An Immersive and Interactive VR Dataset to Elicit Emotions

Weiwei Jiang, Maximiliane Windl, Benjamin Tag, Zhanna Sarsenbayeva, and Sven Mayer,

Abstract— Images and videos are widely used to elicit emotions; however, their visual appeal differs from real-world experiences. With virtual reality becoming more realistic, immersive, and interactive, we envision virtual environments to elicit emotions effectively, rapidly, and with high ecological validity. This work presents the first interactive virtual reality dataset to elicit emotions. We created five interactive virtual environments based on corresponding validated 360° videos and validated their effectiveness with 160 participants. Our results show that our virtual environments successfully elicit targeted emotions. Compared with the existing methods using images or videos, our dataset allows virtual reality researchers and practitioners to integrate their designs effectively with emotion elicitation settings in an immersive and interactive way.

Index Terms—Virtual Reality, Emotions, Dataset (https://github.com/HighTemplar-wjiang/VR-Dataset-Emotions)

1 Introduction

Human emotions are integral to our everyday behavior and decision-making [25]. Emotions are complex and multifaceted experiences, encompassing a range of cognitive, physiological, and behavioral components influenced by various subjective factors such as personal history, cultural background, and situational context [8]. The ability of a person to generate an appropriate emotional response to an environmental stimulus, thus, involves a multitude of affective and cognitive components of emotional intelligence [39]. Emotions can be fleeting, subtle, or overlapping, making it difficult to quantify them accurately [76]. Therefore, eliciting, quantifying, and capturing emotions reliably and robustly presents a significant challenge.

Besides quantifying and capturing true human emotions, eliciting true emotions in research settings is also challenging. Researchers have developed various methods and techniques to elicit specific emotions in response to this challenge. In particular, commonly employed approaches include film clips [55], still images [38], music [33] and sounds [10], memory recall [21], facial expression manipulations [16], feedback [63], and 360° videos in virtual reality (VR) scenes [40]. Specifically, films have been seen as a gold standard in psychology [26]. Nevertheless, it is challenging to elicit emotions in a reliable and standardized way due to the subjective nature of emotional experience.

Even though all these techniques exist, running studies involving eliciting emotions can be arduous and time-consuming for researchers and research participants. Researchers often find themselves compelled to repeat the experiment numerous times to accommodate for factors such as external context, personal background, or individual experiences to ensure consistent emotional responses across participants. Others have recently called for reconsidering the methodology of emotion research, *e.g.*, by working with individuals (*i.e.*, $N = 1$) rather than larger populations [65]. These issues are exacerbated when the experimental protocol for eliciting emotions is intricate and lengthy, extending the study duration.

Consequently, the research community will benefit from a simple, reliably reproducible, and readily customizable method to elicit emotions effectively. As individuals can have different emotional responses to the same stimuli, creating a method that can reliably elicit consistent emotional states across diverse participants becomes more challenging. Moreover, there exists a discrepancy between stimuli in the efficacy of eliciting emotion, as certain stimuli prove to be more efficient [22]. This discrepancy contributes to the complexity of creating a universal experimental method for successfully eliciting emotions, as researchers must meticulously select the appropriate stimulus for their experiments. Finally, as emotion-eliciting experiments virtually always mean manipulating participants' emotions, they may potentially have short-term and long-term effects on the participants' mental state. Therefore, to minimize any potential negative impact, the method for eliciting emotions must account for participants' well-being by allowing for an adequate selection of stimuli and considering appropriate recovery periods. This is where VR plays an important role, as it has advantages over traditional elicitation methods of being effective and immersive compared to 2D images or videos [41, 50].

Today, existing works are dedicated to eliciting particular emotions and cannot be readily integrated for practical VR applications or customized for in-depth studies to understand different factors in emotion elicitation (*e.g.*, viewpoints, colors, interactive objects). In summary, our work makes the following contributions:

- **Dataset:** Our dataset consists of five VR scenes validated to induce five base emotions effectively. The scenes are based on the 360° video from two widely used and validated libraries ([40] and luVRe [57]). To the best of our knowledge, our work is the first to create a validated emotion-elicitation dataset of VR scenes that researchers and practitioners can use. The dataset allows us to replicate and directly compare the findings across different studies, promoting consistency and comparability across different studies, including Affective Computing, Human-Computer Interaction (HCI), Psychology, and other related fields. Our dataset aims to establish a common language and understanding within the research community as a shared methodology allows researchers to communicate, compare, exchange, and advance knowledge in emotion research effectively.
- **Evaluations:** We validated our dataset in a study with 160 participants recruited from Prolific with diverse demographics, including locations, age groups, and cultural backgrounds. Our results show the effectiveness and generalizability of our dataset in eliciting different and rich emotions, including different valence and arousal levels.
- **Insights:** We provide rich insights into the design of virtual environments for integrating emotional experiences in VR, such as customizing emotional experiences for individuals, which is challenging using conventional methods, including 360° videos. We consider various factors in designing emotional experiences in VR, such as interaction with virtual objects, personal preferences, and residual emotions. Our dataset is a tool to rapidly and seamlessly integrate emotional experiences in VR applications.

- *Weiwei Jiang is with Nanjing University of Information Science and Technology. E-mail: weiwei.jiang@nuist.edu.cn.*
- *Maximiliane Windl is with LMU Munich and the Munich Center for Machine Learning (MCML). E-mail: maximiliane.windl@ifi.lmu.de.*
- *Benjamin Tag is with Monash University. E-mail: benjamin.tag@monash.edu.*
- *Zhanna Sarsenbayeva is with the University of Sydney. E-mail: zhanna.sarsenbayeva@sydney.edu.au.*
- *Sven Mayer is with LMU Munich. E-mail: info@sven-mayer.com.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

2 Related Work

Our dataset provides a readily available tool for emotion studies and integrating emotional experiences in VR applications. To create and validate the dataset, we refer to previous work focusing on different methods for measuring, understanding, and eliciting emotions.

2.1 Measuring Emotions Using Self-Reports

Several affect measuring techniques were adopted from Psychology studies in HCI research, with the majority based on self-reports. The widely accepted Circumplex Model of Affect [51] conceptualizes emotions on two dimensions: Valence and arousal. Another model by Ekman [20] classifies basic emotions, though debates exist regarding the universality of emotional expressions and the cultural context of emotions. The Positive and Negative Affect Schedule (PANAS) [69] combines valence and arousal into positive and negative affect measures. Still, researchers argue it may misinterpret pleasure-driven positive affect, particularly in scenarios that induce high levels of anger, leading to confusingly high positive PANAS scores [17, 49, 70]. The Self-Assessment Manikin (SAM) [9] uses three dimensions (pleasure, arousal, dominance) depicted through manikin figures, providing a more comprehensive and intuitive assessment. The Photographic Affect Meter (PAM) [49] employs images to efficiently validate emotions based on valence and arousal, making it suitable for frequent assessments. In particular, recent studies have highlighted the robustness and reliability of online validation methods using SAM for creating affective virtual environments [46]. In our study, we adopt SAM for its multidimensional nature and intuitive icons, as image-based scales are more suitable for VR than text-based questionnaires, including valence, arousal, and dominance that have been widely used in literature for self-reporting emotions [6].

2.2 Sensing Emotions

The development of wearables has stimulated attempts to quantify emotions using sensor data, such as skin conductance for arousal [31]. However, its interpretation can be ambiguous as increased conductance can indicate negative and positive excitement [3, 30]. Also, haptic feedback facilitates emotional expression and recognition [4, 48], with potential enhancement through auditory cues [28]. In addition, researchers found that electric muscle stimulation (EMS) could communicate emotions via gestures [24, 29], and thermal feedback could provide a multidimensional impact, with cold stimuli eliciting more negative responses [28] and a vector model better capturing the relationship between temperature and emotions than the traditional circumplex model [72]. Furthermore, using multiple physiological signals, including electroencephalography (EEG), electrodermal activity (EDA), and heart rate variability (HRV), Gupta et al. [27] developed a real-time emotion recognition model using physiological signals to enhance user experience, demonstrating the effectiveness of empathic virtual environments in improving emotional and cognitive engagement.

While capturing natural emotional experiences in the wild is desirable [25, 65], especially given the increasingly important role of digital devices in our emotional lives [37, 61], *i.e.*, indicated by the reliance on technology during challenging times [42, 67], the complexity of emotions calls for a level of control of the environment. VR provides this control and enables the use of wearables and mobile devices for detecting emotions and moods [52, 59, 66, 74, 75].

2.3 Eliciting Emotions

Numerous techniques exist for eliciting emotions, classified into visual stimuli, auditory stimuli, autobiographical recall, situational procedures, and imagery [60]. Visual stimuli include the International Affective Picture System (IAPS) images [38] with standardized valence scores, and 360° videos [57] providing realistic everyday scenes and encounters. Auditory stimuli like fast-tempo music, major harmonies, and dance-like rhythms can effectively evoke happiness. In contrast, fearful classical music elicits fear, and slow-tempo, low-volume, minor-key music induces sadness [60]. Background music in VR has also been shown to elicit emotions [35]. Other methods include autobiographical recall and imagery for inducing emotions like anger, happiness, fear, disgust, and sadness [60], and situational procedures effective for anger, surprise, fear, and happiness. Factors like color associations [5], facial expressions of avatars [32], and specific VR scenarios [23] can influence emotional states. Phobic stimuli, horror movies, and scary content can elicit fear, while humorous films and comedy induce happiness [60].

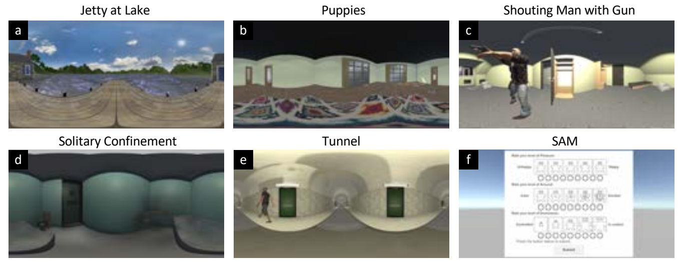

Fig. 1: The modeled scenes for emotion elicitation. The scene names for each corresponding target emotion are a) Jetty at Lake – high valence, low arousal; b) Puppies – high valence, low arousal; c) Shouting Man with Gun – mid-to-low valence, high arousal; d) Solitary Confinement – low valence, low arousal and e) Tunnel – mid-to-low valence, mid-to-low arousal. After experiencing each scene, the participant is asked to asses the scene using the f) Self-Assessment Manikin (SAM).

In addition, recent studies have explored the design of VR environments focusing on eliciting specific emotions. For example, researchers designed interactive virtual environments aimed at inducing complex emotional responses such as awe [15]. Moreover, previous studies have validated immersive videos for their effectiveness in evoking awe [14] and demonstrated the utility of VR in eliciting emotional reactions through carefully designed virtual environments [18, 19, 45, 46].

While these approaches exist, there is a lack of datasets providing an immersive, interactive, and customizable way to elicit emotions in VR. Our dataset aims to fill this gap by offering validated VR scenes that can be readily integrated into applications and studies, allowing for the exploration of different factors in emotion elicitation (*e.g.*, viewpoints, colors, interactive objects). Unlike existing methods such as 360° videos [54], our dataset provides rich emotional experiences and flexibility in emotional experience design, enhancing both research and application development in the field of emotion elicitation.

3 Scene Modeling

To ensure the ecological validity of our virtual environments, we modeled scenes from existing validated 360° videos. This approach allowed us to build upon established emotional triggers while leveraging the enhanced immersive and interactive capabilities of VR. In particular, we first selected five scenes from two validated 360° video datasets widely used in previous studies. We selected the videos from a pool of 158 video clips: A public dataset consisting of 73 video clips [40], and a separate dataset consisting of 85 video clips (luVRe) [57]. Both datasets have been validated in VR studies [40, 57].

The selection was made by five researchers with experience in emotion research and affective computing from different locations (Europe, Asia, and Oceania) and different cultural backgrounds. The selection process was based on the following criteria:

- *Effectiveness: Is the scene effective enough to elicit emotions?* – To elicit emotions effectively, we selected scenes that show pronounced valence and arousal values, *i.e.*, values that are far from the neutral emotion on the Russel Circumplex Model of Affect [51]. We did not consider dominance due to the absence of reference values in one 360° dataset [40]. Instead, we focused on valence and arousal, which are the key dimensions in the SAM surveys.
- *Complexity: Is the scene moderately complex to allow deployment on most VR devices?* – Considering the visual fidelity and computing resource requirements, we aim to select scenes that are not too complex and can be deployed in most VR devices (*i.e.*, without excessive details that are computationally expensive to render).
- *Generalizability: Can the scene elicit emotions in a general sample?* – As we strive to provide an asset that can be used for users with

different demographics, we aim to select scenes that elicit emotions in a general sample from diverse demographics (*e.g.*, gender, age, and cultural background).

- *Diversity: Can each scene elicit a distinct emotion?* – Our dataset aims to provide rich emotional experiences by eliciting different emotions. In particular, we aim to select scenes targeting different emotions in Russell's Circumplex Model of Affect [51].
- *Potential Applications: Can the scene be used in different VR applications?* – As a primary advantage of virtual environments over 360° videos, our dataset enables researchers and practitioners to integrate emotional experiences into their VR applications. Hence, we aim to select scenes that can be used in typical VR applications to enhance their use cases (*e.g.*, education, gaming, healthcare, tourism, etc.).

The selection process was conducted over three virtual meetings and two offline meetings. In particular, the selection process included three phrases: 1) Brainstorming – suggesting as many scenes as possible that may fulfill the selection criteria; 2) filtering – excluding the scenes that are difficult to model as VR scenes; 3) selecting – selecting the scenes that could fulfill the criteria without ambiguity, *i.e.*, consensus must be reached among all researchers. Consequently, we chose five scenes that fulfill the aforementioned criteria and represent different emotions. For the targeted emotions, we referred to the evaluation results (valence and arousal) of the corresponding 360° videos as reported in [40, 57].

To transform these 360° videos into VR scenes, we created 3D models and environments as digital twins of the original video, ensuring that the essence and emotional impact of the original videos were preserved. In particular, we used Blender and Unity to accurately model the scenes, incorporating interactive elements to enhance immersion and engagement. We detail each modeled scene below. In this paper, we refer to "user" as the subject who uses the VR device (*e.g.*, participants) and "player" as their virtual representation in the virtual space.

Jetty at Lake The player is spawned on a jetty by a lake next to a stone house, surrounded by tree-covered hills. This aims to elicit emotions with *high valence* and *low arousal*.

Puppies The player is spawned in a room with three puppies. The puppies walk playfully through the room. The room is inside a furnished house by the hills. This scene aims to elicit emotions with *high valence* and *low arousal*. Although this scene aims to elicit similar emotions as *Jetty at Lake*, researchers reported high valence and high arousal emotions in different datasets such as [40].

Shouting Man with Gun The player is spawned in a furnished and well-illuminated attic in a quiet neighborhood. After a while (adjustable, ~ 25 seconds in our settings), a man breaches the front door while screaming and aims a pistol at the player. This scene aims to elicit emotions with *mid-to-low valence* and *high arousal*.

Solitary Confinement The player is spawned in solitary confinement, with a semi-broken flashing light, a toilet set, and a single bed. The user can hear a heavy door closing and locking outside the room. This scene aims to elicit *low arousal* and *low valence* emotions.

Tunnel The player is spawned in the middle of a long tunnel, illuminated by yellowish lights, with a few pedestrians ($N = 3 \sim 5$, randomly spawned) passing by occasionally. There are two doors on both sides of the tunnel. We selected this scene to elicit emotions with *mid-to-low valence* and *mid-to-low arousal*.

We summarize the affect values of each scene in Table 1 and illustrate the modeled scenes in Fig. 1. In our dataset, we did not include a scene targeting high valence and high arousal due to ethical considerations. In particular, we aim to provide a dataset that can be used in more naturalistic scenarios without researchers' supervision (*e.g.*, out of the laboratory, detailed in Sec. 4). High arousal, high valence scenes are potentially harmful to the participants because they cause cybersickness [44] (*e.g.*, roller coaster, canyon swing [40]) or involve intense body motions that may result in the participant losing their balance or falling [12] (*e.g.*, speed flying, walk the tight rope [40]). While currently not considered, our dataset can be easily expanded by such scenes in the future when these issues can be better addressed.

Tab. 1: Details of the modeled scenes for emotion elicitation.

Scene Name	Visual Description	Sound Effect	Targeted Emotion	Ref
Jetty at Lake	A jetty in front of a stone house by a lake within hills covered by trees and grass.	Water flow	V: High A: Low	[57]
Puppies	A spacious and furnished room with several puppies walking around.	(Quiet)	V: High A: Low	[40, 57]
Shouting Man with Gun	A furnished attic. A man breaches the front door while screaming loudly and aims at the players using a pistol after a certain time.	Man shouting	V: Mid-to-low A: High	[57]
Solitary Confinement	A gloomy cell with a flashing light, a toilet set, and a single bed.	Water dropping	V: Low A: Low	[40]
Tunnel	A long tunnel illuminated by yellowish lights with a few pedestrians passing by.	Footsteps	V: Mid-to-low A: Mid-to-low	[57]

V – Valence, A – Arousal

For each scene, we predefined an area around the spawning position where the players can move around using teleportation. The teleportation area can be adjusted for different VR applications. We also include corresponding sound effects for each scene, *e.g.*, footsteps, water flow, a man shouting, and water dropping, as detailed in Table 1. The scenes are developed using Unity 2021.3 Long-Term-Support (LTS) version for compatibility. We also optimized the graphic performance to avoid cybersickness due to low frame rate [36]. All 3D models are either imported from open-sourced assets licensed under CC BY-SA, or self-modeled using Blender.

Furthermore, to validate the effectiveness of our scenes, we developed a VR study application. The VR application uses the OpenXR framework. For data collection, all data is streamed and stored on a cloud server running a customized service (developed using the Django framework[1]). We open-sourced the scenes and the VR application in https://github.com/HighTemplar-wjiang/VR-Dataset-Emotions.

4 User Study

The main objective of our work is to create a standardized dataset of different scenes represented by virtual environments to elicit specific emotions. In this section, we describe the methodology for creating and evaluating the dataset. In particular, we first selected five scenes from validated 360° videos that elicit different emotions. Then, we modeled the selected scenes as virtual environments that can be adopted in VR. Finally, we designed an out-of-laboratory user study to evaluate our scenes and validate our dataset. The user study was approved by the Ethics Committee of the Faculty of Mathematics, Computer Science and Statistics, LMU Munich (number: EK-MIS-2023-149).

In particular, for a more thorough comparison and validation of our dataset, we run a study using a mixed design. Each participant is asked to experience all the five scenes (within-group). Using a between-group factor, we compared the effect of our dataset against the original 360° videos as different stimuli used for eliciting emotions.

Furthermore, to validate the effectiveness of our dataset, we aim to answer the following two research questions (RQs):

RQ 1 How effectively can the developed VR dataset elicit specific emotional responses in diverse user groups?

RQ 2 What are the unique characteristics of the VR environments in the dataset that contribute to emotion elicitation, compared to traditional 360° video environments?

4.1 Apparatus and Protocol

To validate our dataset, we conducted user studies in non-laboratory settings on the Prolific platform. The participants were recruited via

[1] https://www.djangoproject.com/

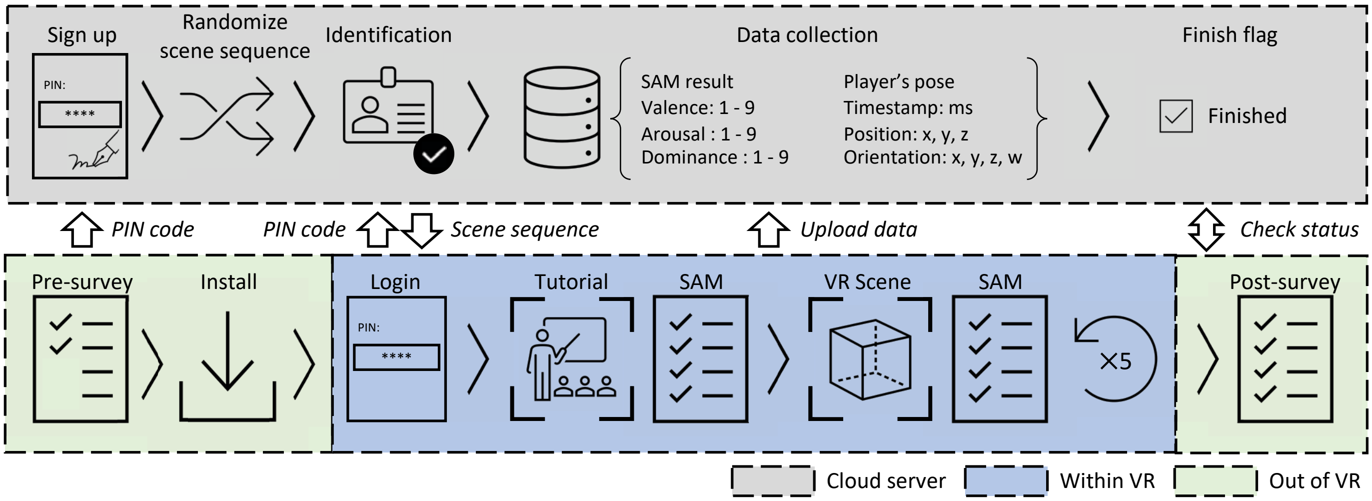

Fig. 2: Illustration of the protocol for the user study (bottom) and the corresponding data collection process with the cloud server (top). Some study elements were conducted outside of the VR environment (bottom, light green color), while the main study parts happened within the VR environment to not break the participants' immersion (bottom, dark blue color).

crowdsourcing with diverse backgrounds, and the study was conducted remotely without our supervision. Further, we used three tools for data collection– Prolific for participants' demographics, Qualtrics for pre- and post-surveys (detailed in the following section), and a server for logging the VR interactions.

The protocol is illustrated in Fig. 2. We asked participants to answer a pre-study questionnaire and install our VR app before the study. Then, the participants experienced all five scenes in random order. Finally, we asked the participants to answer a post-study questionnaire. Without our supervision, we developed and deployed a customized service on a cloud server to monitor the study flow and data collection. We summarize the protocol as follows:

Pre-survey – After receiving the study invitation through Prolific, the participant first answers a pre-survey questionnaire. This is where each participant is informed of their rights and has to confirm that they have a Meta Quest 2 VR headset, that they are willing to install our VR app to run the study, and that they consent to participate. Then, the participant creates a unique identifier (*e.g.*, PIN code) to log into our VR app. The server stores the identifier and then generates a randomized scene sequence for each participant. We also provided participants with instructions on how to answer the SAM questionnaire [9], using icons under the scores that can be intuitively understood, as illustrated in Fig. 1(f) [6, 46].

Install – The participant downloads and installs our VR app. In the process, we include an official video to give detailed instructions.

Login – The participant starts the application in VR and inputs their self-generated unique identifier (*e.g.*, PIN code) to log in to our VR app. The server validates the identifier and sends the corresponding scene sequence to the VR app.

Tutorial & SAM – After successful login, the tutorial instructed them to adjust their headset's volume. Afterwards, they can correctly hear a verification code we provided over the HMD's speakers, learn basic controls, and are informed of the study flow. The tutorial scene includes a 9-point SAM questionnaire for a baseline.

Sequence: VR Scene & SAM – The participant experiences all scenes for at least 30 seconds in random order. Then, the door nearest to the participant opens. The participant can either stay for further exploration or leave the scene by teleporting to the door. For the 360° video condition, the scene automatically ends. After each scene, they had to fill in the SAM scale.

Post-survey – After finishing all the scenes, the participant exits the VR application and is asked to finish a post-study survey where they are asked to describe their feeling about each scene. The survey tool automatically checks the study status with the server before allowing the participant to start the post-study survey.

The total time spent on the user study is 45 minutes per participant.

4.2 Participants

We only recruited participants on Prolific who were 18 years or older, had access to a Meta Quest 2, and were willing to install our applications. Upon finishing the study, we reimbursed our participants at a rate of £9 per hour for the time spent in the user study. In total, we recruited 160 participants (50 female, 110 male), aged between 18 and 66 years ($M = 31.81$, $SD = 10.50$) from various locations (see Fig. 3 for a detailed overview), who could understand in English, as required to use Prolific. We also set up multiple hurdles in our protocol (*e.g.*, PIN code, app install, and post-survey in Fig. 2) to ensure that the participants fully understand our instructions. A completion code was provided for the participants to make a valid submission in Prolific, only after they passed all the steps they completed the study. In addition, 142 participants used VR at least once per month (daily: 30, weekly: 72, monthly: 40), showing that they were familiar with VR.

5 Results

In the following, we detail the findings of our study and provide a comprehensive evaluation of our dataset. First, we analyze the emotion measurements of the self-reported SAM questionnaires. Subsequently, to further understand the emotion elicitation process, we investigate multiple aspects of how the participants interacted with the virtual environments. In particular, we explore the *time* participants engaged with each scene, their *virtual positions and orientations* in the virtual environments, and their *descriptions of their emotions* in each scene through topic modeling.

Fig. 3: Participants' demographics, including locations (left), age (right-top), and VR usage frequency (right bottom).

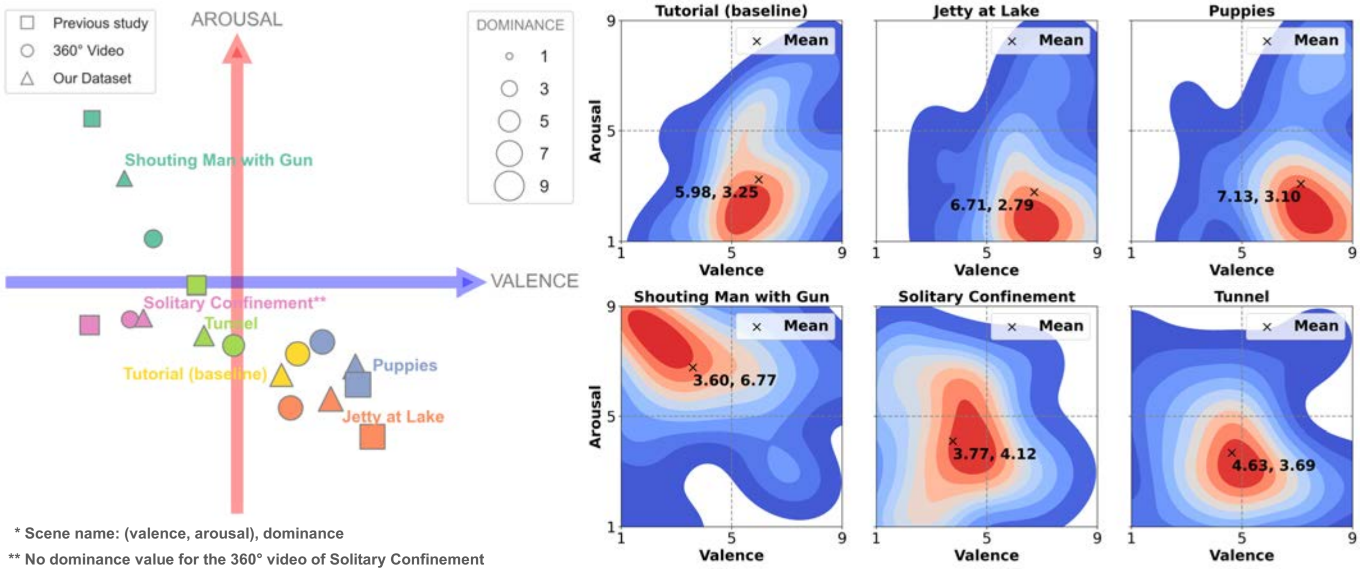

Fig. 4: Illustrations of emotional responses in the valence-arousal space. Left: mean values of valence, arousal, and dominance. Right: kernel density estimate plots of valence and arousal values for our dataset.

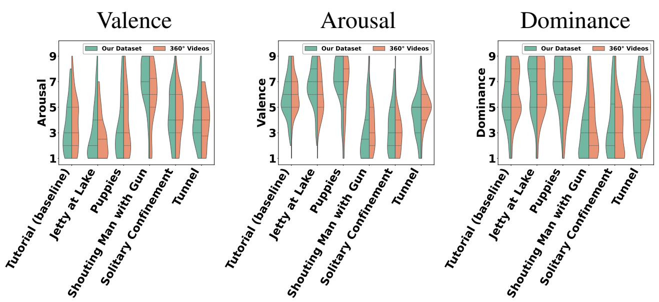

Fig. 5: SAM questionnaire results of within-VR study after experiencing each scene for valence, arousal, and dominance, respectively.

5.1 Emotion Elicitation Measurements

First, we analyzed the emotion measurements from the SAM questionnaires to evaluate how effectively our scenes elicited the emotions. In total, we received 960 SAM measurements: 80 Participants per Condition $\times$ 2 CONDITIONS $\times$ (1 baseline + 5 SCENES). The CONDITIONS refer to the use of either standard 360° videos or our interactive VR environment. The emotional responses in the valence-arousal space are illustrated in Fig. 4 and 5. The results illustrate that our VR environments elicit emotional responses comparable to the 360° videos, indicating the effectiveness of our VR scenes for emotion elicitation. In particular, high-arousal scenes like "Shouting Man with Gun" and positive-valence scenes like "Puppies" and "Jetty at Lake" show that our VR scenes (triangles) closely align with the emotional responses from the 360° videos (circles) validating the feasibility and impact of our dataset.

5.1.1 Impact of 360 Video vs. VR Environment

Firstly, we investigate the impact of CONDITION (levels: *Our Dataset* vs. *360 Videos*) on the SAM scale. After checking for normality, we used ART ANOVAs [73] to analyze the non-normally distributed results of the SAM questionnaires (Shapiro test [58], $p < 0.05$). As expected we found that *Arousal* is significantly influenced by SCENE ($F = (5,790) = 71.096$, $p < .001$), but not by the CONDITION ($F = (1,158) = .008$, $p = .928$) with interaction effects being statistically significant ($F = (5,790) = 3.407$, $p = .004$). Similarly, we found that *Valence* is significantly influenced by SCENE ($F = (5,790) = 191.252$, $p < .001$), but not by the CONDITION ($F = (1,158) = 74102.62$, $p = .978$) with interaction effects being statistically significant ($F = (5,790) = 3.407$, $p = .004$). Lastly, as expected we found that *Dominance* is significantly influenced by SCENE ($F = (5,790) = 71.096$, $p < .001$), but not by the CONDITION ($F = (1,158) = .008$, $p = .928$) with interaction effects being statistically significant ($F = (5,790) = 6.952$, $p < .001$).

Without the statistically significant main effect of CONDITION, we investigated the likelihood of CONDITION not influencing the measurements using Bayesian RM ANOVAs with JASP. The results[2] show a *substantial evidence* that CONDITION (*Our Dataset* vs. *360 Videos*) has no impact on the the measurement.

5.1.2 Impact of Scene

The above results indicate that there is no impact of CONDITION; however, the SCENE impacts the SAM measurements. Consequently, we will analyze the SCENES independently in the following. Thus, we conducted Wilcoxon signed-rank tests with Bonferroni correction for post hoc analysis [71], resulting in 10 post hoc pairs for each SAM dimension.

The post hoc test results show that, for all scenes, at least one SAM dimension has a statistically different median than the baseline (Tutorial), indicating that all the scenes effectively elicited non-neutral emotions. In particular, for our dataset, we made the following observations:

Jetty at Lake For this scene, the participants reported significantly different median values in valence compared to other scenes except for *Puppies* (M = 6.66 ± 1.65, Med = 7.00), showing moderately high valence levels. In the arousal dimension, the participants reported significantly different median values compared to the *Tunnel*, *Shouting Man with Gun* and *Solitary Confinement* scenes (M = 2.94 ± 2.12, Med = 2.00), showing relatively low arousal. In the dominance dimension, the participants reported significantly different median values compared to the *Tunnel*, *Shouting Man with Gun*, and *Solitary Confinement* scenes (M = 6.54 ± 2.11, Med = 7.00), showing moderately high dominance levels.

[2] *Arousal*: CONDITION ($BF_{10} = .171$, $error\% = 1.172$), SCENE ($BF_{10} > 100$, $error\% = .563$), and interaction ($BF_{10} > 100$, $error\% = 1.378$). *Valence*: CONDITION ($BF_{10} = .113$, $error\% = .985$), SCENE ($BF_{10} > 100$, $error\% = .561$), and interaction ($BF_{10} > 100$, $error\% = 1.338$). *Dominance*: CONDITION ($BF_{10} = .118$, $error\% = 1.201$), SCENE ($BF_{10} > 100$, $error\% = .428$), and interaction ($BF_{10} > 100$, $error\% = .953$).

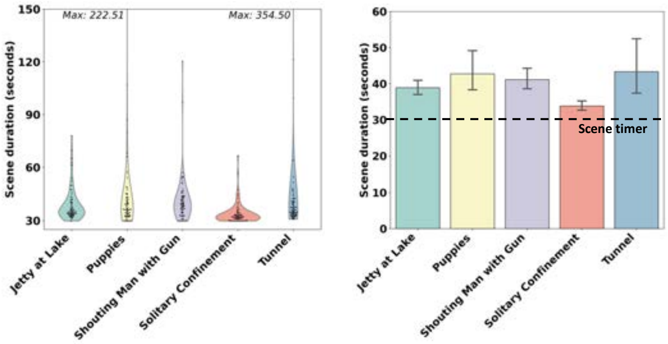

Fig. 6: Engagement time in different scenes. Left: Distributions of engagement time; Right: Mean engagement time.

Puppies First, the medians of reported values in all three dimensions are not significantly different compared to the *Jetty at Lake*, indicating similar emotions elicited in these two scenes. In particular, the valence (M = 7.10 ± 1.78, Med = 7.00) and dominance (M = 6.72 ± 1.97, Med = 7.00) values are similar to the *Jetty at Lake* scene. However, different to *Jetty at Lake*, in the arousal dimension, participants reported significantly different median values compared to only the *Shouting Man with Gun* and *Solitary Confinement* scenes (M = 3.51 ± 2.42, Med = 3.00).

Shouting Man with Gun For this scene, we can only observe no significant difference in median values for the valence and dominance compared to the *Solitary Confinement* scene. For all other reported values, we can find significantly different median values compared to the other scenes in valence (M = 3.00 ± 2.10, Med = 2.50), arousal (M = 6.85 ± 2.00, Med = 7.00) and dominance (M = 3.05 ± 2.11, Med = 3.00), clearly indicating moderately low valence and high arousal, with low dominance levels.

Solitary Confinement Finally, we found a significant difference in median values for *Solitary Confinement* in multiple dimensions, except when compared with *Shouting Man with Gun* of both valence and dominance and with *Tunnel* in arousal. Overall, the participants reported relatively low levels of valence (M = 3.34 ± 1.88, Med = 3.00), arousal (M = 4.38 ± 2.07, Med = 4.00), and dominance (M = 3.71 ± 2.48, Med = 3.00).

Tunnel After experiencing this scene, the participants reported significantly different median valence values compared to other scenes, with a mid-to-low value (M = 4.41 ± 1.81, Med = 5.00). For the arousal dimension, the participants reported significantly different median values than for *Jetty at Lake* and *Shouting Man with Gun* (M = 4.06 ± 1.99, Med = 4.00), showing relatively low arousal levels. The participants reported significantly different median values for the dominance dimension compared to other scenes except for the baseline (M = 4.66 ± 2.20, Med = 5.00), showing neutral dominance levels.

Synthesizing the above, we find that all five scenes successfully elicit the targeted emotions. While some participants reported *high arousal* and *high valence* for the *Puppies scene*, the overall ratings were similar between *Puppies* and *Jetty at Lake*.

5.1.3 Comparison to Prior Work

Compared to previous studies that were conducted in a laboratory, we observe several differences in the mean ratings with our crowdsourced study using the original 360° videos[3]. In particular, for the *Shouting Man with Gun* and *Tunnel* scenes, the valence and arousal values are lower than reported in the previous lab studies [40, 57]. In contrast, for the *Puppies* and *Jetty at the Lake* scenes, our study shows higher arousal values and lower valence values. Nevertheless, the elicited emotions remain in the same quarters of the Circumplex model (Fig. 4).

[3]We cannot run statistical tests since we do not have the experimental data of previous studies.

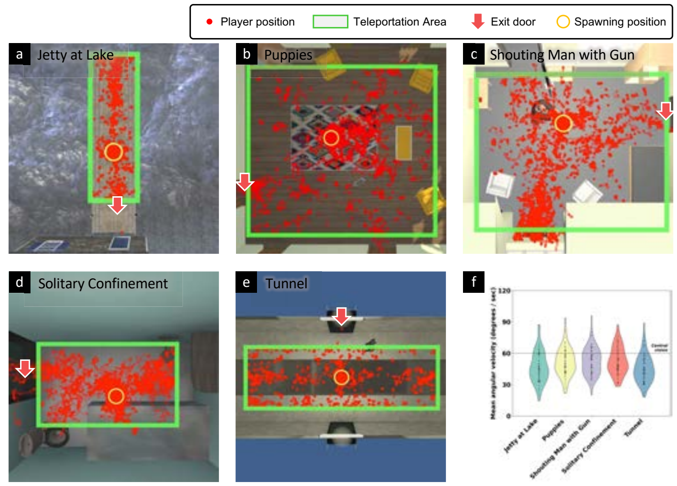

Fig. 7: Virtual positions (a-e) and mean angular speed (f) in different VR scenes. The positions are sampled every 0.1 seconds (sampling frequency = 10 Hz).

5.2 Engagement Time in VR Scenes

Based on the emotion measurement results, we investigate participants' engagement time for each scene to further understand the emotion elicitation process. The engagement time (in seconds, errors: ± 0.1 sec) has been collected during our study for each scene. In particular, we consider the engagement time a proxy for interest and immersion. Following our protocol, each scene had a 30-second minimum engagement time. However, the participants could choose to remain in the scenes longer.

Similarly, as the distribution of our time engagement data is significantly different from a normal distribution (Shapiro test, $p < 0.05$), we ran a Friedman test among all five scenes. The result shows that there are significant differences between the medians. Consequently, we ran Wilcoxon signed-rank tests with Bonferroni corrections applied to identify the scene pairs with different median engagement times. The results are illustrated in Fig. 6.

We can observe that only *Solitary Confinement* has significantly different medians compared to other scenes. We can further observe that the engagement time is lower than in other scenes (M = 33.78 ± 5.83 sec, Med = 32.13 sec). Considering that the minimum engagement time is 30 seconds, this indicates that the participants left the scene immediately after the 30-second timer. This aligns with the previous observation that *Solitary Confinement* elicited emotions at relatively low levels in all SAM dimensions (valence, arousal, and dominance) overall. In comparison, participants reported neutral dominance levels for the *Tunnel* scene with low valence and arousal levels.

5.3 Engagement with Virtual Environments

We further consider how participants engaged with our virtual environments by observing the virtual positions and orientations. First, we show the positions the participants were primarily interested in for each scene. Specifically, we focus on the XZ plane that represents the positions from the top-down view (in Unity, the Y dimension represents the "height"). We illustrate the position samples in Fig. 7. Based on the positions and our subsequent topic modeling, we can make the following observations:

Jetty at Lake Besides the spawning point, we find a clear cluster towards the end of the jetty. This implies that the lake attracted the participants' attention, leading to higher reported valence values.

Puppies Most positions are clustered by the spawning point where the puppies are, indicating that the puppies attracted the attention, which might elicit pleasant feelings expressed in the high-valence emotions.

Shouting Man with Gun We observe that the participants explored the scene actively while also paying more attention to the window. This implies the participants were first confused by the scene and then surprised by the shouting man with a gun afterward. Consequently, this might elicit high-arousal emotions.

Solitary Confinement We find a high-density cluster at the spawning point, indicating small and frequent movements. Unlike the Tunnel scene, the high-density cluster implies that the participants might be anxious, as we can also observe a slightly low valence but relatively higher arousal levels compared to the Tunnel scene.

Tunnel The positions are sparsely distributed over the teleportation area without clear clusters beside the spawning point. This scene did not elicit high arousal levels, and we can infer that this implies the participants were not interested in exploring the area.

Overall, we can observe different position distribution patterns for all scenes. The patterns may imply the underlying process of emotion elicitation. In addition, we compute the mean angular speed for each participant in each scene, representing how frequently they moved their head to adjust their viewpoints (in *radians per second*), see Fig. 7 (f). Specifically, considering 60° as the central visual field [64], we can observe that participants changed their viewpoints frequently ($mean = 49.68 \pm 13.67$ $degrees/sec$), indicating that the participants changed their central vision to engage with the virtual environments.

5.4 Topic Modeling

We further investigate the emotion elicitation process by analyzing the qualitative results from the participants' descriptions of their emotional responses, collected from our post-study Qualtrics questionnaire. Due to the subjective nature of emotional experiences, we aim to show a more in-depth observation of the emotion elicitation process [41] and explore factors that may cause particular emotions.

In particular, we conduct topic modeling to find the most frequent words participants used to describe a scene. The process involves word filtering, lemmatization, and a Latent Dirichlet Allocation (LDA) analysis. First, as we focused on the words that describe emotional experiences, we filtered the words to only keep nouns and adjectives. We also removed stop words and uninformative words, such as *i.e.*, *feeling*, *scene*, *little*, or *bit*. Then, we lemmatized the words (*i.e.*, grouping inflected words and returning them to their dictionary form, *e.g.*, plural to singular) and conducted a LDA analysis to pinpoint the top-3 topics across the participants [43]. Each topic includes several words that frequently appear together, implying a similar topic. Here, we focus on the top five words for better relevance and coherence. We also show a word cloud that illustrates the most frequent words among all participants for each scene.

5.4.1 Topics: Jetty at Lake

The Jetty at Lake scene elicited emotions with high valence (M = 6.66 ± 1.65, Med = 7.00), low arousal (M = 2.94 ± 2.12, Med = 2.00) and high dominance (M = 6.54 ± 2.11, Med = 7.00) levels, see Fig. 8 for participants' reflections. We can observe that participants frequently mentioned "nice" ($N = 15$), "water" ($N = 20$), and "calm" ($N = 23$), showing a strong connection between their emotions and the scene.

Furthermore, the LDA yields three main topics that underscore connections between the scene's ambiance and the user's emotion, predominantly *calmness* and *relaxation*. Participants felt peaceful and happy as immersed in nature – *"the scene made me feel peaceful and happy because there was a beautiful scene of nature all around me"* (P36). Further, these emotions were accentuated by the ambient sound of water, as one remarked, *"this scene was very calming and relaxing. Peaceful, quiet with only splashes of water and sounds of nature in the background to listen to"* (P24). This underscores participants' emotional responses to the scene's ambiance as calm and peaceful.

However, some participants experienced mixed emotions. For example, P11 associated the lake with *fear* due to their personal experiences – *"I am scared of water because I can't swim"*. Also, residual emotions from the previous scene, especially the intense *Shouting Man with Gun*, may also shape the emotional reactions such as *uneasy*, as remarked by P56 – *"I still thought it was too peaceful and something was going to jump at me"*. This dichotomy of experiences underscores the effectiveness of our dataset, aligning with our objective of creating VR environments with high ecological validity.

5.4.2 Topics: Puppies

The *Puppies* scene elicited similar emotions as *Jetty at Lake*, with high valence (M = 7.10 ± 1.78, Med = 7.00), low arousal (M = 3.51 ± 2.42, Med = 3.00) and high dominance levels (M = 6.72 ± 1.97, Med = 7.00). As illustrated in Fig. 8, participants frequently mentioned "room" ($N = 12$), "happy" ($N = 22$) and "cute" ($N = 15$), with their main impression is positive.

Furthermore, LDA yielded three main topics that highlight how both the ambiance and the active objects (puppies) elicit *calm* and *happy* emotions. In particular, participants underscored their fondness for the ambiance of this scene, represented by the cute puppies and the warm feeling in the room. For example, participants mentioned *"I felt like I was at home with some cute dogs and felt peaceful"* (P05), and *"The room felt warm and inviting. Overall, this scene was relaxing"* (P31). This generally reflects that the participants related this scene to positive emotions. This also aligns with the previous observations in Sec. 5.3, showing that participants focused mainly on the puppies.

In addition, we find that although the *Puppies* scene elicited similar emotions as *Jetty at Lake*, the participants perceived it differently. In particular, the participants' remarks on the puppies show that *calm* and *happy* emotions can be elicited beyond environmental settings, as P78 highlighted – *"I felt good because I like puppies."*. Also, participants' mixed emotional responses underscored the residual effect from previous scenes such as *Shouting Man with Gun*, as mentioned by P42 – *"Being the first scene, I felt it was a trick and was on high alert"*. This shows how active objects contribute to the emotion elicitation together with the ambiance and highlights how our dataset provided rich emotional experiences.

5.4.3 Topics: Shouting Man with Gun

As we have seen above, participants frequently highlighted their emotional residues from the *Shouting Man with Gun* scene with low valence (M = 3.00 ± 2.10, Med = 2.50), high arousal (M = 6.85 ± 2.00, Med = 7.00) and low dominance (M = 3.05 ± 2.11, Med = 3.00). Our post-study survey confirms this, where most participants perceived this scene as the emotionally most impactful ($N = 24$). Also, as illustrated in Fig. 8, the participants underlined "room" ($N = 17$), "man" ($N = 19$), "window" ($N = 14$) and "gun" ($N = 14$) most frequently, underpinning their primary impression of this scene.

Based on the LDA results, we can observe three main topics that highlight how the participants perceived this scene as *surprising* and *scary*. Specifically, participants' reflections commonly mentioned being "jumpscared" while exploring the scene, especially near the window, as we observed in Fig. 8. For example, participants mentioned *"I looked out of the window and then heard a yell. I immediately got the chills and whipped around me to see a guy aiming a pistol at me"* (P24).

Furthermore, we observe other effects that may impact the emotional elicitation process, such as the ambiance and the active objects. In particular, our virtual environment was modeled with a peaceful ambiance. This made the participants lower their alertness, as one highlighted – *"I was peaceful admiring the view, it felt nice until the guy entered and yelled, I couldn't relax after that"* (P02). Also, the interactiveness of our dataset contributed to the emotion elicitation, as P73 mentioned – *"As I moved around the room, the man kept pointing the gun at me, it was very disturbing."*

Overall, the *Shouting Man with Gun* scene distinctly stands out in its ability to elicit high-arousal, negative valence emotions, see Fig. 8. In particular, the scene's design, characterized by a sudden and unexpected element (*i.e.*, jumpscare), effectively elicited the targeted emotion. Our participants' further responses, such as the ambiance and active objects, underscored the immersiveness and interactiveness of our dataset for emotion elicitation in VR.

5.4.4 Topics: Solitary Confinement

The Solitary Confinement scene aimed to elicit emotions with low valence (M = 3.34 ± 1.88, Med = 3.00), low arousal (M = 4.38 ± 2.07, Med = 4.00), and low dominance (M = 3.71 ± 2.48, Med = 3.00). As illustrated in Fig. 8, participants had various feelings about this scene,

Fig. 8: Word cloud based on the frequency of participants' descriptions on experiencing our modeled VR scenes.

such as scary ($N = 8$), claustrophobic ($N = 9$), and small ($N = 9$). In general, participants described this scene as negative.

Then, based on the LDA results, we can further observe three main topics that underscore how this scene elicited emotions, including *unhappy*, *scary*, and *uncomfortable*. In particular, participants perceived this scene as *"claustrophobic"* (P11) and also felt low dominance, such as *"no control of my location, or of being able to control or stop the blinking light above"* (P46), or *"not being able to do anything"* (P48). This demonstrates how our virtual environment elicited emotions with low valence, low arousal, and low dominance.

Furthermore, some participants expressed mixed emotions, such as "calm", with an underlying awareness that the confinement was not real, as one remarked, *"I was calm because I knew that it is not real thing"* (P32). Such emotions may also link with their personal experience, as P62 mentioned – *"I was pretty calm, though. I've been there before"*. These experiences underscore how the participants related their emotions to real-life experiences, with cognitive resistance to the negative emotions.

5.4.5 Topics: Tunnel

The *Tunnel* scene elicited more positive emotions than the *Solitary Confinement* scene by placing the participants in a long tunnel with yellowish illuminations and occasional pedestrians. The scene elicited emotions with mid-to-low valence (M = 4.41 ± 1.81, Med = 5.00), mid-to-low arousal (M = 4.06 ± 1.99, Med = 4.00) and neutral dominance (M = 4.66 ± 2.20, Med = 5.00). Through the participants' descriptions, as illustrated in Fig. 8, we observed mixed emotions described as "scary" ($N = 8$), "uneasy" ($N = 9$), "anxious" ($N = 5$) and "calm" ($N = 5$).

Furthermore, the LDA results suggest the three predominant topics that focus on how participants perceived different emotions for this scene. Here, participants highlighted how the physical appearance of the tunnel elicited their feeling of dominance. For example, a participant highlighted: *"I do not like how narrow and long the tunnel is. I feel uneasy"* (P12). Also, the theme of "control" emerged commonly, with remarks such as *"felt confined, a bit anxious, not in control because of the confinement"* (P46). Also, participants found the scene unsettling because of the active objects (pedestrians), as reflected by remarks such as being *"eerie"* due to *"stranger coming through from both sides"* (P24), and *"the people constantly walking and the very little space I could move made me a bit anxious"* (P79). In addition, a few participants mentioned that previous scenes, such as the *Shouting Man with Gun* scene[4], influenced their reactions, making them feel *"uneasy"* and *"like I could get attacked at any time"* (P49).

In summary, participants' feedback on the Tunnel scene highlighted three factors of the emotion elicitation: 1) the physical appearance of the scene, 2) the presence of pedestrians, and 3) previous experiences from other VR scenes, in particular, the *Shouting Man with Gun* scene, as also mentioned in other scenes. Remarkably, some participants had mixed emotions from at least one of these factors, underscoring the variety of the emotion elicitation process for different subjects.

5.4.6 Overview of topic modeling

In summary, across all five scenes, several consistent topics emerged. We found that the virtual environments succeeded in eliciting the targeted emotions. In particular, we can observe how the ambiance and the active objects elicit emotions, demonstrating the immersiveness and interactiveness of our dataset. However, participants might perceive the same scene differently due to their personal experiences or preferences, such as the *Jetty at Lake* and *Solitary Confinement* scenes. This shows the nuanced and subjective nature of emotion elicitation, also in VR settings. We also found emotional residual effects from previous scenes, particularly from the intense *Shouting Man with Gun* scene, underscoring the interconnections of emotional experiences in VR. These observations feature our dataset as an immersive and effective tool for creating rich emotional experiences.

6 Discussion

There are numerous methods for eliciting basic emotions in psychology laboratories. Nevertheless, not all existing methods are equally efficient in inducing target emotions, resulting in a researcher dilemma of choosing the most efficient method [60].

6.1 Effectiveness of Our Dataset

Our dataset elicits various emotions through validated 360° videos. User study observations confirm its effectiveness. Notably, more factors may influence this effectiveness compared to other methods like 360° videos [40], 2D videos [55], and images [38]. Our dataset offers enriched emotional experiences by allowing active exploration of virtual environments, unlike passive methods where participants remain in fixed positions. This interactivity likely enhances the emotional impact, as shown by participant feedback indicating a more immersive and realistic experience. Our dataset generally elicited emotions with higher valence and lower arousal than 360° videos. This may be due to our diverse participant pool from an online platform, leading to a more natural and generalizable study setting compared to lab-based evaluations. The familiar VR setups used by participants likely minimized the observer effect [1], and the interactive nature of our dataset may have contributed to higher valence emotions. In summary, our dataset effectively provides rich emotional experiences for VR applications through its interactive and immersive approach.

6.2 Customizing Emotional Experiences in VR

Furthermore, besides our dataset's interactive and immersive nature, a distinctive feature is its customizability. Existing datasets such as 360° videos are difficult to customize as most of the materials are pre-recorded instead of modeled. In particular, our dataset can be customized to elicit different emotions by simply adjusting the settings of the virtual environments. For example, participants highlighted how the ambiance affected their emotions, such as the overall appearance and the background water sound elicited high valence in the *Jetty at Lake* scene, while the ambiance of the *Tunnel* elicited low valence emotions. This can be easily achieved by manipulating the visual and auditory profiles of the virtual environments in our dataset. For instance, researchers and practitioners can replace the sound file, adjust the darkness for the water, or change the skybox to show different weather, such as a storming day to elicit low-valence emotions or a sunny day to elicit high-valence emotions. Such adjustments can be easily made using VR development tools such as Unity but can be extremely challenging for 360° videos.

Moreover, our user study shows that participants have different emotional responses to the same virtual environment. This is caused by their personal experiences or preferences, as aligned with prior research [2, 68]. For example, in the *Puppies* scene, some participants expressed that they did not like dogs. To elicit targeted emotions (*e.g.*,

[4]Note that the scene sequences were randomly generated (Fig. 2).

high valence) from these participants, we may customize the scene by replacing the dogs with their preferred animal (*e.g.*, cats). Similarly, for other scenes, we can readily adjust the virtual environment to elicit targeted emotions by updating the modeling or the interaction settings. To date, this cannot be achieved for 360° videos as such tools are not available for the public. To this end, our dataset enables researchers and practitioners to design their emotional experiences based on their requirements, allowing rapid and flexible customization and integration of emotion elicitation into practical VR applications.

6.3 External Factors for Emotional Experiences

Beyond the virtual environment settings, we found several external factors that may impact the emotional experiences. Notably, the emotional residual effects, participants' awareness of the artificial nature of virtual environments, and the interaction features may be significant for designing emotional experiences in VR. We found that participants underscored the emotional residual effects of the *Shouting Man with Gun* scene. The high arousal experience caused uneasy or anxious emotions, even in scenes with a calm ambiance, such as *Jetty at Lake* or *Puppies*. In contrast, such effects were not highlighted in other scenes, which aligns with previous work [7, 13, 56]. However, we also find that such effects did not significantly impact the emotion elicitation process for subsequent scenes. This may be caused by the immersive nature of VR, which can better overlap with previous emotions, showing the effectiveness of our dataset.

Nevertheless, we also observed that participants might be aware of the artificial nature of the virtual environments from their feedback, particularly in the *Solitary Confinement* scene. This might be caused by their impulse to escape from the scene, eliciting negative emotions. On the one hand, this may break the immersion of participants' experience due to their resilience to the unpleasant virtual environment, leading to less effective emotion elicitation. On the other hand, this may cause other emotions, such as anxiety, as the participants were confined within the scene for a minimum time until they could escape. As we can also find from the engagement time analysis in Sec. 5.2, participants spent less time in this scene. To this end, the elicited emotions may vary by adjusting the confinement time to avoid immersion breaks by shortening the time or eliciting higher anxiety levels by prolonging the time.

6.4 Implications for Research and Practice

We highlight that our dataset can transform research and practice around emotions. First, as different individuals may have different emotional sensitivities, thresholds, and responses to the same emotional stimuli, accounting for this individual variability poses a challenge when eliciting emotions. Our dataset can alleviate this issue by providing the ability to customize and personalize the emotional experiences for individuals to elicit targeted emotions, allowing us to conduct research more practically and flexibly. For example, researchers can adjust the parameters of models and their materials (*e.g.*, color, size, placement) to study further how corresponding factors affect the emotion elicitation process. Also, our results can be used as the baseline, providing a validated reference for such studies in the future.

Furthermore, our user study demonstrates the feasibility of conducting experiments outside laboratory settings without supervision. This can significantly accelerate research in related areas and open access to more diverse populations. In particular, eliciting emotions is vital in various research fields, while conducting such studies can be costly or time-consuming. Our dataset enables researchers to investigate emotional processes, study emotional disorders, and develop effective therapeutic interventions closer to a real-life experience, paving the way toward practical and realistic study settings.

The ability to elicit various emotions, such as being relaxed, happiness, surprise, and contentment, can significantly contribute to individuals' emotional well-being, quality of life, and overall psychological health. The ability to create immersive experiences that can help regulate emotions and evoke emotional responses opens up new possibilities, *e.g.*, for gaming [53], virtual therapy [47], and training [34].

6.5 Limitations and Future Work

Despite the rich features of our dataset, we found improvements for future exploration, including additional scene design, interaction design, evaluation, and usability. First, our dataset currently includes only five scenes, which can serve as "seed scenes" for creating more customized scenes. We did not include a scene from the first quarter of the valence-arousal space (*i.e.*, high valence and high arousal, such as joy) due to ethical considerations, *e.g.*, the risk of cybersickness. Although some participants exhibited high valence and arousal emotions during our study (see Fig. 4), this was limited to a few individuals. This aligns with previous studies reporting similar emotional responses for scenes like *Puppies* in Li *et al.* [40]. Additionally, our user study showed lower arousal and higher valence levels compared to 360° videos. Future work can leverage the customizability of our dataset to adjust scene parameters and explore factors affecting arousal and valence levels. Our database mitigates the practical challenges of VR scene development, providing a baseline validated dataset that complements existing 360° videos and supports future research in VR emotion elicitation.

Our dataset offers interactive virtual environments, but we did not enable object interaction despite participants expressing a desire to interact with the elements, *e.g.*, puppies, which could enhance or alter emotional elicitation, though they might also increase variations in emotional responses. Future work should explore enabling object interactions and incorporating multisensory stimuli to enhance the emotional experience [18], *e.g.*, tactile and olfactory elements.

We conducted our study online, resulting in high demographic diversity and more naturalistic settings than lab studies. However, this led to a non-gender-balanced sample due to the gender skew on the crowdsourcing platform. Future studies should focus on specific user groups for in-depth analysis of emotion elicitation, considering factors like age and cultural background [11]. We only used self-reported SAM questionnaires to measure emotions; future studies should include physiological measurements, *e.g.*, EEG, which has proven more accurate for emotion analysis [41]. We did not measure immersiveness or interactivity directly, assuming these as intrinsic features of VR scenes [62]. However, the exact levels of these features may affect emotion elicitation, especially in scenarios where external factors them.

7 Conclusion

We present a dataset with five VR scenes to induce five different emotions. We validated our dataset by running an out-of-laboratory user study on Prolific, where we recruited participants with diverse demographics. Our results show that our dataset can effectively elicit targeted emotions while providing rich emotional experiences for the participants. Our dataset allows VR researchers and practitioners to seamlessly and readily integrate emotional experiences into their applications and provides insights into designing VR applications for various emotional responses. Our work provides avenues for future emotion studies in HCI, psychology, and other related fields through immersive and interactive VR environments, accelerating emotion studies through immersive and interactive VR environments.

Open Science

We encourage future research to expand on our dataset, using it as a starting point to develop additional scenes. To support this, we are open-sourcing our VR scenes, study data, and evaluation scripts for reusability and reproducibility. Find the resources at `https://github.com/HighTemplar-wjiang/VR-Dataset-Emotions`.

Acknowledgments

SPONSORED BY THE

Federal Ministry of Education and Research

This work was supported in part by the NSFC under grants No. 62072004, the Startup Foundation for Introducing Talent of NUIST, and funded by the Federal Ministry of Education and Research (BMBF) and the Australia-Germany Joint Research Cooperation Scheme (DAAD grant number: 57600233).

REFERENCES

[1] J. G. Adair. The hawthorne effect: a reconsideration of the methodological artifact. *Journal of applied psychology*, 69(2):334, 1984. 8

[2] K. Annamary, G. S. Prathima, R. Sajeev, G. Kayalvizhi, V. Ramesh, and G. Ezhumalai. Colour preference to emotions in relation to the anxiety level among school children in puducherry–a cross-sectional study. *Journal of clinical and diagnostic research: JCDR*, 10(7):ZC26, 2016. 8

[3] E. Babaei, B. Tag, T. Dingler, and E. Velloso. A critique of electrodermal activity practices at chi. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. 2

[4] J. N. Bailenson, N. Yee, S. Brave, D. Merget, and D. Koslow. Virtual interpersonal touch: expressing and recognizing emotions through haptic devices. *Human–Computer Interaction*, 22(3):325–353, 2007. 2

[5] L. Bartram, A. Patra, and M. Stone. Affective color in visualization. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 1364–1374, 2017. 2

[6] A. Betella and P. F. M. J. Verschure. The affective slider: A digital self-assessment scale for the measurement of human emotions. *PLOS ONE*, 11(2):1–11, 02 2016. 2, 4

[7] J. Bisby and N. Burgess. Differential effects of negative emotion on memory for items and associations, and their relationship to intrusive imagery. *Current Opinion in Behavioral Sciences*, 17:124–132, 2017. 9

[8] G. A. Bonanno and C. L. Burton. Regulatory flexibility: An individual differences perspective on coping and emotion regulation. *Perspectives on psychological science*, 8(6):591–612, 2013. 1

[9] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994. 2, 4

[10] M. M. Bradley and P. J. Lang. Affective reactions to acoustic stimuli. *Psychophysiology*, 37(2):204–215, 2000. 1

[11] E. A. Butler, T. L. Lee, and J. J. Gross. Emotion regulation and culture: Are the social consequences of emotion suppression culture-specific? *Emotion*, 7(1):30, 2007. 9

[12] H. Chander, A. Shojaei, S. Deb, S. N. K. K. Arachchige, C. Hudson, A. C. Knight, and D. W. Carruth. Impact of virtual reality–generated construction environments at different heights on postural stability and fall risk. *Workplace Health & Safety*, 69(1):32–40, 2021. 3

[13] G. Chanel, J. J. Kierkels, M. Soleymani, and T. Pun. Short-term emotion assessment in a recall paradigm. *International Journal of Human-Computer Studies*, 67(8):607–627, 2009. 9

[14] A. Chirico, P. Cipresso, D. B. Yaden, F. Biassoni, G. Riva, and A. Gaggioli. Effectiveness of immersive videos in inducing awe: An experimental study. *Scientific Reports*, 7(1):1218, Apr 2017. 2

[15] A. Chirico, F. Ferrise, L. Cordella, and A. Gaggioli. Designing awe in virtual reality: An experimental study. *Frontiers in Psychology*, 8, 2018. 2

[16] J. A. Coan and J. J. Allen. *Handbook of emotion elicitation and assessment*. Oxford university press, 2007. 1

[17] J. R. Crawford and J. D. Henry. The positive and negative affect schedule (panas): Construct validity, measurement properties and normative data in a large non-clinical sample. *British journal of clinical psychology*, 43(3):245–265, 2004. 2

[18] N. Dozio, F. Marcolin, G. W. Scurati, F. Nonis, L. Ulrich, E. Vezzetti, and F. Ferrise. Development of an affective database made of interactive virtual environments. *Scientific Reports*, 11(1):24108, Dec 2021. 2, 9

[19] N. Dozio, F. Marcolin, G. W. Scurati, L. Ulrich, F. Nonis, E. Vezzetti, G. Marsocci, A. La Rosa, and F. Ferrise. A design methodology for affective virtual reality. *International Journal of Human-Computer Studies*, 162:102791, 2022. 2

[20] P. Ekman. An argument for basic emotions. *Cognition and Emotion*, 6:169–200, 1992. 2

[21] M. FakhrHosseini, M. Jeon, and R. Bose. Estimation of drivers' emotional states based on neuroergonmic equipment: an exploratory study using fnirs. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 38–43, 2015. 1

[22] S. M. Fakhrhosseini and M. Jeon. Affect/emotion induction methods. In *Emotions and affect in human factors and human-computer interaction*, pp. 235–253. Elsevier, 2017. 1

[23] A. Felnhofer, O. D. Kothgassner, M. Schmidt, A.-K. Heinzle, L. Beutl, H. Hlavacs, and I. Kryspin-Exner. Is virtual reality emotionally arousing? investigating five emotion inducing virtual park scenarios. *International journal of human-computer studies*, 82:48–56, 2015. 2

[24] T. Goto, B. Tag, K. Kunze, and T. Dingler. Towards enhancing emotional responses to media using auto-calibrating electric muscle stimulation (ems). In *Proceedings of the 9th Augmented Human International Conference*, pp. 1–2, 2018. 2

[25] J. J. Gross and O. P. John. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology*, 85(2):348, 2003. 1, 2

[26] J. J. Gross and R. W. Levenson. Emotion elicitation using films. *Cognition and Emotion*, 9:87–108, 1 1995. 1

[27] K. Gupta, Y. Zhang, T. S. Gunasekaran, N. Krishna, Y. S. Pai, and M. Billinghurst. Caevr: Biosignals-driven context-aware empathy in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2024. 2

[28] M. Halvey, M. Henderson, S. A. Brewster, G. Wilson, and S. A. Hughes. Augmenting media with thermal stimulation. In *International Conference on Haptic and Audio Interaction Design*, pp. 91–100. Springer, 2012. 2

[29] M. Hassib, M. Pfeiffer, S. Schneegass, M. Rohs, and F. Alt. Emotion actuator: Embodied emotional feedback through electroencephalography and electrical muscle stimulation. In *Proceedings of the 2017 chi conference on human factors in computing systems*, pp. 6133–6146, 2017. 2

[30] J. Healey. 14 physiological sensing of emotion. *The Oxford handbook of affective computing*, p. 204, 2014. 2

[31] N. Howell, L. Devendorf, T. A. Vega Gálvez, R. Tian, and K. Ryokai. Tensions of data-driven reflection: A case study of real-time emotional biosensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018. 2

[32] J. Jun, M. Jung, S.-Y. Kim, and K. Kim. Full-body ownership illusion can change our emotion. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–11, 2018. 2

[33] P. N. Juslin and D. Västfjäll. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences*, 31(5):559–575, 2008. 1

[34] M. E. Kemeny, C. Foltz, J. F. Cavanagh, M. Cullen, J. Giese-Davis, P. Jennings, E. L. Rosenberg, O. Gillath, P. R. Shaver, B. A. Wallace, et al. Contemplative/emotion training reduces negative emotional behavior and promotes prosocial responses. *Emotion*, 12(2):338, 2012. 9

[35] A. C. Kern, W. Ellermeier, and L. Jost. The influence of mood induction by music or a soundscape on presence and emotions in a virtual reality park scenario. In *Proceedings of the 15th International Conference on Audio Mostly*, pp. 233–236, 2020. 2

[36] J. Kim, A. Charbel-Salloum, S. Perry, and S. Palmisano. Effects of display lag on vection and presence in the oculus rift hmd. *Virtual Reality*, 26(2):425–436, Jun 2022. 3

[37] P. Koval, E. K. Kalokerinos, K. H. Greenaway, H. Medland, P. Kuppens, J. B. Nezlek, J. D. Hinton, and J. J. Gross. Emotion regulation in everyday life: Mapping global self-reports to daily processes. *Emotion*, 23(2):357, 2023. 2

[38] P. J. Lang. The emotion probe: Studies of motivation and attention. *American psychologist*, 50(5):372, 1995. 1, 2, 8

[39] J. LeDoux. Rethinking the emotional brain. *Neuron*, 73(4):653–676, 2012. 1

[40] B. J. Li, J. N. Bailenson, A. Pines, W. J. Greenleaf, and L. M. Williams. A public database of immersive vr videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures. *Frontiers in Psychology*, 8, 2017. 1, 2, 3, 6, 8, 9

[41] M. Li, J. Pan, Y. Gao, Y. Shen, F. Luo, J. Dai, A. Hao, and H. Qin. Neurophysiological and subjective analysis of vr emotion induction paradigm. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3832–3842, 2022. 1, 7, 9

[42] A. Lieberoth, S.-Y. Lin, S. Stöckli, H. Han, M. Kowal, R. Gelpi, S. Chrona, T. P. Tran, A. Jeftić, J. Rasmussen, et al. Stress and worry in the 2020 coronavirus pandemic: relationships to trust and compliance with preventive measures across 48 countries in the covidistress global survey. *Royal Society Open Science*, 8:200589, 2021. 2

[43] Y. Ma, Q. Zhou, B. Tag, Z. Sarsenbayeva, J. Knibbe, and J. Goncalves. "hello, fellow villager!": Perceptions and impact of displaying users' locations on weibo. In *IFIP Conference on Human-Computer Interaction*, pp. 511–532. Springer, 2023. 7

[44] C. MacArthur, A. Grinberg, D. Harley, and M. Hancock. You're making me sick: A systematic review of how virtual reality research considers gender & cybersickness. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Comput-

ing Machinery, New York, NY, USA, 2021. 3

[45] F. Marcolin, G. Wally Scurati, L. Ulrich, F. Nonis, E. Vezzetti, N. Dozio, and F. Ferrise. Affective virtual reality: How to design artificial experiences impacting human emotions. *IEEE Computer Graphics and Applications*, 41(6):171–178, 2021. 2

[46] I. Mavridou, E. Balaguer-Ballester, C. Nduka, and E. Seiss. A reliable and robust online validation method for creating a novel 3d affective virtual environment and event library (avel). *PLOS ONE*, 18(4):1–20, 04 2023. 2, 4

[47] N. Morina, H. Ijntema, K. Meyerbröker, and P. M. Emmelkamp. Can virtual reality exposure therapy gains be generalized to real-life? a meta-analysis of studies applying behavioral assessments. *Behaviour Research and Therapy*, 74:18–24, 2015. 9

[48] Y.-W. Park, K.-M. Baek, and T.-J. Nam. The roles of touch during phone conversations: long-distance couples' use of poke in their homes. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1679–1688, 2013. 2

[49] J. P. Pollak, P. Adams, and G. Gay. Pam: a photographic affect meter for frequent, in situ measurement of affect. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 725–734, 2011. 2

[50] G. Riva, R. M. Baños, C. Botella, B. K. Wiederhold, and A. Gaggioli. Positive technology: using interactive technologies to promote positive functioning. *Cyberpsychology, Behavior, and Social Networking*, 15(2):69–77, 2012. 1

[51] J. A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980. 2, 3

[52] Z. Sarsenbayeva, C. Fleming, B. Tag, A. Withana, N. van Berkel, and A. McEwan. A review on mood assessment using smartphones. In J. Abdelnour Nocera, M. Kristín Lárusdóttir, H. Petrie, A. Piccinno, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2023*, pp. 385–413. Springer Nature Switzerland, Cham, 2023. 2

[53] Z. Sarsenbayeva, B. Tag, S. Yan, V. Kostakos, and J. Goncalves. Using video games to regulate emotions. In *Proceedings of the 32nd Australian Conference on Human-Computer Interaction*, OzCHI '20, p. 755–759. Association for Computing Machinery, New York, NY, USA, 2021. 9

[54] L. Sassatelli, M. Winckler, T. Fisichella, and R. Aparicio. User-adaptive editing for 360 degree video streaming with deep reinforcement learning. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, p. 2208–2210. Association for Computing Machinery, New York, NY, USA, 2019. 2

[55] W. Sato, M. Noguchi, and S. Yoshikawa. Emotion elicitation effect of films in a japanese sample. *Social Behavior and Personality: an international journal*, 35(7):863–874, 2007. 1, 8

[56] S. R. Schmidt and C. R. Schmidt. The emotional carryover effect in memory for words. *Memory*, 24(7):916–938, 2016. 9

[57] B. Schöne, J. Kisker, R. S. Sylvester, E. L. Radtke, and T. Gruber. Library for universal virtual reality experiments (luvre): A standardized immersive 3d/360° picture and video database for vr based research. *Current Psychology*, May 2021. 1, 2, 3, 6

[58] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. 5

[59] Y. Shi, P. Koval, V. Kostakos, J. Goncalves, and G. Wadley. "instant happiness": Smartphones as tools for everyday emotion regulation. *International Journal of Human-Computer Studies*, 170:102958, 2023. 2

[60] E. Siedlecka and T. F. Denson. Experimental methods for inducing basic emotions: A qualitative review. *Emotion Review*, 11(1):87–97, 2019. 2, 8

[61] W. Smith, G. Wadley, S. Webber, B. Tag, V. Kostakos, P. Koval, and J. J. Gross. Digital emotion regulation in everyday life. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. 2

[62] R. Somarathna, T. Bednarz, and G. Mohammadi. Virtual reality for emotion elicitation – a review. *IEEE Transactions on Affective Computing*, pp. 1–21, 2022. 9

[63] M. Spering, D. Wagener, and J. Funke. The role of emotions in complex problem solving. *Cognition and Emotion*, 19(8):1252–1261, 2005. 1

[64] H. Strasburger, I. Rentschler, and M. Jüttner. Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5):13–13, 12 2011. 7

[65] B. Tag, J. Goncalves, S. Webber, P. Koval, and V. Kostakos. A retrospective and a look forward: Lessons learned from researching emotions in-the-wild. *IEEE Pervasive Computing*, 2021. 1, 2

[66] B. Tag, Z. Sarsenbayeva, A. L. Cox, G. Wadley, J. Goncalves, and V. Kostakos. Emotion trajectories in smartphone use: Towards recognizing emotion regulation in-the-wild. *International Journal of Human-Computer Studies*, 166:102872, 10 2022. 2

[67] B. Tag, N. van Berkel, A. W. Vargo, Z. Sarsenbayeva, T. Colasante, G. Wadley, S. Webber, W. Smith, P. Koval, T. Hollenstein, J. Goncalves, and V. Kostakos. Impact of the global pandemic upon young people's use of technology for emotion regulation. *Computers in Human Behavior Reports*, 6:100192, 5 2022. 2

[68] M. M. Terwogt and J. B. Hoeksma. Colors and emotions: Preferences and combinations. *The Journal of General Psychology*, 122(1):5–17, 1995. 8

[69] D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063, 1988. 2

[70] D. Watson, D. Wiese, J. Vaidya, and A. Tellegen. The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of personality and social psychology*, 76(5):820, 1999. 2

[71] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945. 5

[72] G. Wilson, D. Dobrev, and S. A. Brewster. Hot under the collar: mapping thermal feedback to dimensional models of emotion. In *proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 4838–4849, 2016. 2

[73] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 143–146. Association for Computing Machinery, New York, NY, USA, 2011. 5

[74] K. Yang, B. Tag, Y. Gu, C. Wang, T. Dingler, G. Wadley, and J. Goncalves. Mobile emotion recognition via multiple physiological signals using convolution-augmented transformer. In *Proceedings of the 2022 International Conference on Multimedia Retrieval*, ICMR '22, p. 562–570. Association for Computing Machinery, New York, NY, USA, 2022. 2

[75] K. Yang, B. Tag, C. Wang, Y. Gu, Z. Sarsenbayeva, T. Dingler, G. Wadley, and J. Goncalves. Survey on emotion sensing using mobile devices. *IEEE Transactions on Affective Computing*, 2022. 2

[76] K. Yang, C. Wang, Y. Gu, Z. Sarsenbayeva, B. Tag, T. Dingler, G. Wadley, and J. Goncalves. Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition. *IEEE Transactions on Affective Computing*, 2021. 1