biosignals due to increases or decreases in feelings of craving. Data was downloaded from the E4 Connect to be processed, detailed below.

EDA - Skin Conductance Response: [174]: SCR (*phasic activation*) was computed using the continuous analysis function (CDA S-Deconv) with default settings and optimization using the LedaLab toolbox in MATLAB, detecting phasic driver activity within response windows of a stimulus, revealing emotional arousal or stress to stimuli [177, 184, 185]. SCR peak amplitude thresholds can be in a range between 0.05 and 0.5, but were ultimately set at 0.2 microsiemens (μS) to avoid false positives [175]. Peaks found when the *Event Marker* was pressed on the E4 has been removed, save for 5 seconds at the end of every scene because of SCR onset latency. The peaks are binarized to show occurrences but not specific amplitudes, as the effect of these are disputed. Peaks are computed in means of peaks per minute (PPM) for individual conditions, scenarios, and demographic groups as per the pre-experience questionnaire (see Appendix F), shown in Fig. 14.

A slight increase was observed in SCR PPM changes throughout the duration of the experiment as a result of exposure, as seen in Fig. 15.

PPG and BVP - Heart Rate and Heart Rate Variability [174]: IBI data were recorded by the Empatica E4 device by measuring the intervals between BVP peaks. HR is computed by finding the peaks per minute (PPM)/beats per minute (BPM) from peak values of IBI data = 60/IBI, averaged over Bar scenes in both conditions and the neutral condition, as well as demographic groups. HRV is computed by taking the HR BPM and using the following formula together with the smoothed Root Mean Square of the Successive Differences (RMSSD) values between adjacent R-R peak intervals (the time elapsed between two successive R-waves of the QRS heart signal) [186].

HR and HRV responses revealed no significant differences between exposure to the three conditions, seen in Fig. 16. Results were tested for normality using the *Shapiro-Wilk* test and for sphericity using

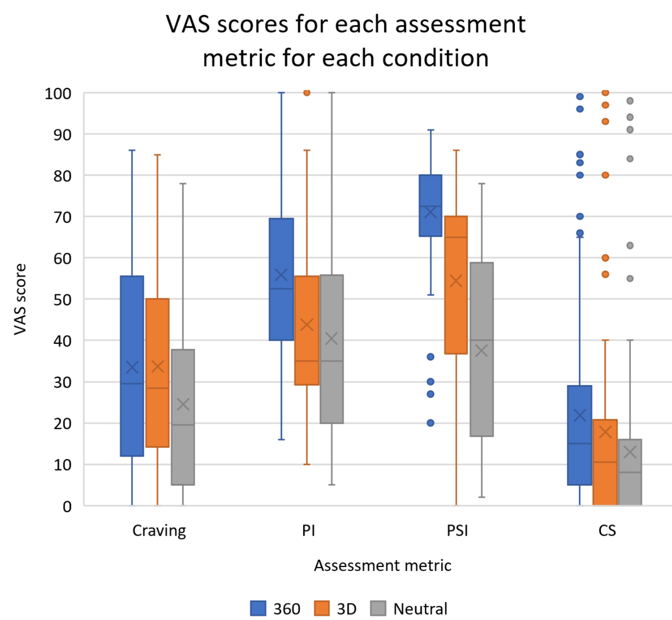


Figure 10: Box plots of VAS scores of each condition for both craving, PI, Psi, and CS.

Mauchly's test, and a subsequent t-test between the 3D and 360-video conditions revealed no statistical significance between group means in either HR (t(25)=-1.17, p=.250) or HRV (t(25)=-2.02, p=.054) responses. A one-way repeated measures ANOVA included the neutral condition in addition to 3D and 360-video scenes, finding no statistical significance between group means as well in either HR (F(2,75)=1.23, p=.300) and HRV (F(2,75)=2.59, p=.085) responses.

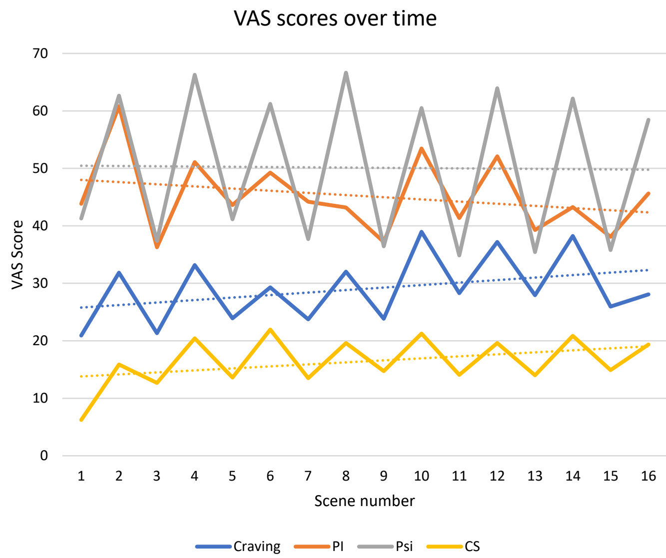
An important note relates to to the heterogeneity of sampled participants, making HRV hard to draw definitive conclusions from as metrics can vary, based on differences in personal characteristics from participant to participant [147]. As such, HRV results in Fig. 17 compares the experiment conditions with respect to four demographic groups based on similar characteristics and a sample size ≥ 3.

Skin Temperature [174]: Skin temperature is measured directly by the device. Since skin temperature changes very slowly, the presented individual scenes were not of a long enough duration to facilitate specific noticeable changes in skin temperature. Instead, a mean of participant skin temperatures over the duration of the experiment is taken (N=26, M=33.00, SD=1.68). Fig. 18 shows values over the duration of the experiment, revealing a slight increase trend in temperature.

		One-way ANOVA			Friedman
		Craving	PI	Psi	CS
Test statistic		F = 13.7	F = 8.38	F = 19.1	Q = 12.7
p-value		p = .001	p = .003	p < .001	p < .001
Effect size		strong η2 = 0.36	strong η2 = 0.46	strong η2 = 0.60	-
Multiple compa-risons	360 3D	p = .995	p = .017	p = .014	p = .317
	360 Neutral	p < .001	p = .003	p < .001	p = .002
	3D Neutral	p < .001	p = .686	p = .012	p = .047

Figure 11: ANOVA and *Friedman test* results with effect sizes and significant difference comparisons between all conditions. Green p values indicate significant difference between the compared conditions.

A



B

Average placement in array of scenes				
Scenario				
Happiness	Anxiety	Anger	Disgust	Neutral
8.84	8.96	8.12	10.08	8
	Condition			
	360	3D	Neutral	
	9.28	8.72	8	

Figure 12: A: VAS scores of craving, PI, Psi, and CS with respect to the number of scenes participants have experienced. *Scene number* on the horizontal axis refers to the number of scenes experienced by the participant, and the vertical axis shows the *VAS Score*. B: The numbers in the Table represent the average placement in the array of all 16 scenes experienced by the participant for each scenario and condition.

5.2.3 Post-Experiment Questions and Comments

The post-experiment comments and responses were mixed, with ten participants preferring the 3D environment, nine preferring the 360 environment, and seven with no preference. However, some general patterns in responses and following comments were seen.

The actors in the 360 scenes were perceived to be visually more realistic in comparison to the 3D scenes causing some participants to feel more engaged with the narrative. Several participants pointed out facial expression, eye contact, and body language as the determining factor in this feeling, as these would help the participant to determine the emotions of the actors and the situation they were placed in. This had some participants feel the experience in the 360 scenes was much more realistic, with some pointing out that it was *"intense"* and *"intimidating"* as opposed to the 3D scenes which were seen as more *"fun"* and *"like a video game"*. By comparison, the virtual agents in the 3D scenes were seen by some participants as slightly uncanny and with less accurate movements. A few participants pointed out that higher fidelity in terms of the models, textures, and animation, could help close the gap in terms of realism between the two versions.

Some participants felt that interaction was lacking in the 360 conditions. It was stated that interaction, including the virtual hands presented to them, was an important factor to be able to feel *"engaged in the environment"*. The contrast between the 360 scenes, with no interaction, and the 3D scenes, with some interactions, made some participants state they were willing to forgive the visual inferiority as long as they were able to interact with objects in the environment. The fact that participants were allowed to manipulate objects in the environment also lead to some of the comments about the 3D scenes being *"more fun"* compared to the 360 scenes, which they stated caused them to focus more on the interaction instead of the narrative unfolding. It is worth noting that some of the participants stating preferences towards the 3D version belonged to the group with much VR experience.

Lastly, two reasons for preferring the 3D version over the 360 version were related not to positive aspects of the 3D version, but instead issues with the 360 version. The first being that the 360 video had an insufficient resolution and distortions in the video caused by stitching seams, meaning they were unable to clearly distinguish objects in the background. The second and most frequent reason was they felt the table presented in front of them would not move as they would expect when they turned their head. Lastly, it was pointed out by some that they did not feel like they were part of the group of friends (actors) by the table due to their placement inside the VE in relation to the table.

6 Discussion

This section will discuss results and bias starting with Experiment A, followed by Experiment B. A section will afterwards briefly discuss additional observations considering both experiments.

6.1 Experiment A: Plausibility

The First Transition: In Table 3, looking at the first transition, the results where the starting configuration is $<0,0,0,0>$ are especially of interest. It is observed the first transition participants perform results in the configuration $<0,0,1,0>$ in most trials (50%), making sounds audible in the VE. This is in line with participants' self-reports and explanations for choosing the pSR parameter as it is, to most participants, the very first aspect they notice is missing inside the environment.

Tendency of Only Upgrading Each Parameter Once: There is a tendency in participants starting each trial by upgrading each parameter once - this is seen in transition k_4 in Table 3. However, pGR is not upgraded in 25% of cases, seen in Table 4. This can possibly be explained by the design of the parameter levels, in which the starting level is intended to be of extremely low fidelity, the next level is designed with the intention of being at an acceptable or above acceptable level of fidelity, and the final level adds only subtle details. Here, pGR_0 might not be of an adequately low fidelity compared to the other parameters, which was also indicated by a few participants stating they could not differentiate between the levels of pGR fidelity, possibly explaining why pGR was not upgraded in 25% of cases.

Most Frequent Order of Upgraded Parameters: The order in which the four parameters are upgraded is most frequently the following: pSR, pAA, pAR, and finally pGR. This is seen in transition k_2 and k_3 in Table 3. The fact that pAA is upgraded prior to pAR was also

	Craving	PI	Psi	CS
Craving	1			
PI	r = -.232 T = -3.15 p = .002	1		
Psi	r = .496 T = 9.11 p < .001	*	1	
CS	r = -.035 T = -.962 p = .489	r = -.047 T = -.625 p = .533	r = .250 T = 4.11 p < .001	1

Figure 13: Correlation between Craving, PI, PSI and CS across all conditions. Box color indicates r statistic.

apparent during the experiment, where it was observed that whenever a new trial was initiated, a lot of participants tended to remain passive, merely observing the VE until they had performed a couple of parameter upgrades. Only then, they begun to interact with the VE, e.g., trying to provoke reactions from the virtual agents.

Tendency to Overlook Subtle Details: As it can be seen in Table 4 there is a tendency for participants to not notice the subtle details in pAA_2, pSR_2, and pGR_2, with match probabilities of 18%, 25%, and 30%, respectively. The tendency to overlook subtle details also shows in transition k_5 and k_6 in Table 3 and might explain why configuration $< 1, 1, 1, 1 >$ was chosen the most with 25% as seen in Fig. 8. The slightly higher probability for a participant to upgrade pGR_2 and pSR_2 might be due to those parameters being more straightforward to assess, compared with pAA_2. That is, pGR and pSR are constantly available for assessment by the participant, whereas pAA only allows assessment sporadically, i.e., when virtual agents blink and talk.

The qualitative assessment of parameter importance and parameter noticeability, presented post-experiment, did reveal differences in the soundscape (pSR) to be both noticed the most, as well as deemed most important to upgrade in the experiment. The change between pSR fidelity level 0 (*no sound* and 1 *bar talk without background sound effects*) was retrospectively found to be too much of a difference compared with the other parameter intervals. The change in levels should, e.g., rather have been including changes of the spatial audio implementation, to still comply with the goal of assessing the fidelity of the combined environment elements, suggesting future experimentation in this regard. The implementation did, however, not allow for such a change to be performed smoothly as to be used in the experiment. The second-most noticed (and deemed-most-important-to-include) parameter of agent animations (pAA) also had a fair-sized interval, compared to the rest of the parameters (animation reactions pAR and especially geometric realism pGR). The parameter pGR would perhaps be ranked much higher, if the somewhat subjective decision on where to draw the line between implementations of fidelity levels were different. For agent reactions (pAR), fidelity levels could be made the same way as seen in the pSR parameter of introducing more sounds, instead adjusting how many agents would react to the participant. In the end, the parameters cannot elude subjective decisions of intervals needed for this type of study to be conducted, warranting future studies with different intervals in fidelity to see if the same results appear.

The two steps of the procedure for the experiment which included not being truthful to the participants about the parameter configurations in an attempt at avoiding participants increasing parameter levels at random until they reached the highest level for each parameter seemingly worked as intended. This is apparent by the highest level configuration being reached only by one participant.

Parameter Level Comparability: As previously mentioned, the fidelity levels of each parameter were constructed to be as close to equal as possible. However, each parameter might still not translate correctly to each other, which might influence results towards the parameters with most obvious changes to the environment. For example, pGR might not have had an adequately noticeable difference between each level, which might lead to participants to not notice and therefore not upgrade this parameter as much as others. This could very well affect all parameters and might also explain why pSR was upgraded the most as the first level for this parameter is the complete absence of it. While pAA_0 and pAR_0 also features the complete absence of their respective elements, the virtual agents themselves, which these parameters affect, are still situated in the environment. Additionally, the pSR parameter is different from the remaining three parameters by changes to it influencing quantity rather than quality in the entire VE. Inspired by [30], a different approach could have been to make the parameter influence auralization in terms of simulating sound reflections of more and more accurately sized rooms. Reflections that exactly match those that a room of the virtual room's dimensions would produce. Finally, pAR only have two fidelity levels, compared with the other parameters of three levels. This further leads to imbalance between parameters.

Tendency to Focus on One Virtual Agent: As it can be seen on Table 4, there is a 6% probability of a match being declared while pAA is at level 0. A possible explanation can be concluded from both peri-exposure observations and the post-experiment questionnaire, where it was clear that many participants focused exclusively on the virtual agent sat by the table in front of them in the IVR experience. This might have resulted in them not noticing the other virtual agents in the environment as much. Further, it can be seen that in cases where a match is declared with pAA_0, pAR was at level 1. This, combined with the virtual agent in front of the participant drawing most of their attention, suggests that some participants accidentally triggered a reaction of said virtual agent and thereby thought animations were activate for all virtual agents and did not need to be upgraded. On the contrary, having a virtual agent seated right in front of the participant also makes it very obvious whether or not pAA needs to be upgraded if no reaction is triggered. The virtual agent might also influence pAR as it becomes more noticeable than it otherwise would. Many participants did not even notice other possible reactions outside of those from the virtual agent in front of them. This led to most participants even abusing the virtual agent in front of them reacting, by spilling the beer, to assess the level of pAR, as was also observed during the experiment.

Another factor to consider regarding pAR is the inverse of the problem presented above, where animations are being mistaken for reactions. As part of the experience, there is a timed event in which the virtual agent in front of the participants would raise their beer and cheer. For some participants, it led to a false impression that he was reacting to them picking up the beer and there was then no need to upgrade reactions.

6.1.1 Bias

First off, the results of each of the trials are assumed to be statistically independent, even though the within-subjects design forced participants to repeat the same trials, indicing carry-over effects. However, the design of randomly presenting different starting configurations to participants should eliminate carry-over effects, as they would be forced to evaluate the current configuration every time they started a new trial.

The demonstration video shown to participants introduces a bias to pAR as the purpose is to show each of the parameters and how the participant is supposed to upgrade them. However, to demonstrate upgrading pAR, only the reactions triggered by spilling the beer and touching the drink menu are shown as examples. This might result in participants thinking these are the only reactions they can trigger and therefore just copies what they saw in the demo video to test reactions rather than explore the environment for more reactions. Other reactions could also have been too subtle for many of the participants to notice as they are triggered by looking at specific objects or virtual agents.

In general, a possible BIP always exists because of how the experiment is conducted, which can in turn harm the illusion of plausibility.

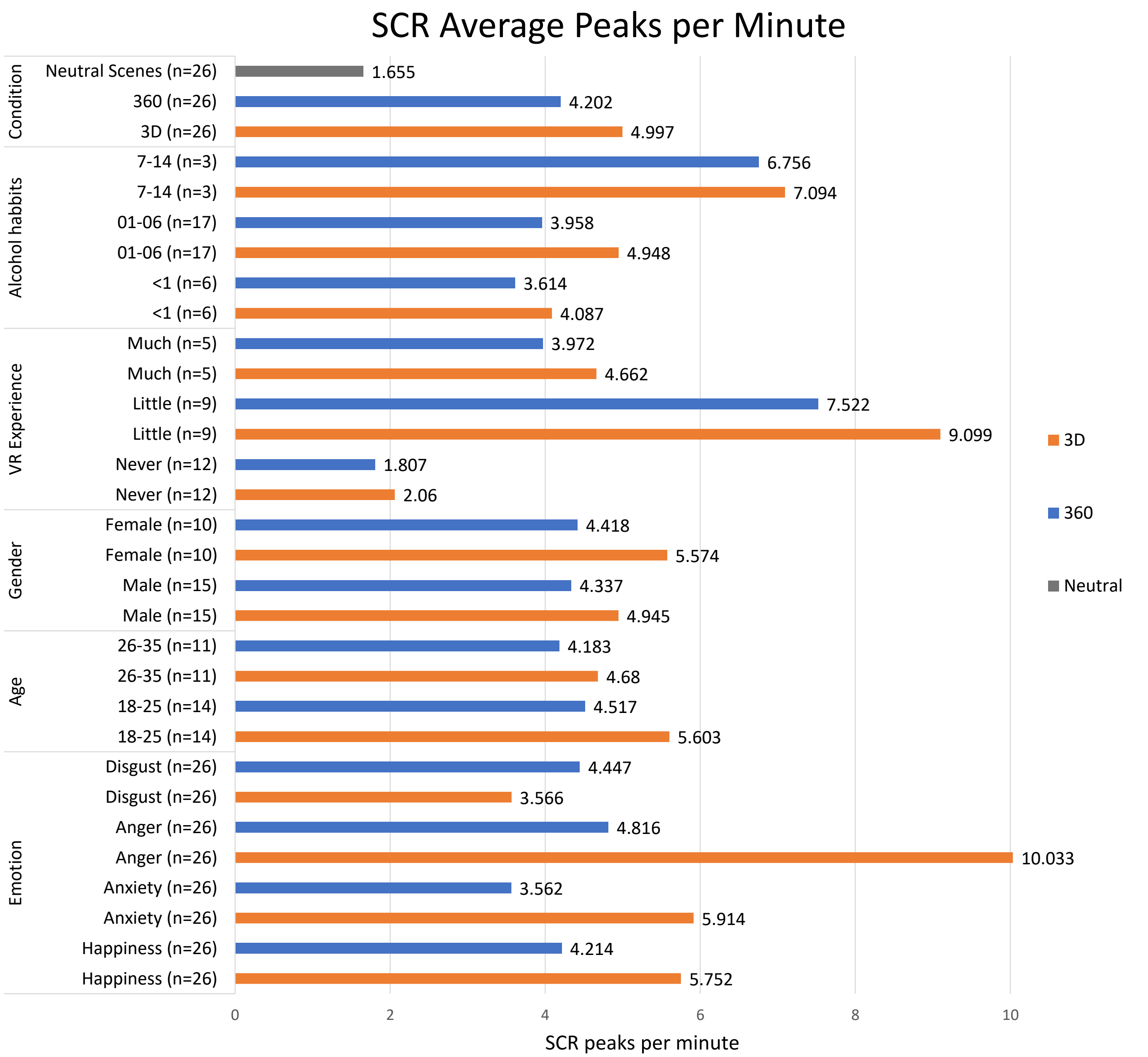


Figure 14: Bar chart of the SCR average significant PPMs. Samples are indicated in the groups to the left (n). Comparisons can be done in specific groups, and with, e.g., the neutral condition result, but cannot necessarily be done between all groups. Because of sample sizes, one nonbinary participant is not plotted in the gender group, and one participant aged 36-45 is not plotted in the age group. The top *Bar scenes* result is a mean of the combined 3D and 360 condition result.

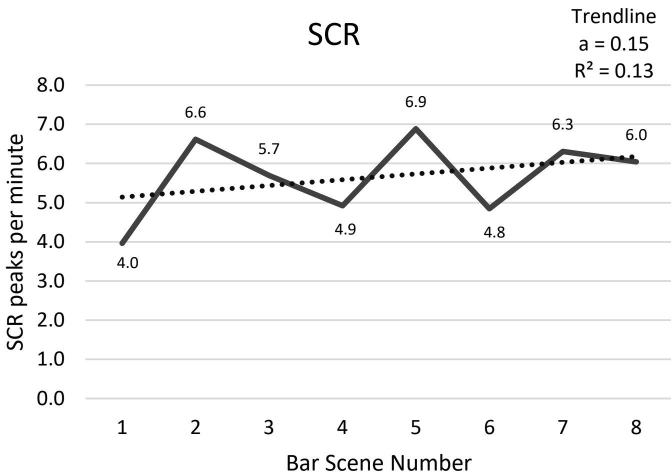


Figure 15: SCR PPM for every bar scene shown in the experiment in order of first to last with no regard to which scene was shown, with a trend (grey line) over the duration of the experiment.

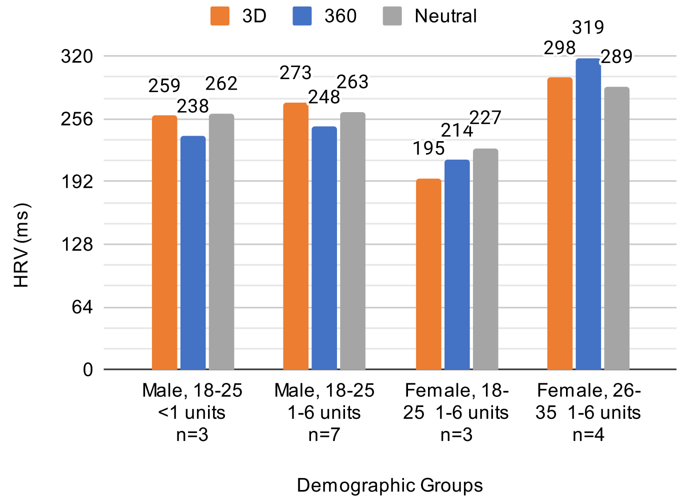


Figure 17: HRV compared between conditions in homogeneous sample groups sized ≥ 3, in terms of gender, age, and weekly alcohol consumption.

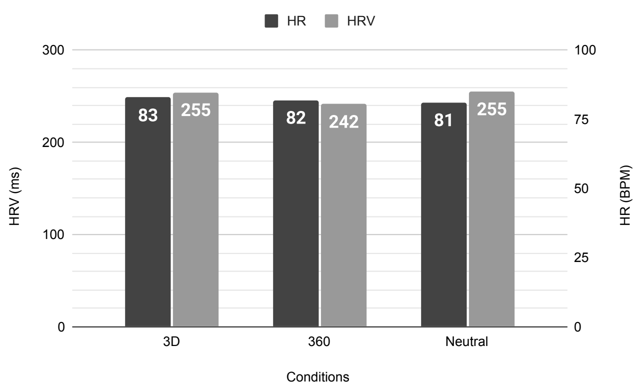


Figure 16: HRV (in ms) and HR (in BPM) values, comparing the three conditions experienced by participants.

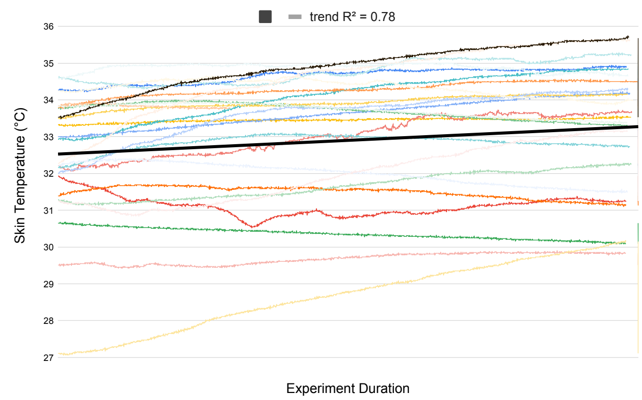


Figure 18: Skin temperature data and trend (black line) over the duration of the experiment.

The participant needs to meta-assess their situation and the surrounding environment evolving constantly over short spans of time, preventing feelings of presence that would simply evolve the longer the participant is immersed in the VE. Furthermore, the instructor needs to interfere every short while to help the participant initiate a transition to a new configuration. The short time spent in each configuration also possibly did not allow participants to pay attention to how real they felt the VE to be, and their parameter upgrades may not have been thoroughly considered each time. Designing a more punishing cost structure to each upgrade or setting a minimum amount of time to be used before being allowed to upgrade a parameter could also have been considered.

6.2 Experiment B: Craving in 3D-VR versus 360-VR

6.2.1 Craving

Responses in 3D and 360 Conditions: From the leftmost box plots in Fig. 10, it can be seen that self-reported levels of craving from the VAS are more similar when comparing the 360 conditions against the 3D condition than when comparing any of these against the neutral condition. This is supported by the indications of significant differences from the ANOVA seen in Fig. 11, where 3D and 360 bar conditions resulted in significantly higher craving than in the neutral condition. From the physiological data, significant differences cannot be seen in HR, HRV, or skin temperature data, but responses of SCR PPM follows this trend and shows a significant difference between bar conditions and the neutral condition, but no significant difference between bar conditions, as seen at the top of Fig. 14. SCR PPM data shows a general trend towards higher PPM values for the 3D condition than for the 360 condition in *all* but the *Disgust* scene. Whether these PPM observations can indicate higher levels of affection leading to higher levels of craving experienced in the 3D condition over the 360 condition, or simply are a result of participants being more active in the 3D condition, cannot be verified with the other physiological metrics or the VAS scores, however, as these show opposite trends. Still, the 3D bar condition with the *Anger* scene also differentiates wildly, reaching just above 10 PPM on average. Why this result is so different from the rest can only be a guess, as data has been computed similarly across conditions and groups, with randomizations of scene exposure order limiting carry-over effect. Perhaps some scores were affected by the increased possibility of interaction in that particular scene, with some participants reaching out for the shot glasses set on the table, possibly skewing data due to this interaction. If this is the case, it could support the explanation of the 3D condition scoring higher than the 360 condition, as a result of participants being more active in the 3D condition. Interestingly, the PPM values also differ substantially between participants with different experience with VR. Values reach as high as 9 PPM in the 3D condition and 7.5 in the 360 condition for participants with little experience with VR (n=9), whereas participants with no experience (n=12) reached much lower values of around 2 PPM. However, this may simply be explained by randomness due to the low sample sizes in this specific comparison.

The significant difference between bar and neutral conditions, however, demonstrates that alcohol cues and stimuli in the bar conditions might be successful in inducing a response, and the neutral condition can be used to deescalate responses back to baseline levels, as is also supported by the VAS craving scores. The results also suggest no difference in terms of how the VE was constructed concerning craving - that is, using either 360 or 3D approaches to creating VEs. Again, this is supported by the ANOVA and the physiological responses of HR, HRV, and skin temperatures, in Figs. 16 and 18, showing no significant difference between the two presented conditions.

High Variation in Craving in All Scenes: Fig. 10 shows the interquartile range indicating a high variation of the craving VAS results, which may be explained by the nature of craving and its tendency to be dependent on both current circumstances and the general mood of each participant. More consistent results might have been attained if such confounding variables were further considered. An obvious factor would be the time of day or day of the week, as different participants underwent the experiment during different times of the day, on the three last days of the workweek. Later in the workweek and later in the day is more common for the consumption of alcohol, especially when considering participant demographics comprising students primarily. Even the time of year can affect the stress level in participants, as the experiment was conducted close to an exam and project hand-in period. Physiological data can also be affected by this confounding variable.

Although the method is recommended and has seen use in various literature, using self-reported craving based on a single-item VAS is heavily affected by the subjective nature of the question, perhaps explaining the wide interquartile range in the craving box plots.

Differences in Values Over Time: From the blue trend-line seen in Fig. 12, we see that craving increases in respect to how many scenes the participant has experienced. Taking this into account, the 360 condition has a slight advantage in terms of experienced craving as it was on average presented to the participant as scene number 9.8, compared with scene number 8.72, and 8 for the 3D and neutral conditions, respectively. Furthermore, the increase in certain metrics over time may be explained by the fact that it is difficult for participants to be entirely unaffected by repeated exposure. Although the experimental design attempts to accommodate this by placing the participant in the neutral condition between each scene, there might still be some carry-over effect. The same trend can be seen in the physiological data of SCR PPM in Fig. 15, and skin temperature seen in Fig. 18, both slightly increasing over time. This may be explained by a heightened state of arousal of being exposed to cues as if participants were in a real bar indicating feelings of craving. A different explanation concerns the experiment duration and needing to wear IVR equipment for an extended period of time, or, finally, the room temperature increasing because the experiment was run in a small room with three people present at a time, with several computers dissipating heat as well.

Craving Measures Compared With Other Studies: Although the experiment was conducted on a sample population consisting of individuals without AUD, it might still be worth considering how much craving was measured in the different conditions, compared with other similar studies testing individuals *with* AUD. Although comparisons may not be directly meaningful, there might still be some similar patterns in the reported craving levels. The average craving measured in the neutral, 360, and 3D conditions in this study is *24.69*, *33.5*, and *33,7*, respectively. A study by Lee et al. found an average score in their neutral condition of approximately *5*, and *20* in their alcohol condition (a party scenario) [27]. The results are relatively similar, although with a larger difference in the neutral condition versus the alcohol condition compared to the findings in this study. In contrast, a study by Traylor et al. found scores of both their neutral and their party condition at around *60* for individuals with AUD. However, for non-AUD individuals, the scores were around *40* and *20* in the party and neutral condition, respectively [49]. Bordnick et al. found an average score of *11* in the neutral condition and *61* in their alcohol condition [20]. The higher craving scores reported by AUD individuals in these studies might be explained by AUD individuals responding much more strongly to alcohol cues. Note that the non-alcoholic individuals in the study by Traylor et al. reported craving levels similar to the ones found in this study.

For the physiological data, other studies often see increases in craving to correlate with changes in, e.g., increased HR, decreased HRV, increased SCR, as well as increased skin temperature [40, 46, 54, 164]. This is, naturally, mostly reported in alcohol-dependent individuals, making comparison or correlation to other works difficult as the measures vary highly because of individual characteristics [147]. Additionally, very few studies assessing the correlation between VR cue exposure and craving induction has utilized physiological measurements. However, the non-VR study by Reid et al. saw significant results from subjective assessments of alcohol craving, but no significant results from the physiological assessments [48], indicating similarity with the experiment of this study.

6.2.2 Place Illusions and Plausibility

From the PI and Psi box plots in Fig. 10, it can be seen that self-reported levels of PI from the VAS are more similar when comparing the 3D condition against the neutral condition than when comparing any of these against the 360 condition. This is supported by the indications of significant differences from the ANOVA seen in Fig. 11, where the 3D and neutral conditions resulted in significantly lower PI than in the 360 condition. When comparing the 360 and the 3D conditions, it is difficult to create a fair comparison due to differences in technology. The two conditions must be similar to be comparable, otherwise there is a risk of introducing confounding variables. For this reason, the conditions developed for this experiment was created to be as similar as possible, entailing the 3D condition being limited in terms of interaction capabilities (valid effectual actions) such as manipulating objects or triggering reactions from virtual agents. All scenarios in the experiment had the participant stationary at the table with the three actors in the 360 condition and virtual agents in the 3D condition. This, however, puts the 3D condition at a disadvantage as 6DOF and locomotion is an important aspect in the effectiveness of 3D virtual worlds. Not utilizing every aspect of the 3D condition might explain the lower PI score reported for the 3D condition than was reported for the 360 condition. This result might also be explained by the attempt of assessing PI using a single-item VAS not being comprehensive enough. Seen in Fig. 12, a slight decrease in PI is also observed with respect to the scene number, perhaps because participants detect certain technical flaws in the experience due to repeated exposure, and, e.g., begin to interact and play with objects, such as seen with participants reaching out for the shot glasses in the *Anger* scene, or in a few cases toppling or throwing the virtual beer.

In terms of Psi, results show a significant difference between all combinations of the three conditions, the 360 condition scoring highest, followed by the 3D condition, and finally the neutral condition. This was expected as the 360 condition facilitated a social context with higher fidelity than that of the 3D and neutral condition as the actors seemed more real to participants than did the virtual agents. Furthermore, the limitations of the 3D version not utilizing interaction capabilities also confirm previous assumptions that the Psi score would be higher in the 360 condition - this assumption was made from Experiment A, finding reactions to be crucial in facilitating Psi in 3D-generated VEs. The limitations caused by the 360 condition also meant the narrative itself had to be passive, as it is naturally not possible to change an environment that is created using pre-recorded footage based on the users' behavior. This limitation may be one of the largest issues faced when implementing a 360-video recording in a VE, as such experiences may provide a lower level of user engagement. Furthermore, not all participants utilized the few interaction possibilities developed for 3D environments - those participants could have had a lessened sense of PI as well as less feelings of craving as a result hereof. As both the VE of the neutral and the 3D condition were constructed using 3D-rendering, the similar results between these VEs in terms of PI *were* expected. Given the relatively passive experience of the bar scenarios, the same set of valid sensorimotor actions were present in both scenes.

6.2.3 Cybersickness

CS results presented in the the rightmost box-plot in Fig. 10 indicate a low level of CS in general, although with a modest amount of outliers, and a tendency of increasing in respect to number of scenes experienced by participants, as seen in Fig. 12. Certain events in scenes may have been provoking feelings of CS, or certain individuals may have been particularly prone to CS, which is in agreement with current research [187]. The 360 condition is seen to induce CS the most, likely due to the limitations in degrees of freedom as the elements in the scene would be stationary when participants moved their head, causing a sensory mismatch between the perceived surroundings and the feeling of vection. Naturally, feelings of CS could potentially impair test participants' ability to follow the narrative of the scenarios and hinder feelings of presence in the environment.

6.2.4 VAS Metrics Correlations

In terms of correlation between the four VAS metrics seen in Fig. 13, the strongest correlation exists between craving and Psi ($r = 0.496$) indicating a moderate positive relation. This supports findings in similar studies and emphasizes the requirement that IVR-CET experiences must be able to facilitate a plausible situation to effectively induce craving in users. The 360 condition promoted the highest level of Psi ($\mu = 71.1, \sigma = 14.9$), maybe due to the way the environment was constructed using video recordings of actors and a physical environment. Especially the movement of actors were notably superior in comparison to the 3D condition containing animated virtual agents. A weak positive relation was detected between CS and Psi ($r = 0.25$), the causality of which is possibly explained by the 360 condition scoring highest on Psi due to the realism of the social interactions, but also scoring highest on CS due to the 3DoF limitation. It might be interesting to investigate the relationship between CS and the feeling of Psi and PI further, as previous studies have suggested there is a negative correlation between CS and feelings of presence [127]. Lastly, a weak negative correlation between PI and craving ($r = -0.232$) is seen. This may be due to the tendency for craving to increase depending on how many scenes has been experienced versus PI which tends to decrease over time. As mentioned earlier, the increase in craving could be related to a carry-over effect between the scenes, whereas the decrease in PI may be because participants detect inconsistencies in the simulation they are presented with.

6.2.5 Bias

Perhaps the most prominent bias is regarding the physiological data, as the reliability and validity of the raw data can be disputed. The Empatica E4 wristband was very sensitive and would decrease in reading accuracy whenever it was moved. Participants were instructed to not move the arm wearing the wristband, but even then, a few participants would move their arm a bit. Furthermore, the wristband had a button pressed manually to mark an event in the data, as no similar feature was present in the accompanying smartphone app, affecting data acquisition and accuracy, as well as introduce a possible BIP for the participant. The participant can have been distracted by the button presses on the wristband made by the researcher, perhaps also influencing data accuracy, and even the VAS measures of plausibility and place illusions. An optimal solution would be to program an event marking function to run every time the scenes changed in the virtual experience, or at least when the operator advanced the experiment rounds. Lastly, it is worth noting the length of each scenario when considering the amount of time required to attain valid readings, both when acquiring baselines and when measuring responses to cue exposure. For baselines, some studies have suggested upwards of 15 minutes before such readings can be obtained, which contrasts the 1-2 minutes used to obtain baseline readings preceding bar exposure in the experiment. Skin temperature data in Fig. 18 can be particularly effected due to its rate of change over time being significantly lower than measures of SCR, HR, and HRV. Further, frequent scene changes also follows shorter duration of scenarios which can lead to possible BIPs.

Other factors more difficult to evaluate such as mood, stress level, and general alcohol attitude could also have been further investigated to ensure the validity of the self-reported craving. However, revealing such factors may raise ethical concerns due to the nature of the information that would have to be collected such as personal life events, drinking habits, etc.

6.3 Additional Observations and Considerations

Based on the findings from the two experiments, virtual environments in general can induce both psychological and physiological changes of feelings of presence and feelings of alcohol craving in participants. These changes relate highly to the exposure to virtual complex alcohol cues and social settings, prompting the potential for future research in implementations of these in IVR-CET tools.

From the results, the 360 environment is observed to induce greater responses in terms of feelings of plausibility and place illusions, but, as seen from some of the post-questionnaire answers, it was mainly due to the perceived photorealism making the people present at the bar feel more real. If sufficient time is invested, and required hardware is available, it would perhaps be possible to create a 3D environment eliciting the same subjective opinion of photorealism as seen when participants were exposed to the 360-video condition. The 3D environment would then not only be as realistic as the 360-video condition, but also include 6DOF and interactable elements in the environment. If this was the case, the 3D environment should surpass the 360 condition in most aspects.

To solidify findings and allow for further triangulation, assessing the behavior of participants could reveal additional important insights in both Experiment A and Experiment B, as detailed in Section 2.5.1 on behavioural assessment. In both experiments, several participants would engage verbally with, or make gestures towards the actors or virtual agents presented in either of the two conditions. In both experiments, measuring, e.g., eye gaze could be used to reveal participant focus and attention relating and perhaps correlating to either the parameters they could change in Experiment A, or to the induced feelings of craving and presence in Experiment B. Seeing if participants looked at either agents or objects, could triangulate, validate, and explain the behaviour and responses by the participant. Also the apparent interaction with objects in front of participants could be quantified, e.g., to measure how many times participants grabbed the beer or the drink menu and focused on either objects or people that were present in the VEs. Certain behaviours could be interpreted as an indication that participants feel as though the agents are real and it might influence their own behaviour so they adhere to social norms.

7 Conclusion and Future Work

In general CET, especially regarding phobias and anxiety, IVR has seen increased use due to its practicality, possibility of control during exposure, and effectiveness as a tool to present situations that elicit stressful responses. Positive effects are observed in individuals who used IVR as a tool in the process of learning to eradicate these responses. In CET for AUD, previous experiments with IVR tools have seen limited effect in eliciting feelings of craving needed to learn and promote the use of coping strategies. This claim is based on the hypothesis that current implementations lack in environment detail and fidelity, preventing feelings of presence through place and plausibility illusions that make the experience feel real. Specifically, the complex cues of social interaction, seen as a substantial element of the induction of alcohol craving, is difficult to implement, but important to portray convincingly in the virtual environment. Furthermore, additional senses can be utilized in inducing feelings of craving as opposed to traditional methods of presenting alcohol cues to participants. The idea of IVR-CET tools is to have the *ex vivo* virtual environment approach *in vivo* exposure to close the so-called transfer gap between clinical and real-world contexts, learning to reduce cue reactivity by being exposed to such cues in simulated high-risk situations, and experience and cope with the induced feelings of craving.

As such, this study presented two experiments with a shared goal of investigating how to best utilize IVR technology for inducing alcohol craving. Due to the cruciality of simulating high risk situation including social contexts, the first experiment focused on discovering which aspects of a 3D-rendered VE contributed most to the experienced plausibility in that environment. The second experiment compared two development approaches of either using 3D-generated or 360-video-based virtual environments on their efficacy in inducing alcohol craving in participants.

From the two experiments, several observations were made in relation to the potential use of an IVR in CET as a tool to induce alcohol craving. Regarding 3D-generated IVR experiences, four parameters consisting of virtual agent animations, virtual agent reactions, soundscape richness, and geometric realism were evaluated in their ability to induce feelings of plausibility in individuals. An experiment utilized a novel presence-matching design, and found that participants would prioritize upgrading parameters relating to, in order, soundscape richness, the animations of virtual agents, their reactions to user behavior, and, lastly, geometric realism. Results also revealed subtle details of the environment to be of much less importance in many cases, and, in general, users would maintain their focus mainly on the virtual agents seated directly in front of them. Finally, the presence-matching method used during the experiment was successfully employed in the investigation of the importance of the established Psi related parameters, proving itself as a valid method for evaluating the Psi of various aspect in IVR environments.

The second experiment found similar craving results comparing fully 3D-generated renditions of virtual environments versus 360-video-based environments, slightly favoring the 360-video experience in subjective assessment results, but the 3D experience elicited stronger responses of arousal in objective physiological measurements. A general tendency for craving to increase over time, as users were exposed to alcohol-related cues in both IVR environments in various social contexts, was observed. Compared to the neutral condition, craving was successfully induced in both bar conditions, eliciting significantly higher levels of craving. Additionally, a moderate correlation between assessments of craving and feelings of plausibility was found, matching findings seen in similar studies. In this study the choice of using VAS measurements for evaluating craving, PI, Psi and CS continuously during the experiment as participant were exposed to the different conditions proved to be a valid approach to determine such metrics. Although to a lower degree of precision, when compared with more extensive methods. Furthermore, the use of physiological measurement were used unsuccessfully as the method proved cumbersome, when compared to the outcome and reliability of results.

The big drawback of the experiment, however, is that participants have no history of alcohol abuse or dependency. As such, the findings of this study are not generalizable to individuals with AUD. However, individuals with AUD react substantially different to exposure to alcohol-related cues. Of course, this warrants further studies on appropriate participants. Yet, findings in the experiments that have been conducted so far will help in evaluating the facilitation of social context in VEs, evaluate cues and to validate assessment methods.

Still, IVR can present itself as a tool allowing for the simulation of a wide variety of situations where the user can feel present and engaged. Furthermore, it allows for therapy to be performed in a controlled environment that more easily enables gradual exposure. The purpose of this study was to evaluate which approach to generating a virtual environment for IVR-CET was superior. The choice between utilizing a fully 3D-generated environment versus a 360-video based virtual environment in a CET context comes down to a trade-off between visual realism and interactivity. Where using 360 video for generating environments will produce a visually superior result, it is impossible to create a dynamic environment that reacts to user action. Using 3D-generated content, however, allows for more dynamic and interactive environments because of the possible implementation of environments featuring both interactive objects and virtual agents that react to user action. Although, more dynamic environments also means a reduction in experimental control, which could be a threat to internal validity as the experience of each participant may differ significantly. Also, a tremendous effort is needed to create 3D content that visually matches 360-generated environments. With the current state of computer graphics technology, it is not possible to produce 3D-generated virtual environments of such high fidelity that competes with recorded 360 video in terms of believability, which in the experiment was found to be an important factor of inducing feelings of plausibility. Once it is possible to produce photorealistic 3D-generated content for VR, 3D-generated content will undoubtedly be superior compared to noninteractive 360 videos as the potential for creating dynamic, reactive, and interactive experiences.

In continuation of this study, it would be interesting to conduct a control group experiment at the physical bar from which the 360 and 3D environments were created. This would allow for comparisons between the data collected during the two experiments to check for validity and to gain further insight in the design of cues and other elements in the simulated environment. Future studies should also assess the metrics of craving, as well as place and plausibility illusions, in experiments that better exploit the functionality of each of the implementation approaches when creating the virtual environment. A 3D environment could be created and tested in an experiment that involves more interaction and locomotion than the experiments in this study, also assessing feelings of presence and craving in the participant. Examining feelings of craving on participants in such environments could be interesting as increased engagement may lead to increased feelings of presence which in turn could prove beneficial when attempting to induce craving. On the opposite, a 360-video implementation can be developed when the experience needs to be stationary and linearly scripted regarding interaction. This could contain a much more complex narrative to engage the user over a longer period, as opposed to the shorter scripted scenarios used in the experiments in this study. Such an experience could allow the user to become more emotionally invested in the individuals they meet, also in part due to increased photorealism. Additionally, it may be interesting to investigate the relationship between plausibility illusions and feelings of craving which were observed to correlate in Experiment B. It is not unreasonable to assume that increased illusions of plausibility may lead to increased feelings of craving, as feelings of presence has been linked to changes in the emotional and psychological states of individuals, possibly enabling inductions of alcohol craving.

8 Acknowledgements

The authors would like to thank Linnea Bjerregaard Pedersen, Benjamin Bække, Maria Anna Krupa, and Machiej Tomasz Odowaz-Sypniewski from 8th semester of Medialogy at Aalborg University for the filming, processing, and sharing of the 360-degree videos used in Experiment B. The authors would also like to thank Daniel Thaysen-Petersen from the Institute of Clinical Medicine at the University of Copenhagen for sharing his expertise in AUD treatment, cue exposure therapy, and theories in psychology. Finally, the authors would like to express their gratitude to Rolf Nordahl, Ali Adjorlu, and Stefania Serafin from the Multisensory Experience Laboratory at Aalborg University Copenhagen for their support during the last few years, helping and bringing new project ideas and directions to the table whenever needed.

References

[1] American Psychiatric Association. "Diagnostic and statistical manual of mental disorders : DSM-5". In: Arlington, VA Washington, D.C: American Psychiatric Association, 2013, pp. 490–493. ISBN: 9780890425558.

[2] Jürgen Rehm et al. "Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders". In: *The Lancet* 373.9682 (2009), pp. 2223–2233. DOI: 10.1016/s0140-6736(09)60746-7.

[3] WHO Europe. *Alcohol use data and statistics*. 2020. URL: https://euro.who.int/en/health-topics/disease-prevention/alcohol-use/data-and-statistics.

[4] Ben L. Da, Gene Y. Im, and Thomas D. Schiano. "Coronavirus Disease 2019 Hangover: A Rising Tide of Alcohol Use Disorder and Alcohol-Associated Liver Disease". In: *Hepatology* 72.3 (2020), pp. 1102–1108. DOI: 10.1002/hep.31307.

[5] William D.S. Killgore et al. "Alcohol dependence during COVID-19 lockdowns". In: *Psychiatry Research* 296 (2021), p. 113676. DOI: 10.1016/j.psychres.2020.113676.

[6] Ariadna Capasso et al. "Increased alcohol use during the COVID-19 pandemic: The effect of mental health and age in a cross-sectional sample of social media users in the U.S." In: *Preventive Medicine* 145 (2021), p. 106422. DOI: 10.1016/j.ypmed.2021.106422.

[7] Sangwoo Cho et al. "Development and Verification of an Alcohol Craving–Induction Tool Using Virtual Reality: Craving Characteristics in Social Pressure Situation". In: *CyberPsychology & Behavior* 11.3 (2008), pp. 302–309. DOI: 10.1089/cpb.2007.0149.

[8] G Alan Marlatt and Dennis M Donovan. *Relapse prevention: Maintenance strategies in the treatment of addictive behaviors*. Guilford press, 2005.

[9] Juris P Mezinskis et al. "The measurement of craving". In: *Journal of addictive diseases* 20.3 (2001), pp. 67–85.

[10] Simon P. Byrne et al. "Cue Exposure Therapy for Alcohol Use Disorders: What Can Be Learned from Exposure Therapy for Anxiety Disorders?" In: *Substance Use & Misuse* 54.12 (2019), pp. 2053–2063. DOI: 10.1080/10826084.2019.1618328.

[11] Alexandra Ghiţă et al. "Identifying Triggers of Alcohol Craving to Develop Effective Virtual Environments for Cue Exposure Therapy". In: *Frontiers in Psychology* 10 (2019). DOI: 10.3389/fpsyg.2019.00074.

[12] Judith S Beck and AT Beck. "Cognitive behavior therapy". In: *New York: Basics and beyond. Guilford Publication* (2011).

[13] Raymond F. Anton et al. "Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence". In: *JAMA* 295.17 (2006), p. 2003. DOI: 10.1001/jama.295.17.2003.

[14] Angelina I. Mellentin et al. "Cue exposure therapy for the treatment of alcohol use disorders: A meta-analytic review". In: *Clinical Psychology Review* 57 (2017), pp. 195–207. DOI: 10.1016/j.cpr.2017.07.006.

[15] Stuart G. Ferguson and Saul Shiffman. "Relation of Craving and Appetitive Behavior". In: *Principles of Addiction*. Elsevier, 2013, pp. 473–479. DOI: 10.1016/b978-0-12-398336-7.00049-8.

[16] Thiagarajan Sitharthan et al. "Cue exposure in moderation drinking: A comparison with cognitive–behavior therapy." In: *Journal of Consulting and Clinical Psychology* 65.5 (1997), pp. 878–882. DOI: 10.1037/0022-006x.65.5.878.

[17] Andreas Junker et al. "Virtual Womb: Experiencing Human Sensory Development From a Fetal Point-of-View in Virtual Reality". In: *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. ACM, 2020. DOI: 10.1145/3419249.3420159.

[18] Walter Greenleaf. "How VR technology will transform healthcare". In: *ACM SIGGRAPH 2016 VR Village*. ACM, 2016. DOI: 10.1145/2929490.2956569.

[19] Wee Sim Khor et al. "Augmented and virtual reality in surgery—the digital surgical environment: applications, limitations and legal pitfalls". In: *Annals of Translational Medicine* 4.23 (2016), pp. 454–454. DOI: 10.21037/atm.2016.12.23.

[20] Patrick S. Bordnick et al. "Assessing reactivity to virtual reality alcohol based cues". In: *Addictive Behaviors* 33.6 (2008), pp. 743–756. DOI: 10.1016/j.addbeh.2007.12.010.

[21] Pedro Gamito et al. "Eliciting Nicotine Craving with Virtual Smoking Cues". In: *Cyberpsychology, Behavior, and Social Networking* 17.8 (2014), pp. 556–561. ISSN: 2152-2715, 2152-2723. DOI: 10.1089/cyber.2013.0329.

[22] Caleb Ward. *Demystifying 360 vs. VR*. Vimeo Blog. 2017. URL: https://vimeo.com/blog/post/virtual-reality-vs-360-degree-video/ (visited on 05/13/2021).

[23] Anurag Syal. *Panorama vs 360°vs 3D vs VR — The Big Difference*. Medium. 2020. URL: https://medium.com/frulix/panorama-vs-360-vs-3d-vs-vr-the-big-difference-d9e490819a0.

[24] Ben Lang. *An Introduction to Positional Tracking and Degrees of Freedom (DOF)*. 2013. URL: https://www.roadtovr.com/introduction-positional-tracking-degrees-freedom-dof/.

[25] Mel Slater. "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009), pp. 3549–3557.

[26] Ahmed Elmezeny, Nina Edenhofer, and Jeffrey Wimmer. "Immersive Storytelling in 360-Degree Videos: An Analysis of Interplay Between Narrative and Technical Immersion". In: *Journal For Virtual Worlds Research* 11.1 (2018). ISSN: 1941-8477. DOI: 10.4101/jvwr.v11i1.7298.

[27] Jung Suk Lee et al. "Social Pressure-Induced Craving in Patients with Alcohol Dependence: Application of Virtual Reality to Coping Skill Training". In: *Psychiatry Investigation* 5.4 (2008), p. 239. DOI: 10.4306/pi.2008.5.4.239.

[28] Mel Slater, Bernhard Spanlang, and David Corominas. "Simulating virtual environments within virtual environments as the basis for a psychophysics of presence". In: *ACM Transactions on Graphics* 29.4 (2010), pp. 1–9. ISSN: 0730-0301, 1557-7368. DOI: 10.1145/1778765.1778829.

[29] Richard Skarbez, Frederick P. Brooks Jr., and Mary C. Whitton. "A Survey of Presence and Related Concepts". In: *ACM Computing Surveys* 50.6 (2017), 96:1–96:39. ISSN: 0360-0300. DOI: 10.1145/3134301.

[30] Ilias Bergström et al. "The plausibility of a string quartet performance in virtual reality". In: *IEEE transactions on visualization and computer graphics* 23.4 (2017), pp. 1352–1359.

[31] M. Murcia-López et al. "Evaluating Virtual Reality Experiences Through Participant Choices". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). 2020, pp. 747–755. DOI: 10.1109/VR46266.2020.00098.

[32] Theresa H. McKim et al. "Translational Research on Habit and Alcohol". In: *Current Addiction Reports* 3.1 (2016), pp. 37–49. DOI: 10.1007/s40429-016-0089-8.

[33] Olga Hernández-Serrano et al. "Predictors of Changes in Alcohol Craving Levels during a Virtual Reality Cue Exposure Treatment among Patients with Alcohol Use Disorder". In: *Journal of Clinical Medicine* 9.9 (2020), p. 3018. DOI: 10.3390/jcm9093018.

[34] Ariel Zvielli, Amit Bernstein, and Ernst H. W. Koster. “Temporal Dynamics of Attentional Bias”. In: *Clinical Psychological Science* 3.5 (2014), pp. 772–788. DOI: 10.1177/2167702614551572.

[35] Yair Bar-Haim et al. “Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study.” In: *Psychological Bulletin* 133.1 (2007), pp. 1–24. DOI: 10.1037/0033-2909.133.1.1.

[36] Jonathan Baron. *Thinking and Deciding*. Cambridge University Press, 2006. DOI: 10.1017/cbo9780511840265.

[37] Kirsten Kaya Roessler, Martin Mau, and Claus Thorn Ekstrøm. “Interpersonal problems of alcohol use disorder patients undergoing a physical exercise intervention – a randomised controlled trial”. In: *Nordic Psychology* 70.3 (2018), pp. 245–255. DOI: 10.1080/19012276.2017.1418414.

[38] Georg Schomerus et al. “Self-stigma in alcohol dependence: Consequences for drinking-refusal self-efficacy”. In: *Drug and Alcohol Dependence* 114.1 (2011), pp. 12–17. DOI: 10.1016/j.drugalcdep.2010.08.013.

[39] Ismene L Petrakis et al. “Comorbidity of alcoholism and psychiatric disorders: an overview”. In: *Alcohol Research & Health* 26.2 (2002), p. 81.

[40] Helen C. Fox et al. “Stress-Induced and Alcohol Cue-Induced Craving in Recently Abstinent Alcohol-Dependent Individuals”. In: *Alcoholism: Clinical and Experimental Research* 31.3 (2007), pp. 395–403. ISSN: 0145-6008, 1530-0277. DOI: 10.1111/j.1530-0277.2006.00320.x.

[41] Michelle Snelleman. *Exploring Factors Related to Craving and Relapse in Alcohol-Dependent Outpatients*. Tech. rep. IVO-reeks. Instituut voor Verslavingsonderzoek, 2017.

[42] Gillian Steckler, Katie Witkiewitz, and G. Alan Marlatt. “Relapse and Lapse”. In: *Principles of Addiction*. Elsevier, 2013, pp. 125–132. DOI: 10.1016/b978-0-12-398336-7.00013-9.

[43] Rudolf H. Moos and Bernice S. Moos. “Rates and predictors of relapse after natural and treated remission from alcohol use disorders”. In: *Addiction* 101.2 (2006), pp. 212–222. DOI: 10.1111/j.1360-0443.2006.01310.x.

[44] Jörg Albers. *EKSPONERINGSBEHANDLING: HÅNDTERING AF TRANG*. Sydgården. Aabenraa Landevej 120, 6100 Haderslev, 2005.

[45] D. Colin Drummond et al. “Craving research: future directions”. In: *Addiction* 95.8 (2000), pp. 247–255. ISSN: 09652140. DOI: 10.1046/j.1360-0443.95.8s2.13.x.

[46] David J Drobes and Suzanne E Thomas. “Assessing craving for alcohol”. In: *Alcohol Research & Health* 23.3 (1999), p. 179.

[47] Cynthia A. Conklin et al. “Proximal versus distal cues to smoke: The effects of environments on smokers’ cue-reactivity.” In: *Experimental and Clinical Psychopharmacology* 16.3 (2008), pp. 207–214. ISSN: 1936-2293, 1064-1297. DOI: 10.1037/1064-1297.16.3.207.

[48] M. S. Reid et al. “Physiological and subjective responding to alcohol cue exposure in alcoholics and control subjects: evidence for appetitive responding”. In: *Journal of Neural Transmission* 113.10 (2006), pp. 1519–1535. ISSN: 0300-9564, 1435-1463. DOI: 10.1007/s00702-005-0439-5.

[49] Amy C. Traylor et al. “Using virtual reality to investigate complex and contextual cue reactivity in nicotine dependent problem drinkers”. In: *Addictive Behaviors* 36.11 (2011), pp. 1068–1075. DOI: 10.1016/j.addbeh.2011.06.014.

[50] Frauke Nees et al. “The role of context in the processing of alcohol-relevant cues”. In: *Addiction Biology* 17.2 (2011), pp. 441–451. DOI: 10.1111/j.1369-1600.2011.00347.x.

[51] Joseph J. Ryan et al. “Virtual Reality Cues for Binge Drinking in College Students”. In: *Cyberpsychology, Behavior, and Social Networking* 13.2 (2010), pp. 159–162. DOI: 10.1089/cyber.2009.0211.

[52] Alexandra Ghiţă and José Gutiérrez-Maldonado. “Applications of virtual reality in individuals with alcohol misuse: A systematic review”. In: *Addictive Behaviors* 81 (2018), pp. 1–11. DOI: 10.1016/j.addbeh.2018.01.036.

[53] Deok-Yong Kim and Jang-Han Lee. “Development of a Virtual Approach–Avoidance Task to Assess Alcohol Cravings”. In: *Cyberpsychology, Behavior, and Social Networking* 18.12 (2015), pp. 763–766. DOI: 10.1089/cyber.2014.0490.

[54] Tomoyuki Segawa et al. “Virtual Reality (VR) in Assessment and Treatment of Addictive Disorders: A Systematic Review”. In: *Frontiers in Neuroscience* 13 (2020), p. 1409. ISSN: 1662-453X. DOI: 10.3389/fnins.2019.01409.

[55] Lara A Ray et al. “Differences between treatment-seeking and non-treatment-seeking participants in medication studies for alcoholism: do they matter?” In: *The American journal of drug and alcohol abuse* 43.6 (2017), pp. 703–710.

[56] Aaron T Beck and Marjorie Weishaar. “Cognitive therapy”. In: *Comprehensive handbook of cognitive therapy*. Springer, 1989, pp. 21–36.

[57] Judith S Beck. “Cognitive therapy”. In: *The Corsini encyclopedia of psychology* (2010), pp. 1–3.

[58] Paul M Salkovskis. *Frontiers of cognitive therapy*. Guilford Press, 1997.

[59] Brad W Lundahl et al. “A meta-analysis of motivational interviewing: Twenty-five years of empirical studies”. In: *Research on social work practice* 20.2 (2010), pp. 137–160.

[60] Katie Witkiewitz et al. “Mindfulness-based relapse prevention and transcranial direct current stimulation to reduce heavy drinking: A double-blind sham-controlled randomized trial”. In: *Alcoholism: Clinical and Experimental Research* 43.6 (2019), pp. 1296–1307.

[61] Marco Cavicchioli, Mariagrazia Movalli, and Cesare Maffei. “Difficulties with emotion regulation, mindfulness, and substance use disorder severity: the mediating role of self-regulation of attention and acceptance attitudes”. In: *The American journal of drug and alcohol abuse* 45.1 (2019), pp. 97–107.

[62] Pei Chen et al. “Acupuncture for alcohol use disorder”. In: *International journal of physiology, pathophysiology and pharmacology* 10.1 (2018), p. 60.

[63] Marie-Thérèse Lussier and Claude Richard. “The motivational interview: in practice”. In: *Canadian Family Physician* 53.12 (2007), pp. 2117–2118.

[64] Anne C Fernandez et al. “Measuring mindfulness and examining its relationship with alcohol use and negative consequences.” In: *Psychology of Addictive Behaviors* 24.4 (2010), p. 608.

[65] National Institute on Alcohol Abuse and Alcoholism. “Treatment for Alcohol Problems: Finding and Getting Help”. In: *NIH Brochure* (2014).

[66] Kathleen M Carroll. “Relapse prevention as a psychosocial treatment: A review of controlled clinical trials.” In: *Experimental and Clinical Psychopharmacology* (1997).

[67] Kathleen M Carroll and Lisa S Onken. “Behavioral therapies for drug abuse”. In: *American Journal of Psychiatry* 162.8 (2005), pp. 1452–1460.

[68] Jennifer E Irvin et al. “Efficacy of relapse prevention: a meta-analytic review.” In: *Journal of consulting and clinical psychology* 67.4 (1999), p. 563. DOI: 10.1037//0022-006x.67.4.563.

[69] Molly Magill and Lara A Ray. "Cognitive-behavioral treatment with adult alcohol and illicit drug users: a meta-analysis of randomized controlled trials". In: *Journal of studies on alcohol and drugs* 70.4 (2009), pp. 516–527.

[70] Molly Magill et al. "The search for mechanisms of cognitive behavioral therapy for alcohol or other drug use disorders: A systematic review". In: *Behaviour Research and Therapy* 131 (2020), p. 103648.

[71] Peter M. Monti and Damaris J. Rohsenow. "Coping-skills training and cue-exposure therapy in the treatment of alcoholism." In: *Alcohol research and health: the journal of the National Institute on Alcohol Abuse and Alcoholism* 23.2 (1999), pp. 107–15.

[72] Peter M. Monti et al. *Treating alcohol dependence: A coping skills training guide, 2nd ed.* Treating alcohol dependence: A coping skills training guide, 2nd ed. New York, NY, US: Guilford Press, 2002. xii, 196. ISBN: 978-1-57230-793-3.

[73] Joscha Böhnlein et al. "Factors influencing the success of exposure therapy for specific phobia: A systematic review". In: *Neuroscience & Biobehavioral Reviews* 108 (2020), pp. 796–820. DOI: 10.1016/j.neubiorev.2019.12.009.

[74] Jeremy S. Joseph and Matt J. Gray. "Exposure therapy for posttraumatic stress disorder." In: *The Journal of Behavior Analysis of Offender and Victim Treatment and Prevention* 1.4 (2008), pp. 69–79. DOI: 10.1037/h0100457.

[75] Reid K. Hester and William R. (William Richard) Miller. *Handbook of alcoholism treatment approaches : effective alternatives*. English. Pergamon general psychology series ; 157. New York: Pergamon Press, 1989.

[76] Dianne L. Chambless and Thomas H. Ollendick. "Empirically Supported Psychological Interventions: Controversies and Evidence". In: *Annual Review of Psychology* 52.1 (2001), pp. 685–716. DOI: 10.1146/annurev.psych.52.1.685.

[77] Sabine Loeber et al. "Cue exposure in the treatment of alcohol dependence: Effects on drinking outcome, craving and self-efficacy". In: *British Journal of Clinical Psychology* 45.4 (2006), pp. 515–529. DOI: 10.1348/014466505x82586.

[78] Remco C. Havermans et al. "The efficacy of cue exposure with response prevention in extinguishing drug and alcohol cue reactivity". In: *Behavioral Interventions* 22.2 (2007), pp. 121–135. DOI: 10.1002/bin.219.

[79] Richard J Gerrig et al. *Psychology and life*. Pearson Higher Education AU, 2015.

[80] Harold Rosenberg. "Controlled Drinking". In: *Encyclopedia of Psychotherapy*. Ed. by Michel Hersen and William Sledge. New York: Academic Press, 2002, pp. 533–544. ISBN: 978-0-12-343010-6. DOI: 10.1016/B0-12-343010-0/00061-1.

[81] Richard J. Longmore and Michael Worrell. "Do we need to challenge thoughts in cognitive behavior therapy?" In: *Clinical Psychology Review* 27.2 (2007), pp. 173–187. DOI: 10.1016/j.cpr.2006.08.001.

[82] Bruce E. Wampold et al. "In pursuit of truth: A critical examination of meta-analyses of cognitive behavior therapy". In: *Psychotherapy Research* 27.1 (2016), pp. 14–32. DOI: 10.1080/10503307.2016.1249433.

[83] David J. Kavanagh et al. "Addition of cue exposure to cognitive-behaviour therapy for alcohol misuse: a randomized trial with dysphoric drinkers". In: *Addiction (Abingdon, England)* 101.8 (2006), pp. 1106–1116. ISSN: 0965-2140. DOI: 10.1111/j.1360-0443.2006.01488.x.

[84] James MacKillop and Stephen A. Lisman. "Effects of a context shift and multiple context extinction on reactivity to alcohol cues." In: *Experimental and Clinical Psychopharmacology* 16.4 (2008), pp. 322–331. DOI: 10.1037/a0012686.

[85] Petra K Staiger, Janet D Greeley, and Stephen D Wallace. "Alcohol exposure therapy: generalisation and changes in responsivity". In: *Drug and Alcohol Dependence* 57.1 (1999), pp. 29–40. DOI: 10.1016/s0376-8716(99)00037-x.

[86] Catherine A. Hartley and Elizabeth A. Phelps. "Extinction Learning". In: *Encyclopedia of the Sciences of Learning*. Springer US, 2012, pp. 1252–1253. DOI: 10.1007/978-1-4419-1428-6_306.

[87] Cynthia A Conklin and Stephen T Tiffany. "Applying extinction research and theory to cue-exposure addiction treatments". In: *Addiction* 97.2 (2002), pp. 155–167. DOI: 10.1046/j.1360-0443.2002.00014.x.

[88] Antoine Hone-Blanchet, Tobias Wensing, and Shirley Fecteau. "The Use of Virtual Reality in Craving Assessment and Cue-Exposure Therapy in Substance Use Disorders". In: *Frontiers in Human Neuroscience* 8 (2014). DOI: 10.3389/fnhum.2014.00844.

[89] Kenneth A. Perkins. "Subjective Reactivity to Smoking Cues as a Predictor of Quitting Success". In: *Nicotine & Tobacco Research* 14.4 (2012), pp. 383–387. ISSN: 1462-2203. DOI: 10.1093/ntr/ntr229.

[90] Emily M. Becker, Ashley M. Smith, and Amanda Jensen-Doss. "Who's using treatment manuals? A national survey of practicing therapists". In: *Behaviour Research and Therapy* 51.10 (2013), pp. 706–710. ISSN: 00057967. DOI: 10/f5c23r.

[91] Christina Hauke et al. "Standardized treatment manuals: Does adherence matter?" In: *Sensoria: A Journal of Mind, Brain & Culture* 10.2 (2014), p. 1. ISSN: 2203-8469. DOI: 10/gjkq9d.

[92] Damaris J. Rohsenow and Megan M. Pinkston-Camp. "Cognitive-behavioral approaches". In: *The Oxford handbook of substance use and substance use disorders, Vol. 2*. Oxford library of psychology. New York, NY, US: Oxford University Press, 2016, pp. 531–566. ISBN: 978-0-19-938172-2.

[93] Merton S. Krause. "How the psychotherapy research community must work toward measurement validity and why". In: *Journal of Clinical Psychology* 61.3 (2005), pp. 269–283. ISSN: 0021-9762. DOI: 10/cfzcw6.

[94] H. Pashler and E. Wagenmakers. "Editors' Introduction to the Special Section on Replicability in Psychological Science". In: *Perspectives on psychological science : a journal of the Association for Psychological Science* (2012). DOI: 10/gckf56.

[95] F. Leichsenring et al. "Biases in research: risk factors for non-replicability in psychotherapy and pharmacotherapy research". In: *Psychological Medicine* 47.6 (2017), pp. 1000–1011. ISSN: 1469-8978. DOI: 10/f9hd38.

[96] T. Stanley, E. Carter, and Hristos Doucouliagos. "What Meta-Analyses Reveal About the Replicability of Psychological Research". In: *Psychological bulletin* (2018). DOI: 10/gfq37h.

[97] Simone Grassini, Karin Laumann, and Martin Rasmussen Skogstad. "The Use of Virtual Reality Alone Does Not Promote Training Performance (but Sense of Presence Does)". In: *Frontiers in Psychology* 11 (2020). DOI: 10.3389/fpsyg.2020.01743.

[98] Thomas D. Parsons. "Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences". In: *Frontiers in Human Neuroscience* 9 (2015). DOI: 10.3389/fnhum.2015.00660.

[99] David H. Barlow et al. "The transfer gap in systematic desensitization: An analogue study". In: *Behaviour Research and Therapy* 7.2 (1969), pp. 191–196. DOI: 10.1016/0005-7967(69)90032-1.

[100] B.O. Rothbaum et al. "Virtual Reality Exposure Therapy". In: *Encyclopedia of Mental Health*. Elsevier, 2016, pp. 370–374. DOI: 10.1016/b978-0-12-397045-9.00196-8.

[101] Mel Slater. “Immersion and the illusion of presence in virtual reality”. In: *British Journal of Psychology* 109.3 (2018), pp. 431–433. DOI: 10.1111/bjop.12305.

[102] Steven Yantis and Richard A Abrams. *Sensation and perception*. Worth Publishers, Macmillan Learning, 2017. ISBN: 978-1-4641-1170-9.

[103] Fabian Hutmacher. “Why Is There So Much More Research on Vision Than on Any Other Sensory Modality?” In: *Frontiers in Psychology* 10 (2019). DOI: 10.3389/fpsyg.2019.02246.

[104] Emil Rosenlund Høeg et al. “System Immersion in Virtual Reality-Based Rehabilitation of Motor Function in Older Adults: A Systematic Review and Meta-Analysis”. In: *Frontiers in Virtual Reality* 2 (2021), p. 647993. ISSN: 2673-4192. DOI: 10.3389/frvir.2021.647993.

[105] Ultra Leap. *Tracking | Leap Motion Controller | Ultraleap*. 2019. URL: https://ultraleap.com/product/leap-motion-controller/.

[106] James J. Cummings and Jeremy N. Bailenson. “How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence”. In: *Media Psychology* 19.2 (2015), pp. 272–309. DOI: 10.1080/15213269.2015.1015740.

[107] K. Pagano, A. Haddad, and T. Crosby. “Virtual Reality-Making Good On the Promise of Immersive Learning: The effectiveness of in-person training, with the logistical and cost-effective benefits of computer-based systems”. In: *IEEE Consumer Electronics Magazine* 6.1 (2017), pp. 45–47. DOI: 10.1109/MCE.2016.2614413.

[108] Martijn J. Schuemie et al. “Research on Presence in Virtual Reality: A Survey”. In: *CyberPsychology & Behavior* 4.2 (2001), pp. 183–201. ISSN: 1094-9313, 1557-8364. DOI: 10.1089/109493101300117884.

[109] Anton Bogdanovych, Tomas Trescak, and Simeon Simoff. “What makes virtual agents believable?” In: *Connection Science* 28.1 (2016), pp. 83–108.

[110] Iis P. Tussyadiah et al. “Virtual reality, presence, and attitude change: Empirical evidence from tourism”. In: *Tourism Management* 66 (2018), pp. 140–154. DOI: 10.1016/j.tourman.2017.12.003.

[111] Mel Slater and Sylvia Wilbur. “A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments”. In: *Presence: Teleoperators & Virtual Environments* 6.6 (1997), pp. 603–616.

[112] Nick Yee and Jeremy Bailenson. “The Proteus Effect: The Effect of Transformed Self-Representation on Behavior”. In: *Human Communication Research* 33.3 (2007), pp. 271–290. ISSN: 0360-3989, 1468-2958. DOI: 10.1111/j.1468-2958.2007.00299.x.

[113] Nami Ogawa et al. “Do You Feel Like Passing Through Walls?: Effect of Self-Avatar Appearance on Facilitating Realistic Behavior in Virtual Environments”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20: CHI Conference on Human Factors in Computing Systems. Honolulu HI USA: ACM, 2020, pp. 1–14. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376562.

[114] Jean-Luc Lugrin et al. “Any “body” there? avatar visibility effects in a virtual reality game”. In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2018, pp. 17–24.

[115] Konstantina Kilteni, Raphaela Groten, and Mel Slater. “The Sense of Embodiment in Virtual Reality”. In: *Presence: Teleoperators and Virtual Environments* 21.4 (2012), pp. 373–387. DOI: 10.1162/pres_a_00124.

[116] Konstantina Kilteni et al. “Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception”. In: *Frontiers in Human Neuroscience* 9 (2015). DOI: 10.3389/fnhum.2015.00141.

[117] Rabindra Ratan et al. “Avatar characteristics induce users’ behavioral conformity with small-to-medium effect sizes: a meta-analysis of the proteus effect”. In: *Media Psychology* 23.5 (2019), pp. 651–675. DOI: 10.1080/15213269.2019.1623698.

[118] Nick Yee, Jeremy N. Bailenson, and Nicolas Ducheneaut. “The Proteus effect: Implications of transformed digital self-representation on online and offline behavior”. In: *Communication Research* 36.2 (2009), pp. 285–312. DOI: 10.1177/0093650208330254.

[119] Massimiliano di Luca et al. “Locomotion vault: the extra mile in analyzing VR locomotion techniques”. In: *ACM CHI*. 2021. URL: https://www.microsoft.com/en-us/research/publication/locomotion-vault-the-extra-mile-in-analyzing-vr-locomotion-techniques/.

[120] Costas Boletsis. “The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology”. In: *Multimodal Technologies and Interaction* 1.4 (2017), p. 24.

[121] Martin Usoh et al. “Walking> walking-in-place> flying, in virtual environments”. In: *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. 1999, pp. 359–364.

[122] Eike Langbehn and Frank Steinicke. “Redirected Walking in Virtual Reality”. In: 2018.

[123] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. “A systematic review of cybersickness”. In: *Proceedings of the 2014 conference on interactive entertainment*. 2014, pp. 1–9. URL: https://doi.org/10.1145/2677758.2677780.

[124] Germán Gálvez-Garcıa, Marion Hay, and Catherine Gabaude. “Alleviating simulator sickness with galvanic cutaneous stimulation”. In: *Human factors* 57.4 (2015), pp. 649–657.

[125] Rebecca J Reed-Jones et al. “Can galvanic vestibular stimulation reduce simulator adaptation syndrome?” In: *Driving Assessment 2007: 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (2007).

[126] Michael J Cevette et al. “Oculo-vestibular recoupling using galvanic vestibular stimulation to mitigate simulator sickness”. In: *Aviation, space, and environmental medicine* 83.6 (2012), pp. 549–555.

[127] Séamas Weech, Sophie Kenny, and Michael Barnett-Cowan. “Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review”. In: *Frontiers in Psychology* 10 (2019), p. 158. DOI: 10.3389/fpsyg.2019.00158.

[128] Max Weber. “The definition of sociology and social action”. In: *Max Weber: Economy and Society, Berkeley, CA: University of California Press* 1978 (1921), pp. 4–26.

[129] Godwin Emmanuel Oyedokun. “Management Thoughts: The Review of Social Action Theory”. In: *SSRN Electronic Journal* (2016). ISSN: 1556-5068. DOI: 10.2139/ssrn.2912274.

[130] Christina L Sieloff and Patricia R Messmer. “And Middle-Range Theory of Goal Attainment”. In: *Nursing Theorists and Their Work-E-Book* (2017), p. 213.

[131] Kristopher J. Blom and Steffi Beckhaus. “The design space of dynamic interactive virtual environments”. In: *Virtual Reality* 18.2 (2014), pp. 101–116. DOI: 10.1007/s10055-013-0232-y.

[132] Mark O Riedl and Andrew Stern. "Failing believably: Toward drama management with autonomous actors in interactive narratives". In: *International Conference on Technologies for Interactive Digital Storytelling and Entertainment*. Springer. 2006, pp. 195–206.

[133] Henrik Warpefelt. "Breaking immersion by creating social unbelievability". In: *The Society for the Study of Artificial Intelligence and the Simulation of Behaviour*. 2013, p. 4.

[134] Aaron B Loyall. *Believable Agents: Building Interactive Personalities.* Tech. rep. CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE, 1997.

[135] James C Lester and Brian A Stone. "Increasing believability in animated pedagogical agents". In: *Proceedings of the first international conference on Autonomous agents*. 1997, pp. 16–21.

[136] Patrick Doyle. "Believability through context using" knowledge in the world" to create intelligent characters". In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*. 2002, pp. 342–349.

[137] Masahiro Mori, Karl F MacDorman, and Norri Kageki. "The uncanny valley [from the field]". In: *IEEE Robotics & Automation Magazine* 19.2 (2012), pp. 98–100.

[138] Chen-Yen Peng and Rung-Chin Chen. "Voice recognition by Google Home and Raspberry Pi for smart socket control". In: *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*. IEEE. 2018, pp. 324–329.

[139] Dirk Heylen et al. "Controlling the gaze of conversational agents". In: *Advances in natural multimodal dialogue systems*. Springer, 2005, pp. 245–262.

[140] Marcus Thiebaux, Brent Lance, and Stacy Marsella. "Real-time expressive gaze animation for virtual humans". In: *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. 2009, pp. 321–328.

[141] Björn Hartmann, Maurizio Mancini, and Catherine Pelachaud. "Implementing expressive gesture synthesis for embodied conversational agents". In: *International Gesture Workshop*. Springer. 2005, pp. 188–199.

[142] Christos Kyrlitsias et al. "Social Conformity in Immersive Virtual Environments: The Impact of Agents' Gaze Behavior". In: *Frontiers in Psychology* 11 (2020). DOI: 10.3389/fpsyg.2020.02254.

[143] Erica E. Nason et al. "Virtual treatment for veteran social anxiety disorder: A comparison of 360° video and 3D virtual reality". In: *Journal of Technology in Human Services* 38.3 (2019), pp. 288–308. DOI: 10.1080/15228835.2019.1692760.

[144] Eleonora Brivio et al. "Virtual reality and 360° panorama technology: a media comparison to study changes in sense of presence, anxiety, and positive emotions". In: *Virtual reality and 360° panorama technology* (2020). ISSN: 1359-4338, 1434-9957. DOI: 10.1007/s10055-020-00453-7.

[145] Patricia Bianca Lyk et al. "Co-Designing an Immersive and Interactive Alcohol Resistance Training Tool Using 360-Degree Video". In: *Proceedings of the 12th European Conference on Game Based Learning*. ACPI, 2019, p. 53. ISBN: 978-1-912764-38-9. DOI: 10.34190/GBL.19.081.

[146] Juan Luis Higuera-Trujillo, Juan López-Tarruella Maldonado, and Carmen Llinares Millán. "Psychological and Physiological Human Responses to Simulated and Real Environments: A Comparison between Photographs, 360° Panoramas, and Virtual Reality". In: *Applied Ergonomics* 65 (2017), pp. 398–409. ISSN: 00036870. DOI: 10.1016/j.apergo.2017.05.006.

[147] R. Skarbez, F. Brooks, and M. Whitton. "Immersion and Coherence: Research Agenda and Early Results". In: *IEEE Transactions on Visualization and Computer Graphics* (2020), pp. 1–1. ISSN: 1941-0506. DOI: 10.1109/TVCG.2020.2983701.

[148] Aitor Rovira. "The use of virtual reality in the study of people's responses to violent incidents". In: *Frontiers in Behavioral Neuroscience* (2009). ISSN: 16625153. DOI: 10.3389/neuro.08.059.2009.

[149] R. Skarbez et al. "A Psychophysical Experiment Regarding Components of the Plausibility Illusion". In: *IEEE Transactions on Visualization and Computer Graphics* 23.4 (2017), pp. 1369–1378. ISSN: 1941-0506. DOI: 10.1109/TVCG.2017.2657158.

[150] Susanne Putze et al. "Breaking The Experience: Effects of Questionnaires in VR User Studies". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20: CHI Conference on Human Factors in Computing Systems. Honolulu HI USA: ACM, 2020, pp. 1–15. DOI: 10.1145/3313831.3376144.

[151] Martin Feick et al. "The Virtual Reality Questionnaire Toolkit". In: *The 33st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*. UIST '20 Adjunct. event-place: Minneapolis, Minnesota, USA. New York, NY, USA: Association for Computing Machinery, 2020. ISBN: 978-1-4503-7515-3. DOI: https://doi.org/10.1145/3379350.3416188.

[152] Georg Regal et al. "VRate: A Unity3D asset for integrating subjective assessment questionnaires in virtual environments". In: *2018 tenth international conference on quality of multimedia experience (QoMEX) (QoMEX 2018)*. Sardinia, Italy: QoMEX, 2018.

[153] Richard Skarbez, Frederick P. Brooks, and Mary C. Whitton. "Immersion and coherence in a stressful virtual environment". In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. VRST '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–11. ISBN: 978-1-4503-6086-9. DOI: 10.1145/3281505.3281530.

[154] Michael Meehan et al. "Physiological measures of presence in stressful virtual environments". In: *Acm transactions on graphics (tog)* 21.3 (2002), pp. 645–652. DOI: 10.1145/566654.566630.

[155] Delphine Caruelle et al. "The use of electrodermal activity (EDA) measurement to understand consumer emotions – A literature review and a call for action". In: *Journal of Business Research* 104 (2019), pp. 146–160. ISSN: 0148-2963. DOI: 10.1016/j.jbusres.2019.06.041.

[156] Fred Shaffer and J. P. Ginsberg. "An Overview of Heart Rate Variability Metrics and Norms". In: *Frontiers in Public Health* 5 (2017), p. 258. ISSN: 2296-2565. DOI: 10.3389/fpubh.2017.00258.

[157] Mel Slater et al. "Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment". In: *Presence: Teleoperators and Virtual Environments* 15.5 (2006), pp. 553–569. ISSN: 1054-7460. DOI: 10.1162/pres.15.5.553.

[158] Alexandra Ghiţă et al. "Craving and Anxiety Responses as Indicators of the Efficacy of Virtual Reality-Cue Exposure Therapy in Patients Diagnosed with Alcohol use Disorder". In: *Annual Review of CyberTherapy and Telemedicine* 17 (2019).

[159] Alexandra Ghiţă et al. "Cue-Elicited Anxiety and Alcohol Craving as Indicators of the Validity of ALCO-VR Software: A Virtual Reality Study". In: *Journal of Clinical Medicine* 8.8 (2019), p. 1153. ISSN: 2077-0383. DOI: 10.3390/jcm8081153.

[160] Alexandra Ghiţă et al. "Attentional Bias, Alcohol Craving, and Anxiety Implications of the Virtual Reality Cue-Exposure Therapy in Severe Alcohol Use Disorder: A Case Report". In: *Frontiers in Psychology* 12 (2021). DOI: 10.3389/fpsyg.2021.543586.

[161] José Guardia Serecigni et al. "[Validation study of the Multidimensional Alcohol Craving Scale (MACS)]". In: *Medicina Clinica* 123.6 (2004), pp. 211–216. ISSN: 0025-7753. DOI: 10.1016/s0025-7753(04)74463-2.

[162] Ghiţă, Alexandra et al. "Attentional bias assessment in patients with alcohol use disorder: an eye-tracking study". In: *Annual Review of CyberTherapy and Telemedicine* 17 (2019), pp. 83–87. ISSN: 1554-8716. URL: http://hdl.handle.net/2445/175589.

[163] Stephen T. Tiffany. "A cognitive model of drug urges and drug-use behavior: Role of automatic and nonautomatic processes." In: *Psychological Review* 97.2 (1990), pp. 147–168. ISSN: 1939-1471, 0033-295X. DOI: 10.1037/0033-295X.97.2.147.

[164] B. L. Carter and S. T. Tiffany. "Meta-analysis of cue-reactivity in addiction research". In: *Addiction (Abingdon, England)* 94.3 (1999), pp. 327–340. ISSN: 0965-2140.

[165] Ignacio Jauregui. "Neurophysiological Basis of Food Craving". In: *State of the Art of Therapeutic Endocrinology*. Ed. by Sameh Magdeldin. InTech, 2012. ISBN: 978-953-51-0772-9. DOI: 10.5772/48717.

[166] William Stone III. "Psychometric evaluation of the Simulator Sickness Questionnaire as a measure of cybersickness". In: *Graduate Theses and Dissertations* (2017). DOI: 10.31274/etd-180810-5050.

[167] Hyun K. Kim et al. "Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment". In: *Applied Ergonomics* 69 (2018), pp. 66–73. ISSN: 0003-6870. DOI: 10.1016/j.apergo.2017.12.016.

[168] Jang-Han Lee et al. "Cue-Exposure Therapy to Decrease Alcohol Craving in Virtual Environment". In: *CyberPsychology & Behavior* 10.5 (2007), pp. 617–623. DOI: 10.1089/cpb.2007.9978.

[169] Jessica Simon et al. "Alcohol Craving in Heavy and Occasional Alcohol Drinkers After Cue Exposure in a Virtual Environment: The Role of the Sense of Presence". In: *Frontiers in Human Neuroscience* 14 (2020). DOI: 10.3389/fnhum.2020.00124.

[170] Joy Stradford et al. "Conducting a VR Clinical Trial in the Era of COVID-19". In: *Frontiers in Virtual Reality* 2 (2021), p. 639478. ISSN: 2673-4192. DOI: 10.3389/frvir.2021.639478.

[171] Ewa Siedlecka and Thomas F Denson. "Experimental methods for inducing basic emotions: A qualitative review". In: *Emotion Review* 11.1 (2019), pp. 87–97.

[172] Stephanie R Fishel, Eric R Muth, and Adam W Hoover. "Establishing appropriate physiological baseline procedures for real-time physiological measurement". In: *Journal of Cognitive Engineering and Decision Making* 1.3 (2007), pp. 286–308.

[173] Empatica. *Empatica E4 [Physiological Measuring Device]*. 2021. URL: https://www.empatica.com/en-int/research/e4/.

[174] Empatica. *Empatica E4 Wristband Data Support*. 2021. URL: https://support.empatica.com/hc/en-us/sections/200582445-E4-wristband-data.

[175] Wolfram Boucsein. *Electrodermal Activity*. Boston, MA: Springer US, 2012. ISBN: 978-1-4614-1125-3. DOI: 10.1007/978-1-4614-1126-0.

[176] Luca Menghini et al. "Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions". In: *Psychophysiology* 56.11 (2019). ISSN: 0048-5772, 1469-8986. DOI: 10.1111/psyp.13441.

[177] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. "Publication recommendations for electrodermal measurements: Publication standards for EDA". In: *Psychophysiology* 49.8 (2012), pp. 1017–1034. ISSN: 00485772. DOI: 10.1111/j.1469-8986.2012.01384.x.

[178] Unity. *Unity [Software]*. 2019. URL: https://unity3d.com/unity.

[179] Steam. *SteamVR Framework [Software]*. 2019. URL: https://store.steampowered.com/app/250820/SteamVR/.

[180] Google. *Resonance Audio - Unity*. 2018. URL: https://resonance-audio.github.io/resonance-audio/develop/unity/getting-started.

[181] Glycon3D. *Glycon3D - Instant Motion Capture*. URL: www.glycon3d.com.

[182] LipSyncPro. *LipSync Pro | Animation Tools | Unity Asset Store*. URL: https://assetstore.unity.com/packages/tools/animation/lipsync-pro-32117.

[183] Humelo Inc. and Byung-Ju Choi. *LAVSTAR(Lip-Audio-Visualization-STAR)*. Unity Asset Store. 2020. URL: j.mp/lavstar.

[184] Mathias Benedek. "Analysis of EDA data using Ledalab". Presentation. University of Graz, Austria, 2010. URL: http://www.ledalab.de/download/Analysis%20of%20EDA%20data%20using%20Ledalab.pdf.

[185] Mathias Benedek and Christian Kaernbach. "A continuous measure of phasic electrodermal activity". In: *Journal of Neuroscience Methods* 190.1 (2010), pp. 80–91. ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2010.04.028.

[186] Jegama. *Jegama/empaticaHRV*. 2019. URL: https://github.com/Jegama/empaticaHRV.

[187] Sarah Nichols and Harshada Patel. "Health and safety implications of virtual reality: a review of empirical evidence". In: *Applied ergonomics* 33.3 (2002), pp. 251–271.

Appendix A

Consent form for the preliminary experiment.

The experiment investigates which parts of a virtual environment contribute to how real it feels to you. We have designed a virtual bar where you have to adjust the quality of different aspects of the environment you deem most important to make it feel real for you. The duration of the experiment will be 20 minutes approximately.

We ask for your permission to:
- collect anonymous data such as gender, age, and your prior experience with computer games and VR.
- equip you with VR gear and record your actions inside the virtual bar, also by recording the screen.

All data remains anonymous and will be used only for analysis in this project and later eventual publication of a research paper. You can withdraw from the study at any time during the experiment. You may also ask questions at any time. Thank you for participating in this experiment!

Appendix B

Pre-experience questionnaire for the preliminary experiment.

Pre-experience Questionnaire

1) What is your age?

☐ 18-25 ☐ 26-35 ☐ 36-45

☐ 46-55 ☐ 56+

2) What gender do you identify as?

☐ Male ☐ Female ☐ Other

☐ Prefer not to say

3) How many times have you played videogames involving virtual 3D environments?
(Examples of such videogames are Grand Theft Auto, Call of Duty, World of Warcraft, Minecraft, The Sims, etc.)

☐ Never ☐ Few times ☐ Many times

4) How many times have you tried VR before?

☐ Never ☐ Few times ☐ Many times

Post-experience VRSQ questionnaire for the two experiments.

Post experiment

Please select to which degree you experience the following symptoms right now: *

	None	Slight	Moderate	Severe
General discomfort	○	○	○	○
Fatigue	○	○	○	○
Eyestrain	○	○	○	○
Difficulty focusing	○	○	○	○
Headache	○	○	○	○
Fullness of head	○	○	○	○
Blurred vision	○	○	○	○
Dizzy (eyes closed)	○	○	○	○
Vertigo	○	○	○	○

Which aspects of the environment did you pay most attention to? *

Your answer

Which parts of the environment did you find most important to include to feel like reality? *

Your answer

Appendix C

An extended step-by-step procedure for *Experiment A* including the script used for instructing and guiding participants.

- Step 1: The test participant enters the laboratory and is greeted and thanked by both researchers for participating in the study, and INSTRUCTOR explains the role of each researcher, and then explains the overall premise of the study with the following script:

  [DA]
  "Vi prøver i dette studie at finde ud af hvilke dele af et virtuelt miljø, der bidrager mest til hvor virkeligt man oplever miljøet. Med andre ord undersøger vi hvilke aspekter i et virtuelt miljø, der gør at det føles virkeligt. For at undersøge det, har vi lavet en bar i virtual reality, der er lavet efter en virkelig bar i København. I baren er der både musik, en bartender, andre gæster, og hvad man nu ellers finder i en typisk bar. Vi har designet baren på en måde der gør at vi kan justere kvaliteten på forskellige dele af baren. Så måden vi vil finde ud af hvad der, forhåbentlig, gør at baren føles virkelig, er ved at du her om lidt skal justere de dele som du føler er vigtigt for hvor virkelig baren føles for dig. Men det skal vi nok komme mere i dybden med om lidt."

  [EN]
  "In this study, we are investigating which parts of a virtual environment contribute the most to how real the environment is experienced. In other words, we are examining which aspects in a virtual environment that makes it feel real. To examine this, we made a bar in virtual reality, which is a recreation of a real bar in Copenhagen. In the bar there is both music, a bartender, other guests, and anything else commonly found in a typical bar. We have designed the bar in a way that allows us to adjust the quality of different aspects of the bar. So the way we want to figure out what, hopefully, makes the bar feel real, is that, in a moment, you must adjust the aspects that you feel are important for how real the bar feels to you. But we will explain more about that in detail in a moment."

- Step 2: The test participant is asked to sign the consent form and questionnaire in the *Questionnaires* area on Fig. 1:

  [DA]
  "Inden vi starter experimentet bedes du venligst underskrive denne samtykkeærklering og svare på det korte spørgeskema på bærbaren herovre."

  [EN]
  "Before we start the experiment, we would ask you to sign this written consent form and answer the short questionnaire on the laptop over here."

- Step 3: After answering the questionnaire, INSTRUCTOR shows the demonstration video on the laptop in the *Questionnaires* area on Fig. 1, continuing explaining from the script:

  [DA]
  "Vi snakkede før om, at vi har taget dele af baren og gjort så man kan justere kvaliteten på dem. Vi kalder disse dele for parametre, og der er i alt fire parametre man kan justere kvaliteten på, og de fire parametre har vi valgt at kalde *animationer*, *reaktioner*, *lyd* og *grafik*. Du vil nu se en video, der demonstrerer de forskellige kvaliteter for hvert parameter."

  [EN]
  "Before we talked about how we have selected different parts of the bar and made it so you can adjust the quality of them. We refer to these as *parameters*, and there are a total of four parameters you can adjust the quality of, and those four parameters are *animations*, *reactions*, *sound*, and *grahpics*. You will now watch a video demonstrating each level of each parameter."

- Step 4: INSTRUCTOR answers potential questions and then guides the test participant to the table in the *VR Experience* area on Fig. 1, continuing explaining from the script:

  [DA]
  "Du skal nu sætte dig på stolen ved bordet og får derefter VR udstyr på. Når du har fået VR udstyret på skal vi gennemgå de parametre du lige har set i videoen, og du skal selv prøve at justere dem, så du ved hvordan det fungerer før selve eksperimentet starter. Måden du justerer en parameters kvalitet er ved at sige det højt til os, og så vil TECHNICIAN justere dem med det samme. Dvs. du f.eks. kan sige opgradér animationerne, eller opgradér lyden, eller opgradér interaktionen. Har du nogle spørgsmål til det?"

  [EN]
  "You will now sit on the chair by the table and afterwards get the VR equipment put on. When you have put on the VR equipment, we will go through the parameters you have just seen in the video, and you have to adjust them yourself so you know how it works before the experiment itself starts. The way you adjust a parameter's quality is by saying it aloud to us, and then TECHNICIAN will adjust them immediately. That means that you for example can say upgrade the animations, or upgrade the sound, or upgrade the interactions. Do you have any questions regarding that?"

- Step 5: The participant equips themselves with the VR apparatus. INSTRUCTOR instructs the participant in how to use the controllers and HMD for interacting with objects and looking around:

  [DA]
  "I VR kan du kigge rundt hvis du bare drejer dit hoved, og controllerne du har i hænderne styrer dine hænder i VR. Hvis du trykker på triggeren med din pegefinger kan du knytte hånden, og hvis du gør det imens din hånd er tæt på f.eks. menukortet, så kan du tage fat i det... Føler du, du har styr på hvordan det fungerer, eller har du nogle spørgsmål?"

  [EN]
  "In VR you can look around just by turning your head, and the controllers you have in your hands controls your hands in VR. If you squeeze the trigger with your index finger you can close your virtual hand, and if you do it when your hand is close to for example the menu card, you can grab it... Do you feel like you know how it works or do you have some questions?"

- Step 6: Upon familiarization with the VR environment, TECHNICIAN resets the virtual environment to a random starting configuration, and INSTRUCTOR continues explaining from the script:

  [DA]
  "Nu vil jeg forklare selve eksperimentet - hvis du har spørgsmål undervejs så spørg endelig. Eksperimentet går ud på at alle parametrene først vil være sat til en tilfældig kvalitet - vi kalder dette miljø for reference-miljøet. Her vil jeg bede dig om at bruge to minutter på at lægge mærke til og huske hvor virkelig du føler denne virtuelle verden er, og du må gerne bruge dine hænder og kigge rundt imens. Grunden til du skal lægge mærke til og huske hvor virkeligt du føler miljøet er, er fordi at bagefter, så bliver alle parametrene sat til en ny tilfældig kvalitet, og din opgave bliver at opgradere de parametre du føler bidrager mest til at matche, eller opnå samme virkeligheds-følelse, som i reference-miljøet. Der vil ligge en seddel foran dig på bordet i baren med alle 5 parametre så de er nemmere at huske. Hver gang du har lavet en opgradering spørger jeg dig om du føler miljøet matcher den følelse af virkelighed du havde i reference-miljøet. Hvis ikke, så skal du foretage endnu en opgradering du føler vil bidrage mest til at opnå et match, og du må gerne opgradere det samme parameter flere gange i træk. Denne matching-opgave laver vi 6 gange i alt, hvor du starter med et tilfældigt udgangspunkt hver gang. Giver det mening hvad opgaven går ud på?"

  [EN]
  "I will now explain the experiment itself - if you have any questions along the way, feel free to ask. "In this experiment, all parameters in the environment are first assigned a random level - we call this environment for the reference-environment. Here, I will ask you to spend some time noticing and remembering how real you feel this virtual environment is, and you are free to use your hands and look around during this. The reason you must pay attention to and remember how real you feel the environment is, is because afterwards, all the parameters will be set to a new random quality, and your task becomes to upgrade the parameters you feel contributes the most to matching or reaching the same sense-of-reality as in the reference-environment. There will be a piece of paper in front of you on the table in the bar with the names of all 4 parameters so they are easier to remember. Every time you have made an upgrade, I will ask you if you feel the environment matches the feeling of reality you had in the reference-environment. If not, then you are asked to make another upgrade that you feel will contribute the most to reaching a match, and you are free to upgrade the same parameter multiple times in a row. We will perform this matching-task 6 times in total, where you start with a random starting point every time. Does the task make sense to you?"

- Step 7: After the participant has performed the matching-task five times, the test conductor thanks the participant for taking part in the experiment.

Appendix D

Raw data from the preliminary post-experiment subjective assessment.

Which aspects of the environment did you pay most attention to?
Sound and human animations.
The sound, human reactions and the things you could grab.
Background chatter, football on TV, loud music.
Sounds and human animations
The sound was what I noticed first. But maybe that was because it faded in before the visuals.
Sound, human reaction
The person in front of me, and the sound.
The sound and the mood, the people, the reactions and the movement.
The table in front of me and the person in front of me.
Voices, music, the man, beer.
The man sitting across from me - The bar - My hands - The sounds.
The sounds as well as the person in front of me and how the people in the background moved.
Sounds and movements.
Visuals - behaviour.
The lady behind the bar with Fawsett hair, the ridiculously muscular man behind me, the only black lady, the man with the empty glass who wouldn't stop staring at me
Sound and the visuals.
The person sitting across from me, how he reacted when you grabbed the menu and the beer.
The television and the drink menu.
The screens with the football match.
Sound.
Sound, the silence was conspicuous, but equally it was most recognisable from reality, in the scenarios where it was already present.
Mostly the person that was in front of me because he was the one I interacted with the most. And then also the sound because it was a bit awkward when it was silent and the sound added some nice ambience.
The sound.
The sounds of the bar, the overall visual quality, and the movement of the other people
The character sitting in front of me and the items I picked up and the sounds.

Which parts of the environment did you find most important to include to feel like reality?
Sound and human animations
Sound, human reactions and the things you could grab.
Human animations and sound.
Sounds, human animations and human reactions
The sound and the animations. I had a hard time seeing what happened when you upgraded the visuals.
In order from most important to least important - Sound, human reactions, animations, visuals
The sound, as well as the human animations.
The sound, the people, the movement.
The person in front of me.
Voices, music and beer.
The guy sitting across from me - the sounds.
The small movements of people and the small sounds in the background of glass for example.
Realistic surroundings, sounds and movements.
A combination of all. If some is missing, it affects the other aspects.
I would say the things I directly interact with up close are the most important, both with sound and general interaction.
Sound.
Reactions, sounds and animations.
The human atmosphere (other tables).
The sound of conversation and music.
Sound.
Sound first, and human animations second. If humans move but do not emit sound, they appear artificial.
I think all of them were pretty important, but especially the human animation because otherwise everything would be frozen and it wouldnt feel real. But tbh its a difficult choice because i think it feels most real when all of them are upgraded and working together.
The sound and reactions give me a great feeling of being there.
The sounds, the visual quality and the movement of humans.
Without the sounds the environment was empty and felt unreal, but the amount of sounds were not a big difference. The visual quality of the character in front of me was also very noticeable I think. The more detailed the animations were the most lively the bar was.

Appendix E

An extended step-by-step procedure for *Experiment B* including the script used for instructing and guiding participants.

- Step 1: The test participant enters the laboratory and is greeted and thanked by both researchers for participating in the study, and INSTRUCTOR explains the role of each researcher, and then explains the overall premise of the study with the following script:

  [DA]
  "Vi undersøger i dette studie hvordan vi kan gøre en virtuel bar mest effektiv i form af troværdigheden af den og i form af hvor meget den kan give folk lyst til alkohol. I den forbindelse har vi designet 8 forskellige bar-scener som du her om lidt skal opleve i VR. Men inden vi starter experimentet vil vi bede dig om venligst at underskrive denne samtykkeærklering og svare på det korte spørgeskema på bærbaren herovre."

  [EN]
  "In this study, we are investigating how we can make a virtual bar more effective in terms of believability and in terms of to which degree it can make users want alcohol. To examine this, we have designed 8 different bar scenes which you will experience in VR in a moment. But before we start the experiment, we would ask you to sign this written consent form and answer the short questionnaire on the laptop over here."

- Step 2: After signing the consent form and answering the questionnaire, INSTRUCTOR answers potential questions and then guides the test participant to the table in the, continuing explaining from the script:

  [DA]
  "Som du også læste af samtykkeerklæringen, måler vi noget fysiologisk data, som f.eks. din hjerterytme, og til det skal vi have dig til at tage dette armbånd på, på din ikke-dominerende hånd."

  [EN]
  "As you read from the consent form, we are measuring some physiological data such as your heart rate, and to do so, we would like you to equip this wristband on your non-dominant hand."

- Step 3: INSTRUCTOR helps the participant equip the Empatica E4 wristband appropriately, then continuing explaining from the script:

  [DA]
  "Det er vigtigt du sidder stille med den arm og hånd som armbåndet sidder på under hele eksperimentet. Som du kan se, er der en knap [the *Event Marking* button] her – den vil jeg trykke på før og efter hver scene imens du er i VR, men det skal du ikke tænke på. Men som sagt har vi altså lavet 8 bar-scener som du nu skal opleve i tilfældig rækkefølge, og mellem hver bar-scene vil du opleve en skov-scene, hvor du egentlig bare skal slappe af – det er for at kalibrere armbåndet. Sidst men ikke mindst vil du efter hver scene blive stillet 3 spørgsmål som du skal besvare med et tal fra 0 til 100.

  [EN]
  "It is important that you try not to move your arm and hand with the wristband on it during the entire experiment. As you can see, there is a button [the *Event Marking* button] here - I will press it before and after each scene while you are in VR, but you should not think about that. So... as mentioned earlier, we have designed 8 bar scenes which you will now experience in a random order, and between each bar scene there will be a forest scene, in which you should simply relax - this is to calibrate the wristband. Last but not least, after each scene you will be asked 3 questions which you should answer with a number between 0 and 100.

- Step 4: INSTRUCTOR explains to the participant how to use the VAS, showing a printed paper of it simultaneously:

  [DA]
  For at vælge tallet skal trykke på pilene med din virtuelle hånd som du styrer ved bare at bevæge denne controller, og så skal du trykke på 'OK'-knappen for at fortsætte. Spørgsmålene vil være på engelsk, så for at sikre at du forstår spørgsmålene vil jeg lige præsentere dig for dem her.

  [EN]
  To select the number, you must press the arrows with your virtual hand, which is controlled by simply moving this controller, and you must press the 'OK' button to continue. To make sure you understand the questions, I will now present them to you.

- Step 5: INSTRUCTOR shows the questions to the participant on a printed paper with the VAS containing the questions and makes sure the participant understands them. INSTRUCTOR then continues from the script:

  [DA]
  "Så, helt enkelt forklaret er din opgave sådan set bare at opleve det der nu sker i bar-scenerne og svare på de 3 spørgsmål mellem hver scene – og husk at vi ikke tester dig, så vi vil sætte pris på hvis du svarer 100% ærligt på spørgsmålene og kun forholder dig til én scene ad gangen, og ikke tænker over hvad du har svaret på tidligere spørgsmål. Og jeg vil lige minde dig om, at du skal prøve at sidde stille med armen med armbåndet på gennem eksperimentet. Har du nogle spørgsmål?"

  [EN]
  "So, simply put, your task is to experience whatever happens in the bar scenes and answer the 3 questions between each scene – remember that we are not testing you, so we would appreciate if you answer the questions with 100% honesty, and only relate to one scene at a time, and not think about your answers to previous questions. And I will just remind you to try to not move your arm with the wristband on during the experiment. Do you have any questions?"

- Step 6: The participant is assisted in equipping the VR equipment.

- Step 7: While the participant is placed in the Neutral scene, INSTRUCTOR monitors the physiological data. Once a baseline reading is obtained, INSTRUCTOR signals TECHNICIAN, who will then progress the VRE to the VAS scene, where the participant reports their craving level and then their PI/PSI level. From here, the participant experiences one of the eight bar scenes, after which they answer the three questions in the VAS scene again, before returning to the Neutral scene. This cycle is repeated 8 times - until the participant has experienced all bar scenes - however, after experiencing the final bar scene, the participant does not return to the Neutral scene, but instead takes off the VR equipment and the Empatica E4 wristband.

- Step 8: Finally, the participant is asked the following question, which is recorded by a microphone:

  [DA]
  "Som du nok lagde mærke til var nogle af bar scenerne lavet ved en video-optagelse og de andre lavet udelukkende af virtuelle ting. Forestrækker du den ene version over den anden, og i så fald, hvorfor?"

  [EN]
  "As you probably noticed, some of the bar scenes were constructed by video recordings and the other scenes contructed entirely by virtual objects. Do you prefer one version over the other, and if so, why?"

Appendix F

Pre-experience questionnaire for the primary experiment.

Pre-experience Questionnaire

1) What is your age?

☐ 18-25 ☐ 26-35 ☐ 36-45

☐ 46-55 ☐ 56+

2) What gender do you identify as?

☐ Male ☐ Female ☐ Other

☐ Prefer not to say

3) How many times have you used VR previously?

☐ Never ☐ Few times ☐ Many times

4) Do you consume alcohol?

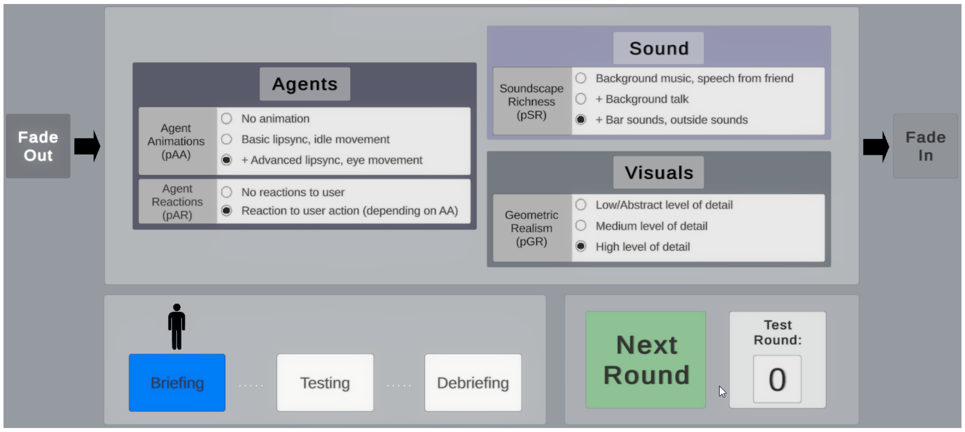
☐ Yes ☐ No

5) Based on the previous 6 months, on average, how many standard drinks of alcohol do you consume on a weekly basis?
(1 standard drink is a can of beer, a glass of wine, or 4 cl of hard liquor)

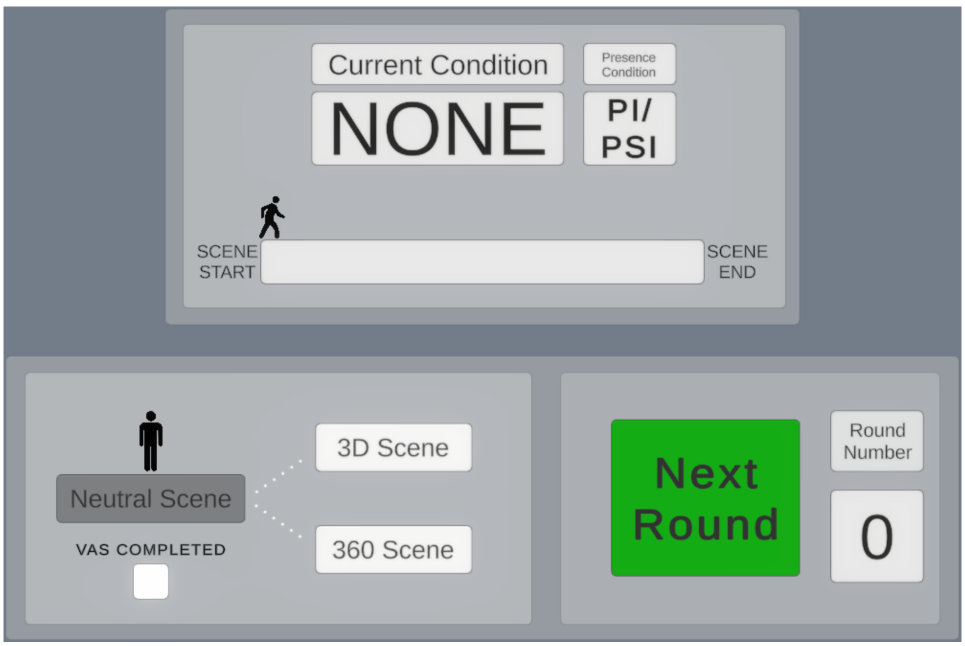
☐ Never ☐ Few times ☐ Many times

Appendix G

UIs created for the operator to control the participant experiences in both experiments.



UI that was used during Experiment A.



UI that was used during Experiment B.