Towards Gamified Alcohol Use Disorder Therapy in Virtual Reality: A Preliminary Usability Study

Fariba Mostajeran * Aila Kirsten † Frank Steinicke ‡ Jürgen Gallinat § Simone Kühn ¶
Universität Hamburg Universität Hamburg Universität Hamburg University Medical Center Hamburg-Eppendorf
University Medical Center Hamburg-Eppendorf
Max Planck Institute for Human Development
Lise Meitner Group for Environmental Neuroscience

ABSTRACT

The combination of virtual reality (VR) and gamification opens up new vistas for innovative forms of therapy for alcohol use disorder (AUD) and have enormous potential to improve traditional therapy methods. In this paper, three gamified and one non-gamified AUD therapy applications for VR are introduced and evaluated. The games are based on two behavioral therapy methods, which are Cue Exposure Therapy (CET) and Approach Avoidance Training (AAT). The games are realized in the context of a virtual supermarket, which is considered as a relapse-risky environment. The aim is to help AUD patients practice avoiding alcohol first in a VR-based simulation and later in a real supermarket. In preparation for a long-term clinical study, a usability study was conducted with 13 healthy participants. The results show that the VR game was enjoyed, increased the motivation, and fewer errors were made than in the comparable non-gamified application.

Keywords: Virtual reality therapy, alcohol use disorder, approach avoidance training, cue exposure therapy.

Index Terms: Human-centered computing—Interaction paradigms—Virtual Reality; Applied Computing—Life and medical sciences—consumer health

1 INTRODUCTION

According to the 2015 National Survey on Drug Use and Health (NSDUH) [1], 15.1 million adults above 18 years old and an estimated 623,000 adolescents ages 12–17 fulfilled the criteria of alcohol use disorder (AUD) in the USA. However, less than 7% of the adults and less than 6% of the adolescents received treatment. More than 85,000 people die from alcohol-related causes annually, making alcohol the third leading preventable cause of death in the US, but other countries are affected by a high alcohol consumption as well [9]. High alcohol consumption on a regular basis is considered risky and can cause serious health problems. For example, approximately half a million hospital stays in Germany resulted from alcohol in 2012. These numbers emphasize the need for innovative alcohol therapy methods, which are much more accessible and effective compared to the traditional ones available today.

Different therapy methods can help patients overcome their AUD. A common behavioral therapy method is called Cue Exposure Therapy (CET). In a therapy session patients are confronted with stimuli that trigger their craving. The goal is to repress salience of those stimuli and to prevent or reduce the dysfunctional approaching behavior [13]. The second often used therapeutic method is Approach Avoidance Training (AAT). Within this method the patient sees multiple drinks as photos on a PC screen and has to push the alcoholic beverages away and pull the non-alcoholic towards himself to learn to avoid alcohol [17]. These methods, however, are associated with some limitations, which have to be addressed while designing innovative therapy methods. For example, therapy sessions using the CET method with the goal to cope with alcohol cues in every day scenarios such as grocery shopping can be expensive and time-consuming. Furthermore, the AAT may be perceived as rather tedious and may not yet be as enjoyable as they it could be.

This shortcoming can be addressed by implementing these behavioral therapy methods in virtual reality (VR). In a virtual environment (VE), craving can be successfully induced and has been shown to be more effective than traditional 2D cues as photographs [8]. Since the VE can easily be customized and adjusted to fit the patient’s needs. In addition, the patient might prefer a VE to an in vivo therapy setting [8]. Furthermore, in a VR session, the patients wear a Head-Mounted Display (HMD) and can immerse themselves in another setting, whereas in reality they are still in a “safe” environment, in which alcohol is not available. Another advantage is that in VR the stimuli (or cues) can be interactive. Finally, gamification has enormous potential to increase the intrinsic motivation of users [16]. Different gamification elements satisfy the key elements that create motivation such as the feeling of competence, experiencing autonomy, and social relatedness [15]. Moreover, recent studies showed that learning effects can be increased through gaming [5], and even more interestingly that gaming on a regular basis can have a positive impact on the structure of the brain [11, 10].

The goal of this study is to develop and evaluate a VR-based game that may potentially enhance alcohol addiction therapy. The VE is a supermarket, which is not only a place which patients will have to visit on a regular basis in their everyday life, but also bears a high risk for a relapse. The patients need to be prepared for this scenario. This summarizes an additional requirement of this game: The VE needs to look and feel as realistic as possible. The game will consist of different mini-games each of which includes one behavioral therapy method. The main goal of the game is to find and buy items that are written on a shopping list. Before paying for the items, the player has to earn money. Earning money can be done through sorting items into a shelf (CET) and clearing a shelf by placing the alcoholic beverages into a trash bin and the non-alcoholic beverages in a shopping-cart (AAT). In preparation for a long-term clinical study, a preliminary study was conducted to evaluate the usability and perceived enjoyment and motivation for training.
2 Background and Related Work

Originally the AAT was developed as a diagnostic instrument and as a variant of an Implicit Association Test (IAT). An IAT measures the relative strength of associations between concepts [19]. The task is to sort the stimuli into four categories but only using two possible responses. Two of the four concepts are called attributes (usually positive-negative) and the other two are the targets. The task is to react as quickly as possible to which category (attribute or target) the stimulus belongs. A stimulus can be an image or word. The measured effect is the different response time for the two possible combinations of attributes and targets. The AAT was implemented as an approach-avoidance-IAT and its results show strong connections between alcohol and approach bias in heavy drinkers [19]. Wiers et al. [18] conducted a clinical AAT training study with 214 alcohol dependent inpatients, who were assigned to either training or control conditions. The control groups either did not receive training or received sham training. The participants in the training group had to push a joystick to make an avoidance movement when pictures of alcohol were presented and to make an approach movement to pictures of soft-drinks (pull a joystick). The results show that four brief sessions (15 minutes) on consecutive days changed implicit approach responses to alcohol. Moreover, a year later the treatment outcomes for patients in the training group was better (i.e., 16% less relapse rate).

Another type of cognitive behavioural therapy method for AUD is CET which is a special form of exposure therapy (ET). In ET, which has been successfully applied for anxiety disorders [13], patients are carefully exposed to specific stimuli which trigger their symptoms. This will overtime help them to gradually control their reactions. Similarly in CET, patients are exposed to certain stimuli (or cues) to trigger their craving (disturbed approach behavior). Exposure to cues increases the craving, which can also be observed physiologically e.g. in a change of the cardiac frequency, electrodermal activity, or the salivation. CET can be used for treatment of various types of substance (e.g., alcohol, nicotine, cocaine, etc.) addiction as well as other psychological disorders such as bulimia, binge eating, shopping addiction, and pathological gambling. For treatment of AUD, alcohol related cues are presented and the patient is instructed to let the craving arise and acknowledge it. The patient gives regularly feedback on his subjective rating of the intensity of the craving. A training session is completed once the intensity is considerably decreased. In later therapy sessions CET can be combined with a training of coping skills, so the patient learns how to decline alcohol offers [13].

In addition to these classical therapy methods, some scientists have reported applying VR for AUD therapy [3, 12, 14]. For example, Brodnick et al. [3] analyzed the subjective craving for alcohol in different VEs. The results of a controlled experiment with 40 AUD patients showed that craving for alcohol was increased in VEs in which alcoholic beverages were present (e.g., party). These results can be helpful for VR-CET and in the study [12] applied VR-CET to eight members of an Alcoholics Anonymous group. The VEs were a Japanese-style pub and a western-style bar. The results of training for eight 30min-sessions showed a decrease in the subjectively reported craving. Lastly, Metcalfe et al. [14] developed a cue refusal VR-video game based on Kinect and Xbox with the goal to support cigarette and alcohol recovery. The player had to hit or kick the addiction cue images which flew towards them. The results of an experiment with 61 participants recovering from alcohol or tobacco addiction showed that on average, reported substance use decreased during the intervention period. Moreover, AUD participants in recovery showed a statistically significant increase in self-efficacy, attitude, and behavior during the intervention and a decrease in alcohol use by 75% after the study.

3 Design and Implementation

We designed and implemented three mini-games for CET and AAT as well as a non-gamified AAT application to be played in VR. All VEs were implemented using Unity3D game engine and were customized for HTC Vive HMD and controllers. Furthermore, players did not have to move in VR as all games could be played while standing still at one place. The three mini-games were featured in a virtual supermarket with a wide range of grocery items, a typical supermarket background noise, and virtual customers walking around the supermarket. The overall game’s goal was to find and buy the correct items of a shopping list. But before being able to buy any item, the player had to earn money by playing two mini-games of which one was inspired by AAT and the other by CET. All interactions with the virtual items in these VEs were accomplished using the HTC vive controllers (e.g., grabbing an object by pressing the trigger button).

The AAT game (see Figure 1 (a) and (b)) was played in front of an unsorted shelf in the alcohol section of the virtual supermarket. The shelf contains alcoholic and non-alcoholic beverages. The goal was to sort out the items by putting the non-alcoholic beverages into the shopping-cart and throwing the alcoholic ones into a trash bin. When an item was sorted correctly, the player earned 0.5€, the target container was colored in green, and a positive audio feedback is played when a non-alcoholic beverage was successfully thrown into it. Misplacement of an item (wrong container) resulted in losing 0.5€ and receiving negative auditory and visual (wrong container appears in red) feedback. Moreover, there was a time limit, so that the player was offered only a certain amount of time to play this game.

Next, the CET mini-game (see Figure 1(c) and (d)) started with the goal of exposing the players for a longer time with alcoholic beverages. To do so, they had to take one bottle at a time out of a bottle crate next to them and place it in a randomly assigned position in an empty shelf. Taking one bottle at the time implied that the player cannot interact with both hands at the same time. This was implemented to increase the focus onto each single bottle. Each placement resulted in receiving 0.5€. Since the bottles could not be placed anywhere else, losing money was not possible in this mini-game. Also similar to the AAT mini-game, the player was only granted a certain amount of time to play.

The shopping game (see Figure 1(e)) started by the player receiving a shopping-list whose items had to be purchased using the money which has been earned while playing the AAT and CET games. The player is supposed to buy only the items on this list while most of them do not refer to a specific product, but rather a group of items (e.g., fruit, hot drinks, bread, etc.) leaving some freedom of choice (e.g., type of bread). The items in the shelves or fridges can be grabbed and placed in a shopping-cart or taken out of it and back in their original place. If the player puts an item to the shopping-list into the shopping-cart, the shopping-cart was colored in red and a negative sound was played to give an immediate feedback on the player’s action. The shopping-list was also updated accordingly to give the player an overview of the purchased items and their prices. If the player puts an item into the cart, which was not on the list, the shopping-cart was colored in red and a negative sound was played to give an immediate feedback on the player’s action. The shopping-list was also updated accordingly to give the player an overview of the purchased items and their prices. If the player puts an item into the cart, which was not on the list, the shopping-cart was colored in red and a negative sound was played to give an immediate feedback on the player’s action. The shopping-list was also updated accordingly to give the player an overview of the purchased items and their prices.

In addition, a non-gamified AAT VR application (see Figure 1(f)) was developed to be compared with the AAT mini-game. The interaction in this environment was designed to be similar to the AAT implementation of Wiers et al. for AUD therapy [18]. Since we could not find a comparable classical version for the other two mini-games (i.e., CET and shopping) we decided on a non-gamified...
Figure 1: Proposed gamified (a-e) and non-gamified (f) AUD therapy VR applications. The games (a-e) were realized in the context of a virtual supermarket and the non-gamified application in a mountain cabin. (a) AAT mini-game (G-AAT): sorting alcoholic beverages into a trash bin and non-alcoholic beverages into a shopping-cart. (b) CET mini-game: taking alcoholic beverages from a bottle crate and placing them on the marked positions. (c) Shopping mini-game: the items on a shopping list should be found and put into a shopping-cart. The names of the correct items were shown in green and incorrect ones (i.e., alcoholic beverages) in red on the shopping list. (d) Non-gamified AAT (N-AAT): alcoholic beverages were supposed to be pushed away and non-alcoholic beverages were to be pulled closer to the participant.

Implementation only for the AAT mini-game. This version was situated in a mountain cabin with a minimal interior design to keep the VE neutral. The bottles appeared at the center of a table. The controllers in this VE appeared as two virtual hands with white gloves. The instructions in this game were to push the alcoholic beverages away and pull the non-alcoholic beverages closer. Since the bottles were not implemented to be grabbed, the participant was only able to interact with the bottles by touching and then moving them either forwards or backwards. For each bottle, the participant had two seconds to react. After a short break of two seconds, a new bottle appeared at the center of the table.
4 User Study

In preparation for a long-term clinical study, we performed a pilot user study, which was approved by the local ethical committee of the Computer Science Department. The aim of this study was to evaluate the usability of three conditions: (i) Non-Gamified AAT (N-AAT), (ii) Gamified AAT (G-AAT), and the (iii) Whole Game (WG, containing all three mini-games: CET, AAT, and shopping). In addition, the comparison between N-AAT and G-AAT could help us understand the effects of the gamification on performance (i.e., the number of errors made while sorting the alcoholic and non-alcoholic beverages), motivation, and enjoyment of the training within each condition. Moreover, the comparison between G-AAT and WG could give us an insight into the experienced level of enjoyment by playing multiple mini-games which have a story behind them (i.e., earn money by sorting to buy your items on the shopping list). Thus, the following hypotheses were formulated:

- $H_1$: G-AAT is more motivating than N-AAT (due to the gamification element)
- $H_2$: G-AAT produces less errors than N-AAT in sorting alcoholic and non-alcoholic beverages
- $H_3$: WG is enjoyed more than G-AAT (due to change introduced by the collection of mini-games and the story behind it, i.e., earn money and do the shopping. This mission is missing in G-AAT alone, as it is not clear for what purpose the money is collected.)

4.1 Participants

13 healthy participants (9 male and 4 female) between 22 and 35 years of age (avg. 25.76) took part in this study. Half of the participants did not need any visual corrections and of the other half, half used glasses and the other half contact lenses to correct their vision. Most participants have experienced VR before. 61.5% were students of the local Department of Computer Science, who received course credit for their participation. All participants signed a consent form prior to the study. They were also free to have breaks or quit the study at any time.

4.2 Procedure

The study took place in a laboratory room of approx. 16m² with a dim light. The VR tracking space was approximately 3m × 3m. For rendering, system control, and logging an Intel computer running Windows 10 (graphics card: GeForce GTX 780 Ti; processor: Intel Core i7; RAM: 16GB) was used. In addition to the HTC Vive HMD, participants wore DT 770 Pro headphones for sound and noise-canceling. They also received an introduction to the whole study and instructions for each condition. The order of the conditions was randomized. Questionnaires were given before and after the whole study and after each condition. In addition to the questionnaire, during G-AAT and N-AAT, the data on the total number of sorted bottles and committed errors were logged. After a short training phase, the experiment was started.

4.3 Results

In order to examine our hypotheses, we used several questionnaires whose results will be presented in this section. Given that the Shapiro-Wilk test indicated non-normally distributed data, all differences were tested for significance using the non-parametric Wilcoxon Signed-Rank Test with an $\alpha$ level of 0.05. If the variations between pairs were insufficient ($n<10$), the critical value for $W$ was used instead of $Z$ and $p$ to evaluate the significance.

First, we used the System Usability Scale (SUS) [4], which is a standardized survey to evaluate the usability of a system. It consists of 10 Likert Scale items, where different statements can be answered from Strongly Agree to Strongly Disagree. The SUS-score for each condition was calculated as: N-AAT=87.3, G-AAT=93.5, and WG=86.7. A SUS-score can rank from 0-100 and a score above 68 is considered to be above average, meaning that all conditions received above average usability scores [2]. A Wilcoxon Signed-Rank Test ($Z = -0.4077$, $p = 0.6818$) showed no significant differences in the SUS-scores between N-AAT and G-AAT or between G-AAT and WG ($Z = -0.1529$, $p = 0.88076$). Tests at the individual item level showed only for the first statement $I$ think I would like to use this system frequently a significant preference of G-AAT over N-AAT ($W = 0.00$, critical value = 3.00).

We also used AttrakDiff [7] for evaluating usability as well as the user experience. The survey consists of three semantic differential questions each containing 9 or 10 word pairs. With the help of a seven point scale, the participant were asked to choose which word of the word pair was more appropriate to describe the system on four dimensions Pragmatic Quality (PQ), Hedonic Quality-Identity (HQ-I), Hedonic Quality-Stimulation (HQ-S), Attractiveness (ATT). The word-pairs and the mean scores for the three conditions are displayed in Figure 2. The individual items reveal that the participants found G-AAT to be significantly more inventive ($W = -0.4543$, critical value = 9.00), more creative ($Z = -2.4973$, $p = 0.01242$), bolder ($W = 0.00$ critical value = 0.00), more innovative ($W = 0.00$, critical value = 2.00), more captivating ($W = 0.00$, critical value = 5.00), more challenging ($Z = -2.9341$, $p = 0.00338$) and more novel ($Z = -2.8031$, $p = 0.00512$) than N-AAT. Comparing G-AAT to WG showed that WG was perceived as significantly more human ($Z = -2.9025$, $p = 0.00374$) and more simple ($Z = -2.4463$, $p = 0.01428$) than G-AAT. And G-AAT was perceived as more innovative ($W = 0.00$, critical value = 8.00) than WG.

With the help of AttrakDiff, the PQ, HQ-I, HQ-S, and ATT could be calculated for each condition (see Figure 3). PQ describes the usability of a system and how achievable its goals were. Comparing directly N-AAT and G-AAT showed that the participants found N-AAT significantly better at PQ ($Z = -2.4318$, $p = 0.0151$). HQ-I shows how much the users were able to identify with the product. The differences between N-AAT and G-AAT on this dimension were not significant ($Z = -0.4543$, $p = 0.65272$). HQ-S reveals how much they were stimulated by the system and how much they felt the system could support the user in improving. On this dimension the direct comparison between N-AAT and G-AAT showed that the participants were significantly more stimulated by G-AAT than N-AAT ($Z = -3.1798$, $p = 0.00148$). ATT describes the overall attractiveness of the system based on the perceived quality. For ATT no significant difference was found when comparing N-AAT and G-AAT ($Z = -0.4543$, $p = 0.65272$). The comparison of G-AAT and WG only showed a significant difference in PQ ($Z = -2.9003$, $p = 0.00374$).

Furthermore, the NASA Task Load Index (NASA-TLX) [6] was used in this study to evaluate the workload of each condition. This questionnaire is a multidimensional scale containing questions about different aspects. The participants can rate on a 100 point scale in steps of 5 the mental, physical and temporal demand and also their performance, effort and frustration. Since the targeted patient group may be limited in their cognitive or even physical abilities, it is important to evaluate whether or not the developed VR application were too demanding or difficult. The results can be seen in Figure 4. Here the average score and standard deviation is displayed for each item and experimental condition. When comparing G-AAT and N-AAT, multiple significant effects were found: G-AAT has a significantly higher mental demand ($Z = -2.534$, $p = 0.0114$), a significantly higher physical demand ($Z = -3.0594$, $p = 0.00222$), a significantly higher temporal demand ($Z = -2.3142$, $p = 0.02088$) and needs significantly more effort to be put into to accomplish the desired level of performance ($Z = -3.0594$, $p = 0.00222$). The overall demand was significantly
higher in G-AAT ($Z = -2.8241$, $p = 0.0048$). For performance ($Z = -0.0392$, $p = 0.9681$) and frustration ($W = 15.00$, critical value = 3.00) no significant differences were found. Comparing G-AAT to WG showed that G-AAT was significantly more mentally demanding ($W = 4.5$, critical value = 9.00). The other aspects did not show any significant difference.

Within the last subjective questionnaire, we asked the participants to rate the perceived enjoyment, motivation and realism in a 7-point Likert scale; ranging from Not at all to A lot. The statements were How much did you enjoy [the last condition]?, How motivated were you to perform well? and How realistic did you find [the last condition]?. As it can be seen in Figure 5, the participants significantly enjoyed the G-AAT more ($W = 0.00$, critical value = 5.00) and felt significantly more motivated to perform well ($W = 0.00$, critical value = 0.00) compared with the N-AAT. No significant effect was found in the enjoyment and motivation when comparing G-AAT to WG. The question about realism showed no significant effects. Thus, $H_1$ can and $H_3$ are not supported by these results.

Finally, the logged data revealed that the participants managed to push or pull approx. 44 bottles on average in the N-AAT (max: 55, min: 41, $\sigma = 3.43$) of which approx. 1.76 were incorrectly pushed or pulled (max: 8, min: 0, $\sigma = 2.09$). Whereas in the G-AAT the participants achieved on average a complete count of 142 bottles (max: 220, min: 89, $\sigma = 39.6$) with only 0.8 errors on average (max: 4, min: 0, $\sigma = 1.2$). The effect proved to be significant ($Z = -2.8451$, $p = 0.00438$). The participants also managed to significantly sort more bottles in G-AAT than in N-AAT ($Z = -3.1798$, $p = 0.00148$). This finding is in support of $H_2$.

### 4.4 Additional Comments

The participants enjoyed G-AAT and found it to be a straightforward and simple mini-game. Multiple participants emphasized how much they enjoyed throwing bottles. They also felt motivated by the time pressure. They appreciated that despite the time pressure, they were able to set the pace. Additionally, they suggested the following improvements: The height of the shelves was not optimal for tall participants, making them have to bend down rather low. Some participants would have found it more realistic if more virtual humans were shopping in the supermarket.

The general feedback about the N-AAT was less positive compared to G-AAT. It was perceived as more monotone and not mentally challenging enough. The participants wished for an acceleration, creating more time pressure (they had constantly two seconds to react per bottle). Some complained about the missing sound and auditory feedback.

Finally, the participants gave combined feedback on the WG. They generally enjoyed the money-aspect and that they were able to earn more money, if they played faster and that they had an influence on the spending: different items varied in costs. They enjoyed the freedom of choice and the progress they made. The different feedback types (the visual hint on the bottle, the summarized earned money, and the auditory feedback) accompanying each movement. The feedback was perceived as motivating especially in the G-AAT and CET. Of all the three mini-games, 7 participants liked the G-AAT the most and 4 participants the Shopping game. No participant favored the CET game. The participants suggested the following improvements: They wished for an option to quit the G-AAT and CET game if they thought they have earned enough money to...
complete the Shopping game. They found the placement of the alcoholic beverages in the Shopping game slightly irritating, because they were always placed between other items on the shelves.

5 Discussions and Conclusion

In this study three mini-games (G-AAT, CET, shopping) and a non-gamified application (N-AAT) for alcohol-addiction therapy in VR was presented. They were designed based on classical therapy methods such as AAT and CET, which are associated with some limitations such as the lack of context and motivation for long-term training. That is the patients might find these therapy methods hard and tedious. A gamified VR can not only add the context and user preferences, but also motivation to continue to the intervention program. Furthermore, in preparation for a long-term clinical study, we conducted a user study with healthy participants to evaluate the usability and motivation for training. Before starting such a long-term study, which is associated with more challenges such as access to the AUD-patients and therapists for the supervision, it was necessary to check whether anything needed to be changed. Moreover, it is plausible that the AUD-patients might perceive the VR-settings differently, meaning that the collected data needs to be treated carefully and may not generalize to all populations.

Our results suggest that all conditions (N-AAT, G-AAT, WG) had a high usability score (SUS scores were above average). Moreover, in comparison with N-AAT, G-AAT was significantly more demanding (mentally, physically and temporally), but also more motivating. Furthermore, the participants made significantly less errors in the G-AAT than N-AAT. When directly comparing G-AAT and WG it is important to note that G-AAT was a part of WG. However, more than half of our participants liked the G-AAT the best of all three mini-games presented within WG, while rather disliking the CET mini-game. Since many participants also enjoyed the Shopping mini-game, it could be assumed that the motivation and enjoyment would be higher when the final game used in a clinical context would only consist of the G-AAT and Shopping mini-game. This will also provide the game with a clear goal through the narrative (earning money to buy groceries).

Finally, further improvements may be relevant for creating effective gamified alcohol addiction therapy in VR. Customization based on the patient needs can be named as an example. Each patient has a different context in which he/she has a high risk of relapse. This context could be the own living room or a certain situation with friends, where the patient might face group pressure. Including the context may increase the effectiveness of modern therapy methods. To create an even more immersive experience and to induce even more craving, olfactory could also be implemented and incorporated into the gamified application. In short, gamified therapy in VR has the potential to revolutionize the recovery process of AUD patients by increasing their intrinsic motivation to abstain and allowing them to train at anytime and anywhere.

References