Emotion Elicitation Techniques in Virtual Reality

Radiah Rivu[1], Ruoyu Jiang[2], Ville Mäkelä[1,2,3], Mariam Hassib[1], and Florian Alt[1]

[1] Universität der Bundeswehr München, Germany, sheikh.rivu@unibw.de, mariam.hassib@unibw.de, florian.alt@unibw.de
[2] LMU Munich, Germany, jryisme@gmail.com, ville.maekelae@ifi.lmu.de
[3] University of Waterloo, Canada

Abstract. In this paper, we explore how state-of-the-art methods of emotion elicitation can be adapted in virtual reality (VR). We envision that emotion research could be conducted in VR for various benefits, such as switching study conditions and settings on the fly, and conducting studies using stimuli that are not easily accessible in the real world such as to induce fear. To this end, we conducted a user study (N=39) where we measured how different emotion elicitation methods (audio, video, image, autobiographical memory recall) perform in VR compared to the real world. We found that elicitation methods produce largely comparable results between the virtual and real world, but overall participants experience slightly stronger valence and arousal in VR. Emotions faded over time following the same pattern in both worlds. Our findings are beneficial to researchers and practitioners studying or using emotional user interfaces in VR.

Keywords: Emotions, Elicitation Methods, Virtual Reality, User Studies

1 Introduction

Emotion is a fundamental characteristic of human behavior [25] and crucial for communication. Hence, emotion is subject to extensive research not only in psychology but also in human-computer interaction (HCI) [8]. Researchers have explored various methods to elicit, measure, and quantify emotion, ranging from anger and sadness to excitement and happiness [27, 30, 36, 54]. The methods used to invoke specific emotions on humans are called *emotion elicitation methods* (or *techniques*). These methods include using external stimuli such as video, images, and sound, as well as recalling past events [10, 27, 38].

In this paper, we explore how various emotion elicitation methods function in virtual reality (VR), and how they compare to real-world use. We see several benefits from this work. First, we envision that VR could be used as a substitute for some real-world studies on emotion. VR can provide researchers with opportunities that they would not have in the real world, for example, by being able to manipulate the virtual environment on the fly [35], and researching scenarios that would be difficult or even dangerous in the real world and may put participants at risk(e.g., certain automotive scenarios [15]). Running studies in VR may also be a more cost-effective solution for complex studies. Second, understanding emotions in VR is beneficial to many VR experiences and games as emotions play a key part in the user experience. In particular, designers and practitioners can use this knowledge to effectively invoke the desired emotions in VR users, while also understanding the lasting effects from these emotions. Third, understanding emotions in VR would also be beneficial in medical treatments as VR can aid patients with treatment and therapy, for example treating anxiety in VR [3, 5, 46].

Existing research has gained only limited understanding on emotion elicitation in VR. In particular, we do not have substantial understanding on whether the different emotion elicitation methods in VR invoke emotions comparable to the real world. Emotion elicitation methods in VR is an under-research area in HCI. Therefore, in this paper, we provide the first exploratory assessment of emotion elicitation methods in VR. Our goal is to understand how different elicitation methods [10, 38, 27] affect a range of emotions in VR. We also investigate how emotion intensity changes over time and how these results compare to the results from an identical real-world study. To this end, we conducted a mixed-design user study (N = 39), where participants experienced four emotion elicitation methods (audio, video, image, autobiographical recall) and four basic emotions (happy, sad, excited, angry) in a virtual lab setting and an identical real-world setting. Our main research questions were:

- **RQ1:** Can established elicitation methods used in the real world elicit similar emotions in VR?
- **RQ2:** How do the elicited emotions wear off over time?
- **RQ3:** What are the unique characteristics of different elicitation techniques?

Our main results are: (1) Emotion elicitation methods work in a similar manner in both VR and the real world, producing largely comparable results, (2) All elicited emotions take time to fade out, thus there exists a minimum waiting period, and (3) All four elicitation techniques had very small differences between the real-world and virtual conditions. The experienced emotions were generally very slightly stronger in VR. Autobiographical memory recall had a unique challenge, as participants found it difficult to stop recalling their memories further.

We believe this understanding is valuable to researchers, as our work provides evidence that emotions can be researched in VR through emotion elicitation techniques. Our results are also useful for designers and practitioners, as they can inform how certain emotions can best be invoked in VR.

2 Background

Our work draws from previous work on 1) emotion elicitation methods, 2) understanding emotions in VR, and 3) the use of VR for research purposes.

2.1 Emotions and Emotion Elicitation Methods

Emotions have received considerable attention from psychology researchers for more than a century. Researchers have identified various basic emotions [22, 34, 45], the most common ones being happiness/joy, sadness, fear, and anger.

Psychologists have also proposed models that use continuous axes to model emotions. Two common axes are *arousal* and *valence*. Arousal refers to the amount of activation in an emotion, in other words, how much calmness or excitement is in the emotion. Valence refers to how positive or negative the emotion is. A third axis often shared by these theories is the *dominance* or control axis. This refers to how much control one has of the emotion, and if it is caused by an internal or external source [44]. Russell presented his Circumplex Model of affect with emotions scattered along both the arousal and valence axis [51]. This model is now commonly used in Affective Computing and HCI research.

To conduct emotion experiments, a broad spectrum of emotions need to be elicited through external stimuli. Exploring emotion elicitation methods is a crucial aspect towards understanding human behaviour and consists of a large research body [50]. Not only do researchers investigate the established approaches towards inducing emotion, but also the opportunity to explore new emotion induction methods such as music [59]. Researchers have also discussed the challenges and concerns of emotion elicitation [57].

Common elicitation methods include showing users audio [58], videos [27], and pictures [57], as well as invoking emotions through past memories using autobiographical recall [31]. Autobiographical recall encourages participants to recall a specific event or memory that can evoke a certain emotion [6, 11, 21]. Recalling past events can be done via writing, talking aloud or remembering. Autobiographical recall has been found to be the preferred method for emotional elicitation in scenarios such as automotive research [10], where using videos or pictures is not possible. Researchers have long compared each traditional method of elicitation to understand it's effectiveness and specific application [18, 30].

2.2 Utilizing VR as a Research Platform

Recently, advances in VR technologies have brought forth the question of whether studies involving human subjects could be conducted in virtual reality [35]. This approach aims, for example, to produce results that apply in similar real-world conditions [1, 19, 35], to generate real-world skills through virtual training [16, 39, 55], or to support real-world behavior change through VR [47].

Conducting studies in VR brings many benefits. Researchers have full control of the virtual world and may modify the environment, conditions and variables at will [35]. VR also gives a higher degree of freedom to researchers as it might enable user studies to be conducted remotely [49]. As VR headsets become more ubiquitous, consumers could participate in user studies remotely from the comfort of the users' own homes. In addition, VR has the potential to be used for studies that would be difficult or unethical to conduct in the real world, such as certain scenarios in automotive research [28, 15, 29] or in usable security and privacy research [2]. We take inspiration from these contexts and explore if VR can further enhance emotion research.

Some studies have briefly touched on emotions in VR. Dickinson et al. [20] investigated the feasibility of using VR to study gambling behaviour. Their findings suggest that participants experience higher levels of arousal and immersion in VR compared to the real world. Guidelines for research in human social interaction using VR is discussed by Pan et al. [40]. The authors compared arousal, immersion, task workload between the real and the virtual world; proposing VR simulations to be beneficial for the study of behaviour. This motivates our research as we comprehend how different emotions are affected in VR in terms of valence, arousal and dominance.

2.3 Emotion Elicitation in VR

Using VR to manipulate and induce emotions is explored by Estupinan et al. [23]. The authors investigated in a pilot study how valence and arousal in VR correlate with the real world. Their findings show that arousal is higher in virtual reality when using images to elicit emotions. Banos et al. [7] explored inducing mood through VR. They created a VR park scene which can adapt based on how the authors wanted to control an emotion after the elicitation phase. The strengths of VR as a platform for investigating emotion are identified by Chirico et al. [13], where they explored how an intense emotion such as *awe* can be elicited in VR. The authors argue in favor of VR allowing experimenters to generate a vast range of complex stimuli to elicit emotions, which are otherwise difficult in the real world. While the authors in this work looked into whether emotions could be elicited in VR, we apply and compare different emotion elicitation methods commonly used in the real word in VR to holistically understand their suitability for VR research and beyond. Riva et al. [48] explored the use of VR to evoke *anxiety* and *relaxation* and analysed the relationship between emotions and presence. Felnhofer et al. [24] covered a range of affective states (joy, sadness, boredom, anger, and anxiety) and found emotional VR scenarios to evoke a higher degree of presence, which in turn affects emotion intensity.

2.4 Summary and Research Approach

The previous sections demonstrate the importance of research on emotion and the need to understand emotion elicitation methods. With VR being an emerging technology which is rapidly becoming more ubiquitous, it is timely to explore how well emotion elicitation methods function in VR. To the best of our knowledge, this is the first work to explore and compare different elicitation methods in VR and the real world in parallel.

3 User Study

We designed a user study to understand the capabilities of different emotion elicitation methods in the virtual world (VW) compared to the real world. Participants experienced four different emotions through four different elicitation methods in the real world as well as in a virtual replica of the real-world space.

3.1 Participants

We recruited 39 participants (24 males) with an average age of 28.1 ($SD = 6.9$). We advertised the study via university mailing lists, social media platforms and company notice boards. 25 participants reported to have prior experience with VR. The participants experienced two different environments (real and virtual) on two separate sessions with each session lasting around 40 minutes. Participants were compensated with 20 EUR in cash for their time.

3.2 Study Design

We conducted a mixed-design experiment with three independent variables:

- **Setting (2 levels):** real world (RW) and virtual world (VW).
- **Elicitation Method (4 levels):** audio, video, image, autobio. recall (ABR)
- **Emotion (4 levels):** happiness, sadness, anger, excitement

Due to the many possible combinations (32), we opted for a Graeco-Latin Square to cover the majority of the combinations. Thus, our study was a mixed design, where participants experienced all elicitation methods and all emotions once in both settings, but they did not experience all combinations. Each participant experienced the same combination of emotion and emotion elicitation methods between the real and virtual world to compare the results. This resulted in four combinations in both settings per participant, in total eight combinations.

We selected four emotions based on Russell's Circumplex model of affect with arousal and valence dimensions as an appropriate model for our investigation, which has been demonstrated to work well in a variety of different studies [51]. The model has two dimensions: *Valence* indicates the positiveness or negativeness of the feeling; *Arousal* indicates how strong the feeling is [9]. We chose four emotions, one from each quartile of the model of affect [52]: Happiness (low arousal, positive valence), Sadness (low arousal, negative valence), Excitement (high arousal, positive valence), and Anger (high arousal, negative valence).

We selected the elicitation methods based on literature reviews which report these as commonly used approaches [14, 50]. For three of the four chosen elicitation methods (video, audio, picture), we needed material that were validated to elicit the desired emotions. These materials were chosen from validated datasets based on literature which shows prolific usage of the datasets in various research directions [32, 43, 56, 4, 53, 42]. The materials were selected based on the emotion label (we only chose materials that evoke our desired emotions) and the highest

Fig. 1. Office environment setup used in the study: a) real world, b) virtual world. We designed the VR environment to closely resemble its real-world counterpart. Participants' view in sitting position in c) real world, and d) Virtual world.

valence and arousal values. For autobiographical recall we did not need external materials. Instead, we asked participants to recall past events that had invoked certain emotions in them and it lasted on average two minutes. We showed video and images for 60 seconds and the audio clips lasted approximately 40-60 seconds. For audio, video and image we used the following databases:

Video Database for Emotion Analysis Using Physiological Signals (DEAP) [33]
Audio The Musical Emotional Bursts (MEB) [41]
Picture The Geneva affective picture database (GAPED) [17]

For collecting subjective measurements as ground truth, we used the 9-point Self-Assessment Manikin scale (SAM) [9] which constitutes our *emotion intensity* dependent variable. The SAM scale uses a visual representation as an indicator of the current emotion while separating the abstract mood into three dimensions: valence, arousal, as described above, and dominance, which indicates how much control the person has over the current feeling. We selected this pictorial representation as this may encourage participants to make reliable assessment of perceived emotion. The databases selected use the same scale (9 point SAM) to visualize Russell's model. We also measured whether emotion induction is dependent on *personality* using the Ten Item Personality Measure (TIPI) [26]. TIPI is a ten-item questionnaire where participants rate statements on a 7-Point Likert scale (1=strongly disagree; 7=strongly agree). The ten questions reflect the Big Five dimensions (Extroversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness) using two questions per dimension.

Sad - Video	Angry - Audio	Happy - Image	Exciting - AbR
Angry - AbR	Happy - Video	Exciting - Audio	Sad - Image
Happy - Audio	Exciting - Image	Sad - AbR	Angry - Video
Exciting - Image	Sad - AbR	Angry - Video	Happy - Audio

Table 1. The orders for each combination of emotions and methods based on Graeco-Latin Square (AbR = Autobiographical Recall).

3.3 Apparatus

For the real-world condition, we chose an office space including two tables, chairs, one cabinet, and windows on one side (Figure 1a). For the VR condition, we built a virtual replica of the same office (Figure 1b). In the real world, the participants were seated on a chair and the elicitation stimuli were shown through the computer screen placed on the table in front of them (Figure 1c). In the virtual replica, the exact same procedure was followed. The VR environment was programmed using Unity, running on HTC Vive via SteamVR.

3.4 Study Procedure

The participants experienced each setting (RW and VW) on separate days to eliminate any lasting effects from the first session. We counterbalanced the setting order. In both settings, each participant experienced all four emotion elicitation methods and all four emotions. The combinations of emotions and elicitation methods were kept the same between the settings, but the order of combination was counterbalanced. This was done to ensure that the data remain comparable between RW and VW. For counterbalancing, we used Graeco-Latin Square to cover the majority of the combinations (Table 1); each participant was exposed to four unique combinations of emotions and elicitation methods.

At the beginning of the first session, we explained the purpose of our study to the participants and they signed a consent form. They then filled out their demographic information (age, gender, background, experience with VR), personality questionnaire (TIPI). Before the study, participants were briefed about the elicitation materials and the SAM scale. Participants were then shown how to use the VR equipment and how fill in the SAM scale in VR using controllers.

In each session, participants experienced the elicitation material one at a time, after which they filled in the SAM scale five times with breaks of 30 seconds in between. We collected five SAM readings to analyze how emotion changes over time and at the same time maintain a reasonable study duration. To the best of our knowledge, no best practices exist regarding the optimal number of readings to study this. After the initial briefing, the participants completed the elicitation without the experimenter in the room. This was done to ensure that the life events narrated by the participants for the autobiographical recall method remained confidential. After finishing both sessions, they filled in the final questionnaire and were interviewed briefly. Each session lasted approximately 40 minutes. The entire study procedure is visualized in Figure 2.

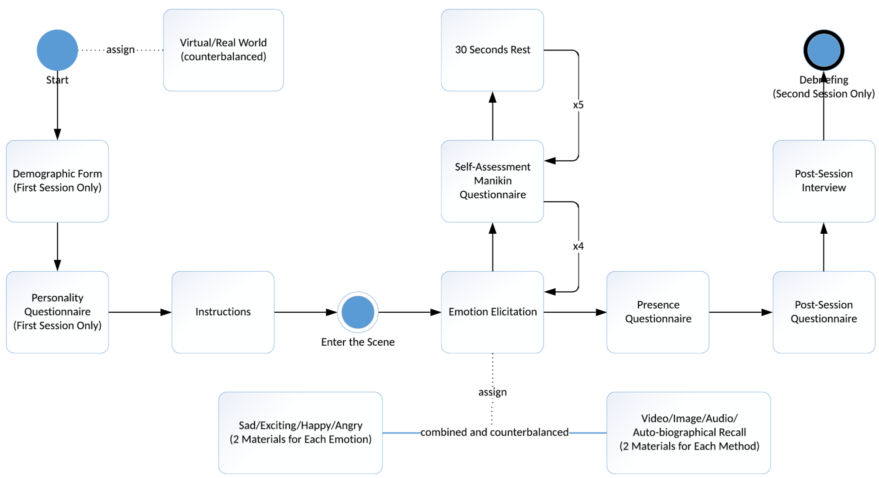

Fig. 2. An overview of the study procedure.

4 Results

We present the results from our comparative study. We considered all five SAM readings only for analysing how emotion intensity drops over time. For all the other analyses we only use the first SAM reading after each elicitation.

4.1 Correlations between Emotions and Personality, Age, Gender, and VR Experience

Using Kruskal-Wallis H Test, we compared the SAM results with the participants' VR experience, their age, gender and TIPI results in each combination of elicitation method, elicited emotion and environment. There was no statistically significant differences between the SAM results and any of the factors ($p > .05$). However, this result may differ if investigated with a larger sample.

4.2 Emotions between the Real World and the Virtual World

We used non-parametric test as the Shapiro-Wilk Test indicated the SAM values to deviate significantly from normal distribution. Table 2 shows the median values of valence, arousal and dominance from the SAM scale. The median analysis of each scale indicates that the elicitation methods are capable of eliciting emotions in VR similar to the real world. Friedman test did not show any significant differences between VR and RW with any of the dimensions. The only near-significant difference was observed in *arousal* regarding *Happiness* ($p = .058$). Therefore, the overall results regarding emotional states are the same between the real world and VR.

	Sadness (n=39)	Anger (n=39)	Happiness (n=39)	Excitement (n=39)
Valence Scale				
Virtual	4 ± 2.01	5 ± 1.90	7 ± 1.44	7 ± 1.66
Real	4 ± 1.79	5 ± 1.86	6 ± 1.02	6 ± 1.52
p-value	0.631	0.346	0.24	0.492
Arousal Scale				
Virtual	6 ± 2.05	5 ± 2.05	5 ± 1.97	6 ± 2.31
Real	5 ± 1.93	6 ± 1.80	4 ± 1.88	5 ± 1.81
p-value	0.375	0.766	0.058	0.267
Dominance Scale				
Virtual	6 ± 1.90	6 ± 2.38	7 ± 1.64	8 ± 1.53
Real	7 ± 2.16	7 ± 1.98	7 ± 1.4	7 ± 1.63
p-value	0.693	0.324	0.852	0.421

Table 2. Median and p-values of SAM scores for each elicited emotion between virtual and real world.

Figure 3 shows box plots for each SAM dimension, environment, and elicited emotion. A Wilcoxon Signed-Rank Test was conducted between virtual and real world regarding valence, arousal and dominance value for each elicited emotion.

For *Anger*, the Wilcoxon Signed-Rank test showed no significant difference in valence (Z = -.943, p = .346), arousal (Z = -.297, p = .766) and dominance (Z = -.987, p = .324) between both environments . Both elicited same median value for valence scale (median = 5). The median values of arousal and dominance dimensions are lower in VR (median of Arousal = 5, median of Dominance = 6) compared to real world (median of Arousal = 6, median of Dominance = 7).

For *Sadness*, no significant difference exists for valence (Z = -.481, p = .631), arousal (Z = -.886, p = .375) and dominance (Z = -.395, p = .693). Though the median values of valence are the same (median = 4), the median arousal value in VR (median = 6) is higher than in the real world (median = 5). The median dominance in VR (median = 6) is lower than in the real world (median = 7).

For *Excitement*, no significant difference exists for valence (Z = -.687, p = .492), arousal (Z = -1.109, p = .267) and dominance (Z = -.804, p = .421). Median values are higher in the virtual (median of Valence = 7, median of Arousal = 6, median of Dominance = 8) compared to the real world (median of Valence = 6, median of Arousal = 5, median of Dominance = 7).

For *Happiness*, valence (Z = -1.176, p = .240) and dominance (Z = -.186, p = .852) show no significant differences. Valence in VR (median = 7) has higher median value than the real world (median = 6). While arousal (Z = -1.898, p = .058) scale after emotion Happiness elicitation indicates the two environments being marginally significant. The median value for arousal in VR (median = 5) is higher (median = 4). The median values for dominance are the same for both environments (median = 7).

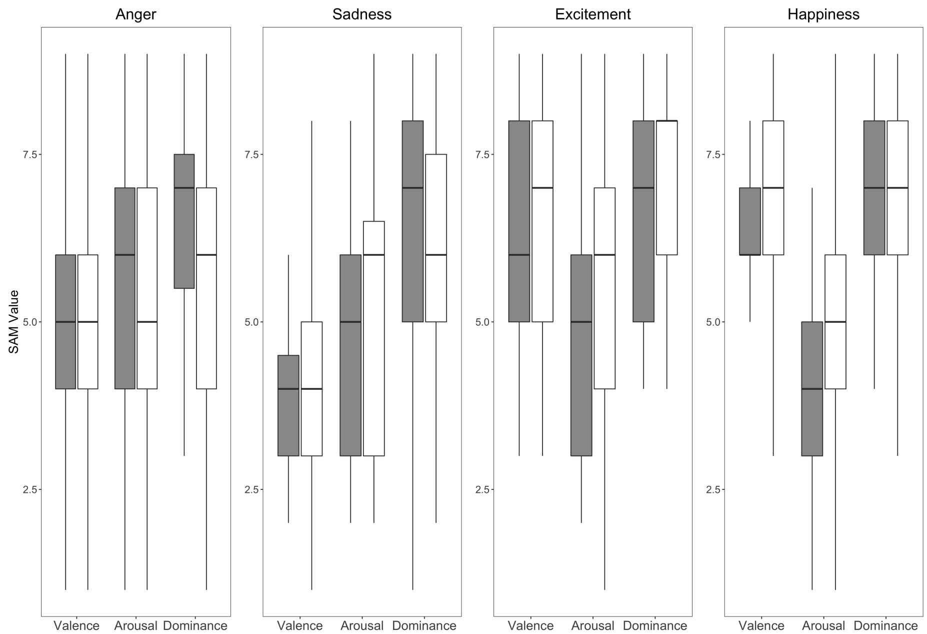

Fig. 3. SAM scores for each elicited emotion in the virtual world and the real world.

4.3 Emotion Elicitation Methods

In the next four tables (Tables 3, 4, 5, 6), we present the comparisons of the elicitation methods for *Sadness*, *Anger*, *Happiness*, and *Excitement* respectively.

Video The results from eliciting *Sadness* through videos show no significant difference between valence and arousal scores between the settings. The median values for arousal were the same across both worlds, though valence was slightly lower in VR. However, dominance was significantly lower in VR than in the real world ($p = 0.044$). Eliciting *Anger* showed no significant differences among the valence, arousal and dominance scales. The median values for valence and arousal were the same, but dominance was slightly higher in VR. Eliciting *Happiness* also showed no significant differences in valence, arousal or dominance, although the median values for valence and arousal were slightly higher in VR.

Image Using images to elicit *Sadness* did not show significant differences in any dimension between the settings. The medians for both valence and dominance were the same between the real world and the virtual world. Arousal was slightly higher in the virtual world. Eliciting *Happiness* showed no significant differences in any dimension between the settings. Eliciting *Excitement* showed a significant difference in arousal ($p = 0.024$), but not in dominance or valence, although valence was slightly higher without significance in the virtual world.

Valence Scale				
	Sadness	Anger	Happiness	Excitement
Virtual World	6.00	6.00	7.50	-
Real World	6.50	6.00	6.00	-
p-value	0.572	0.293	0.251	-
Arousal Scale				
	Sadness	Anger	Happiness	Excitement
Virtual World	6.00	5.00	4.50	-
Real World	6.00	5.00	3.00	-
p-value	0.666	0.572	0.404	-
Dominance Scale				
	Sadness	Anger	Happiness	Excitement
Virtual World	6.50	7.00	6.00	-
Real World	8.50	6.00	6.00	-
p-value	0.044	0.206	0.832	-

Table 3. Median and p-value of SAM ratings for Video elicitation method between virtual and real world.

Valence Scale				
	Sadness	Excitement	Happiness	Anger
Virtual World	4.00	7.00	7.00	-
Real World	4.00	6.00	7.00	-
p-value	0.172	0.374	0.388	-
Arousal Scale				
	Sadness	Excitement	Happiness	Anger
Virtual World	5.00	6.00	4.00	-
Real World	4.00	5.00	4.00	-
p-value	0.397	0.024	0.407	-
Dominance Scale				
	Sadness	Excitement	Happiness	Anger
Virtual World	5.5	7.00	8.00	-
Real World	5.5	7.00	8.00	-
p-value	0.886	0.390	0.389	-

Table 4. Median and p-value of SAM ratings for Image elicitation method between virtual and real world.

Autobiographical Recall Eliciting *Sadness* using Autobiographical Recall showed no significant difference between the real and virtual world for valence, arousal or dominance. The median arousal value in VR is higher and the median dominance is lower. Eliciting *Anger* again showed no difference across the dimensions. Median arousal is lower in the virtual world and the median dominance is higher. Eliciting *Excitement* showed no significant difference though the median valence and arousal values are lower in virtual world and dominance is higher.

Valence Scale				
	Sadness	Anger	Excitement	Happiness
Virtual World	4.00	3.00	7.50	-
Real World	4.00	3.00	8.00	-
p-value	0.273	0.121	0.465	-
Arousal Scale				
	Sadness	Anger	Excitement	Happiness
Virtual World	6.00	5.50	4.50	-
Real World	5.00	7.00	6.00	-
p-value	0.433	0.380	0.634	-
Dominance Scale				
	Sadness	Anger	Excitement	Happiness
Virtual World	5.00	6.50	8.00	-
Real World	6.00	6.00	7.00	-
p-value	0.407	0.546	0.722	-

Table 5. Median and p-value of SAM ratings for Autobiographical Recall elicitation method between virtual and real world.

Valence Scale				
	Anger	Happiness	Excitement	Sadness
Virtual World	5.00	7.00	5.00	-
Real World	5.00	7.00	5.50	-
p-value	0.391	0.552	0.192	-
Arousal Scale				
	Anger	Happiness	Excitement	Sadness
Virtual World	5.00	6.00	5.00	-
Real World	4.00	5.00	4.00	-
p-value	0.666	0.110	0.192	-
Dominance Scale				
	Anger	Happiness	Excitement	Sadness
Virtual World	6.00	6.00	8.00	-
Real World	7.00	7.00	7.00	-
p-value	0.180	0.626	0.917	-

Table 6. Median and p-value of SAM ratings for Audio elicitation method between virtual and real world.

Audio Eliciting *Anger* showed no significant difference across the dimensions with the median arousal value being slightly higher and dominance value being slightly lower in the virtual world. For *Happiness* we saw no statistical significance between the worlds for the SAM dimensions. Similar to *Anger*, the median arousal value is higher and the dominance value is lower. For *Excitement*, there was no statistical significance though valence is slightly lower in virtual world and dominance, arousal slightly higher in virtual world.

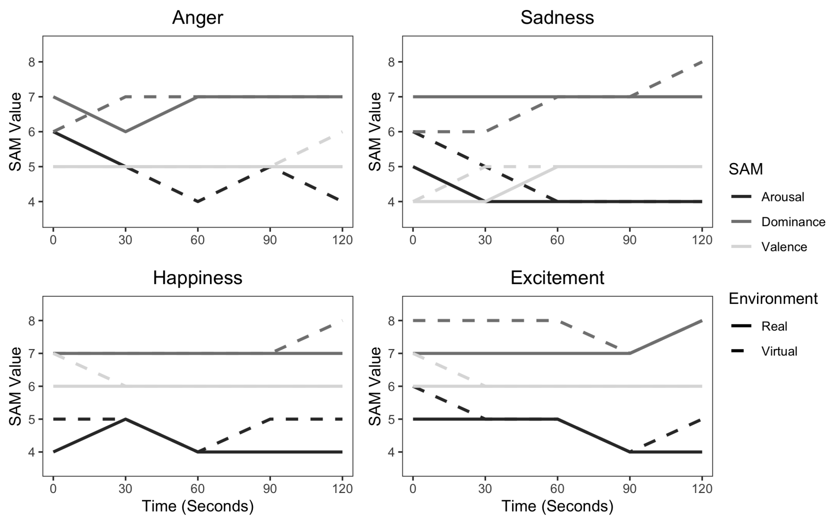

Fig. 4. The lasting effect of arousal, dominance and valence for each emotion in the real world and the virtual world.

4.4 Lasting Effects of Emotions

We analyzed how the SAM values for valence, arousal and dominance change over time for each emotion (Figure 4). SAM readings were taken five times at 30-second intervals. While the median valence values after emotion *Sadness* in the real world showed an ascending trend, the median valence values for the other three emotions remained still over time. For all emotions across both dimensions, the median arousal values became weaker over time. The median dominance values for *Anger* and *Excitement* in both environments showed slight increase over time. For *Sadness* and *Happiness* dominance in virtual world showed an increasing trend while it remained at a constant value in the real world.

We also compared how the emotions vary over time for each elicitation method in RW and VR (Figure 5). Each elicitation method followed a similar pattern except *Audio*, which showed a greater deviation between VR and the real world for all dimensions, compared to other methods.

4.5 Participants' Perspective

At the end of the study, we interviewed participants about their experiences in the RW and VR settings, particularly about whether or not they found these two experiences similar. Most participants stated that they did indeed perceive the experience between the settings very similar. However, some participants mentioned that the real world had more objects, like noticing extra cables on the table in the real world which was absent in VR, in the room which the virtual world was missing. Nonetheless, this did not seem to be a major factor in the experience, as the layout and most critical objects were the same.

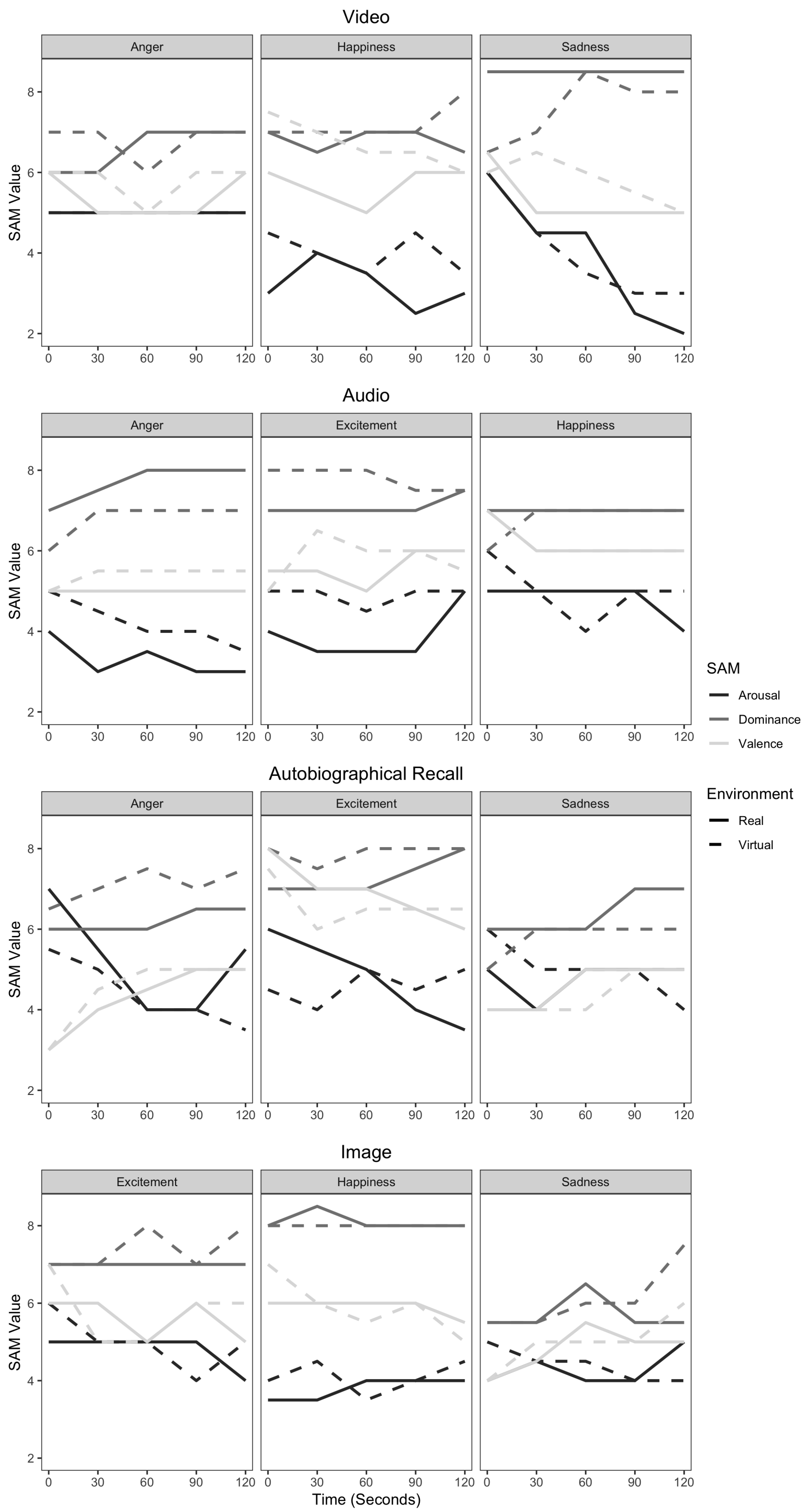

Fig. 5. The lasting effect of arousal, dominance and valence for each emotion with each elicitation method in the real world and the virtual world.

We gained interesting insights from participants regarding autobiographical memory recall. 16 participants mentioned that for emotion elicitation using memory recall, they preferred the real world since VR was perceived more as a game experience. However, 17 participants felt that the virtual world had fewer distractions (such as absence of environmental noise) and helped them concentrate better to recall past events. A few participants also felt that the time given (two minutes) to recall an event was short and they were worried they might not be finished, which may have distracted them.

Several participants mentioned that once they recalled a memory using ABR, continuous memories kept coming back to them even after the emotion elicitation task was over. This might indicate that emotion elicitation through autobiographical memory recall might affect the person for longer periods than other methods, because it is more difficult to fully stop the elicitation when prompted. One participant also mentioned that positive memory recall led to thinking negative memories only in the virtual environment. This indicates possible deviation in behavior and we believe this requires further exploration in VR.

5 Discussion

In this section, we discuss the feasibility of conducting studies entailing emotions and emotion elicitation methods in VR by drawing upon the conducted comparative study. We discuss the differences between the induced emotions in the real and the virtual world. We also present a comparison of the different emotion elicitation methods and the lessons learned.

5.1 Elicitation Methods in VR

We explored four emotion elicitation methods (audio, video, image, autobiographical recall) which have been actively used for research in the real world. We observed that all four elicitation methods were capable of inducing emotions in VR similar to the real world, indicating similar emotional states from participants in both environments.

Taking a closer look at each method, we note that emotion elicitation using video induced stronger negative emotions (*Anger*, *Sadness*) in the virtual world. Similar to video elicitation method, image also is capable of inducing the desired emotions in both the worlds with stronger intensity in the virtual world. Elicitation method using audio failed to induce emotions of the same intensity compared to other methods. Using autobiographical memory recall induces stronger emotions for *Sadness* and *Excitement* in VR, but stronger emotions in the real world for *Anger*.

5.2 Elicited Emotions

We measured the induced emotions using the SAM scale, where participants rated the emotions in three dimensions: valence, arousal and dominance. Generally, our results show that the emotions induced in both RW and VR followed a very similar pattern. None of the measured emotions showed any statistically significant differences between the settings in any of the dimensions. At more detailed analysis of the elicitation methods, emotions, and dimensions shows that only 2 combinations out of the total 36 showed significant differences between the settings (Video-Sadness-Dominance and Image-Excitement-Arousal).

After each elicitation, we measured the emotion five times in 30-second intervals to investigate how quickly they would fade. Prior work investigated if virtual environment itself induces the same emotions when compared to the real world [12] which shows that emotions remain similar in both the real and virtual world. In our paper, we investigated the different emotion elicitation techniques and the differences found between VW and RW were very small, either with no difference or a difference of one point on the 9-point scale. The only exception was the first SAM measure for *Sadness*, with a difference of two points between VW and RW. Therefore, the way the emotions developed over time after elicitation was also very similar between the real world and the virtual world.

5.3 The Lasting Effects of the Elicited Emotions

By taking the SAM scale measures five times with 30-second intervals, we measured how each elicited emotion developed during the two minutes that followed the elicitation. In many cases, the emotions still existed after this period, indicating that two minutes may not be enough to diminish the elicited emotion. This is both an opportunity and a challenge. For research purposes, researchers might want to consider a longer waiting period between the elicited emotions, or look for ways to diminish the emotions more quickly. For designers of VR experiences, this could be useful information when designing the experience and the specific scenes within, knowing that elicited emotions can last for some time.

5.4 Lessons Learned

Our results suggest that VR could be used to research emotions and emotion elicitation techniques, and acquire results that translate into the real world. We note that the discrepancies between the RW and VW itself might influence the experienced emotion and that this could be explored in-depth in future work.

Through this study, we gained insight into the differences between emotion elicitation methods. One highlight is the availability of resources for each method. For example, using autobiographical method does not require any additional resources as this is dependent on the participants only. But using other methods would require researchers to have access to verified and established resources to elicit the desired emotion. Obtaining the dataset is also a time consuming process. The other prime difference is the implementation of methods in VR. For example, writing in VR is harder to implement for autobiographical recall compared to "think-aloud" method.

Autobiographical recall has great potential in specific applications where using materials to elicit emotions is challenging (for example, automotive research). However, we observed that many participants found it difficult to stop recalling their life events once they initiated the process. This may affect further actions from the user, thus care must be taken to ensure that users reach a neutral emotional state when needed.

Our results can be translated to clear actionables, allowing researchers to choose the most fitting method of emotion elicitation based on the task at hand and the resulting preoccupation. For example, where visual senses are occupied such as automotive studies, autobiographical recall or audio method is a fitting approach. Where auditory senses are occupied, such as in a speech interface study autobiographical recall (without think-aloud), image or video (without sound) is a fitting approach. It can also help us to know which methods to eliminate from a set of available options. For example, in a task where cognition is occupied, one should avoid autobiographical recall.

Designers can use the results to inform the design of various VR experiences such as Cinematic Virtual Reality (CVR) as well as VR games, where designers want their experience to elicit specific emotions. For example, in a segment where VR players admire a view and designers want to elicit happiness, auditory cues could be used that do not interfere with the visual experience.

5.5 Limitations and Future Work

To measure the elicited emotions, we used a subjective assessment (SAM scale). Emotions could be more thoroughly measured with additional physiological measures such as heart rate and blood pressure [37].

We identified three areas that could be addressed in the future:

1) Diminishing elicited emotions in VR. In our work we observed how the elicited emotions fluctuate over a brief period of time. In the future, it would be valuable to investigate how long emotions truly last in VR, and which methods could be reasonably used to accelerate this process (for example, making participants fill in a non-relevant questionnaire). Also, we could investigate how exposure time to emotion-inducing material affects the emotions' duration.

2) Exploring different implementations for autobiographical memory recall. In our study, participants recalled life events without anyone else present in the room. In the real world, this method is often conducted by asking participants to write about their life events or talk about them to another person. Therefore, in the future, it might be valuable to investigate different implementations of autobiographical recall in VR (e.g, writing in VR, reporting the experience to a virtual avatar).

3) Emotion elicitation during tasks. In our work, we focused on studying emotion and emotion elicitation techniques at large. Because most VR experiences and studies conducted in VR involve users being active and engaging in various interactive tasks, it could be valuable to investigate further how tasks of various nature might affect the performance of elicitation methods and the strength of the invoked emotions.

6 Conclusion

In this paper, we explored how established emotion elicitation methods work in virtual reality. In particular, we used four emotion elicitation techniques (audio, video, image, and autobiographical recall) to elicit four different emotions (anger, excitement, happiness, and sadness).

We conducted a study, where participants experienced the elicitation methods in the real world and in an identical VR setting. We then compared the results between the real world and the virtual world. We learned that: **(1)** Emotion elicitation methods work in a similar manner in both VR and the real world, producing largely comparable results, **(2)** All of the elicited emotions have a lingering effect and it takes time for the emotion to diminish, and **(3)** All four elicitation techniques had very small differences between the real-world and the virtual world. The experienced emotions were in some cases slightly stronger in VR. Autobiographical memory recall had a unique challenge, as participants found it difficult to stop recalling their memories further.

Through our work, we have gained an understanding of the potential, strengths, and weaknesses of emotion elicitation methods in VR. Our work can inform how elicitation methods can be utilized in VR, and in particular, whether or not results from virtual emotion studies can translate into the real world.

Acknowledgements. This work was funded by the German Research Foundation (DFG), project no. 316457582 and 425869382, and by dtec.bw — Digitalization and Technology Research Center of the Bundeswehr [Voice of Wisdom].

References

1. Philipp Agethen, Viswa Subramanian Sekar, Felix Gaisbauer, Thies Pfeiffer, Michael Otto, and Enrico Rukzio. 2018. Behavior Analysis of Human Locomotion in the Real World and Virtual Reality for the Manufacturing Industry. *ACM Trans. Appl. Percept.* 15, 3, Article 20 (July 2018), 19 pages. DOI: http://dx.doi.org/10.1145/3230648
2. Florian Alt. 2021. Out of the Lab Resarch in Usable Security and Privacy, In Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization. *Adjunct Proceedings of the 29th ACM-Conference on User Modeling, Adaptation and Personalization* (2021). DOI: http://dx.doi.org/10.1145/3450614.3464468 to appear.
3. Page Anderson, Barbara O Rothbaum, and Larry F Hodges. 2003. Virtual reality exposure in the treatment of social anxiety. *Cognitive and Behavioral Practice* 10, 3 (2003), 240–247.
4. John Atkinson and Daniel Campos. 2016. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Systems with Applications* 47 (2016), 35–41.
5. Laura Aymerich-Franch and J Bailenson. 2014. The use of doppelgangers in virtual reality to treat public speaking anxiety: a gender comparison. In *Proceedings of the International Society for Presence Research Annual Conference*. Citeseer, 173–186.

6. Richard C. Baker and Daniel O. Guttfreund. 1993. The effects of written autobiographical recollection induction procedures on mood. *Journal of Clinical Psychology* 49, 4 (1993), 563–568. DOI:http://dx.doi.org/10.1002/1097-4679(199307)49:4¡563::AID-JCLP2270490414¿3.0.CO;2-W
7. Rosa María Baños, Víctor Liaño, Cristina Botella, Mariano Alcañiz, Belén Guerrero, and Beatriz Rey. 2006. Changing induced moods via virtual reality. In *International Conference on Persuasive Technology*. Springer, 7–15.
8. Russell Beale and Christian Peter. 2008. The role of affect and emotion in HCI. In *Affect and emotion in human-computer interaction*. Springer, 1–11.
9. Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59.
10. Michael Braun, Simon Weiser, Bastian Pfleging, and Florian Alt. 2018. A Comparison of Emotion Elicitation Methods for Affective Driving Studies. In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '18)*. Association for Computing Machinery, New York, NY, USA, 77–81. DOI: http://dx.doi.org/10.1145/3239092.3265945
11. Alice Chirico, Pietro Cipresso, and Andrea Gaggioli. 2018. Psychophysiological specificity of four basic emotions through autobiographical recall and videos. In *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer, 1–8.
12. Alice Chirico and Andrea Gaggioli. 2019. When virtual feels real: comparing emotional responses and presence in virtual and natural environments. *Cyberpsychology, Behavior, and Social Networking* 22, 3 (2019), 220–226.
13. Alice Chirico, David B Yaden, Giuseppe Riva, and Andrea Gaggioli. 2016. The potential of virtual reality for the investigation of awe. *Frontiers in psychology* 7 (2016), 1766.
14. James A Coan and John JB Allen. 2007. *Handbook of emotion elicitation and assessment*. Oxford university press.
15. Mark Colley, Marcel Walch, and Enrico Rukzio. 2019. For a Better (Simulated) World: Considerations for VR in External Communication Research. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings (AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 442–449. DOI: http://dx.doi.org/10.1145/3349263.3351523
16. Leanne Coyne, Jody K Takemoto, Brittany L Parmentier, Thayer Merritt, and Rachel A Sharpton. 2018. Exploring virtual reality as a platform for distance team-based learning. *Currents in Pharmacy Teaching and Learning* 10, 10 (2018), 1384–1390.
17. Elise S Dan-Glauser and Klaus R Scherer. 2011. The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behavior research methods* 43, 2 (2011), 468.
18. Moira Dean, Anne Arvola, Marco Vassallo, Liisa Lähteenmäki, Monique M Raats, Anna Saba, and Richard Shepherd. 2006. Comparison of elicitation methods for moral and affective beliefs in the theory of planned behaviour. *Appetite* 47, 2 (2006), 244–252.
19. Shuchisnigdha Deb, Daniel W. Carruth, Richard Sween, Lesley Strawderman, and Teena M. Garrison. 2017. Efficacy of virtual reality in pedestrian safety research. *Applied Ergonomics* 65 (2017), 449 – 460. DOI: http://dx.doi.org/10.1016/j.apergo.2017.03.007

20. Patrick Dickinson, Kathrin Gerling, Liam Wilson, and Adrian Parke. 2020. Virtual reality as a platform for research in gambling behaviour. *Computers in Human Behavior* 107 (2020), 106293.
21. Sasa Drace. 2013. Evidence for the role of affect in mood congruent recall of autobiographic memories. *Motivation and emotion* 37, 3 (2013), 623–628.
22. Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.
23. Sergio Estupiñán, Francisco Rebelo, Paulo Noriega, Carlos Ferreira, and Emília Duarte. 2014. Can virtual reality increase emotional responses (Arousal and Valence)? A pilot study. In *International conference of design, user experience, and usability*. Springer, 541–549.
24. Anna Felnhofer, Oswald D Kothgassner, Mareike Schmidt, Anna-Katharina Heinzle, Leon Beutl, Helmut Hlavacs, and Ilse Kryspin-Exner. 2015. Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International journal of human-computer studies* 82 (2015), 48–56.
25. Maria Gendron. 2010. Defining emotion: A brief history. *Emotion Review* 2, 4 (2010), 371–372.
26. Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.
27. James J Gross and Robert W Levenson. 1995. Emotion elicitation using films. *Cognition & emotion* 9, 1 (1995), 87–108.
28. Mariam Hassib, Michael Braun, Bastian Pfleging, and Florian Alt. 2019. Detecting and influencing driver emotions using psycho-physiological sensors and ambient light. In *Proceedings of the 17th IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT '19)*. Springer, Berlin-Heidelberg, Germany.
29. Kai Holländer, Ashley Colley, Christian Mai, Jonna Häkkilä, Florian Alt, and Bastian Pfleging. 2019. Investigating the Influence of External Car Displays on Pedestrians' Crossing Behavior in Virtual Reality. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 27, 11 pages. DOI:http://dx.doi.org/10.1145/3338286.3340138
30. Christophe Jallais and Anne-Laure Gilet. 2010. Inducing changes in arousal and valence: Comparison of two mood induction procedures. *Behavior research methods* 42, 1 (2010), 318–325.
31. Joris H Janssen, Paul Tacken, JJG de Vries, Egon L van den Broek, Joyce HDM Westerink, Pim Haselager, and Wijnand A IJsselsteijn. 2013. Machines outperform laypersons in recognizing emotions elicited by autobiographical recollection. *Human–Computer Interaction* 28, 6 (2013), 479–517.
32. Noppadon Jatupaiboon, Setha Pan-ngum, and Pasin Israsena. 2013. Real-time EEG-based happiness detection system. *The Scientific World Journal* 2013 (2013).
33. Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing* 3, 1 (2011), 18–31.
34. Kees Leidelmeijer. 1991. *Emotions: An experimental approach.* Tilburg University Press.
35. Ville Mäkelä, Rivu Radiah, Saleh Alsherif, Mohamed Khamis, Chong Xiao, Lisa Borchert, Albrecht Schmidt, and Florian Alt. 2020. Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In *Proceedings of the 2020*

CHI Conference on Human Factors in Computing Systems (CHI '20). ACM, New York, NY, USA. DOI:http://dx.doi.org/10.1145/3313831.3376796
36. W. Mardini, G. A. Ali, E. Magdady, and S. Al-momani. 2018. Detecting human emotions using electroencephalography (EEG) using dynamic programming approach. In *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*. 1–5. DOI:http://dx.doi.org/10.1109/ISDFS.2018.8355324
37. Iris B Mauss and Michael D Robinson. 2009. Measures of emotion: A review. *Cognition and emotion* 23, 2 (2009), 209–237.
38. Caitlin Mills and Sidney D'Mello. 2014. On the validity of the autobiographical emotional memory task for emotion induction. *PloS one* 9, 4 (2014), e95837.
39. Mehdi Moussaïd, Mubbasir Kapadia, Tyler Thrash, Robert W Sumner, Markus Gross, Dirk Helbing, and Christoph Hölscher. 2016. Crowd behaviour during high-stress evacuations in an immersive virtual environment. *Journal of The Royal Society Interface* 13, 122 (2016), 20160414.
40. Xueni Pan and Antonia F de C Hamilton. 2018. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology* 109, 3 (2018), 395–417.
41. Sébastien Paquette, Isabelle Peretz, and Pascal Belin. 2013. The "Musical Emotional Bursts": a validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in psychology* 4 (2013), 509.
42. Sébastien Paquette, Sylvain Takerkart, Shinji Saget, Isabelle Peretz, and Pascal Belin. 2018. Cross-classification of musical and vocal emotions in the auditory cortex. *Ann. NY Acad. Sci* 1423 (2018), 329–337.
43. Kuan-Chuan Peng, Tsuhan Chen, Amir Sadovnik, and Andrew C Gallagher. 2015. A mixed bag of emotions: Model, predict, and transfer emotion distributions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 860–868.
44. Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence* 23, 10 (2001), 1175–1191.
45. Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*. Elsevier, 3–33.
46. Claudia Repetto, Andrea Gaggioli, Federica Pallavicini, Pietro Cipresso, Simona Raspelli, and Giuseppe Riva. 2013. Virtual reality and mobile phones in the treatment of generalized anxiety disorders: a phase-2 clinical trial. *Personal and Ubiquitous Computing* 17, 2 (2013), 253–260.
47. Giuseppe Riva, Rosa M Baños, Cristina Botella, Fabrizia Mantovani, and Andrea Gaggioli. 2016. Transforming experience: the potential of augmented reality and virtual reality for enhancing personal and clinical change. *Frontiers in psychiatry* 7 (2016), 164.
48. Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. 2007. Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior* 10, 1 (2007), 45–56.
49. Radiah Rivu, Ville Mäkelä, Sarah Prange, Sarah Delgado Rodriguez, Robin Piening, Yumeng Zhou, Kay Köhle, Ken Pfeuffer, Yomna Abdelrahman, Matthias Hoppe, Albrecht Schmidt, and Florian Alt. 2021. Remote VR Studies - A Framework for Running Virtual Reality Studies Remotely Via Participant-Owned HMDs. *CoRR* abs/2102.11207 (2021). https://arxiv.org/abs/2102.11207
50. Erika L Rosenberg and Paul Ekman. 2000. Emotion: Methods of study. (2000).

51. James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.
52. James A Russell and Lisa Feldman Barrett. 1999. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology* 76, 5 (1999), 805.
53. Matthew E Sachs, Assal Habibi, Antonio Damasio, and Jonas T Kaplan. 2018. Decoding the neural signatures of emotions expressed through sound. *NeuroImage* 174 (2018), 1–10.
54. Kathleen Searles and Kyle Mattes. 2015. It's a Mad, Mad World: Using Emotion Inductions in a Survey. *Journal of Experimental Political Science* 2, 2 (2015), 172–182. DOI:http://dx.doi.org/10.1017/XPS.2015.5
55. Azadeh Shariati, Mojtaba Shahab, Ali Meghdari, Ali Amoozandeh Nobaveh, Raman Rafatnejad, and Behrad Mozafari. 2018. Virtual Reality Social Robot Platform: A Case Study on Arash Social Robot. In *International Conference on Social Robotics*. Springer, 551–560.
56. M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic. 2012. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Transactions on Affective Computing* 3, 1 (2012), 42–55.
57. Meike K Uhrig, Nadine Trautmann, Ulf Baumgärtner, Rolf-Detlef Treede, Florian Henrich, Wolfgang Hiller, and Susanne Marschall. 2016. Emotion elicitation: A comparison of pictures and films. *Frontiers in psychology* 7 (2016), 180.
58. Daniel Västfjäll. 2001. Emotion induction through music: A review of the musical mood induction procedure. *Musicae Scientiae* 5, 1_suppl (2001), 173–211.
59. Seunghyun Woo, Dong-Seon Chang, Daeyun An, Dongjin Hyun, Christian Wallraven, and Manfred Dangelmaier. 2017. Emotion Induction in Virtual Environments: A Novel Paradigm Using Immersive Scenarios in Driving Simulators. In *SIGGRAPH Asia 2017 Posters (SA '17)*. Association for Computing Machinery, New York, NY, USA, Article Article 49, 2 pages. DOI:http://dx.doi.org/10.1145/3145690.3145693