Designing Immersive Affective Tasks for Emotion Elicitation in Virtual Reality

Jason W. Woodworth
Christoph W. Borst
jason.woodworth1@louisiana.edu
cwborst@gmail.com
University of Louisiana at Lafayette
Lafayette, Louisiana, USA

ABSTRACT

The growing ubiquity of virtual reality presents an opportunity to explore uniquely active emotion elicitation stimuli with the potential to elicit specific emotions more strongly when compared to the traditional passive stimuli such as film, images, or sound. We discuss the design and initial evaluation of four affective tasks in virtual reality that take advantage of the medium's interactive nature to elicit frustration, confusion, boredom, and pleasure. We report results suggesting their ability to elicit these emotions effectively, as well as participant feedback regarding what task elements contributed most to their felt emotions. Finally, we discuss the potential for future works to compare the strengths of emotion elicitation through passive and active stimuli, and our ultimate goal of understanding student emotions in a VR classroom.

ACM Reference Format:
Jason W. Woodworth and Christoph W. Borst. 2021. Designing Immersive Affective Tasks for Emotion Elicitation in Virtual Reality. In *2nd Momentary Emotion Elicitation and Capture Workshop, May 09, 2021.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Lab-based emotion elicitation is traditionally done using passive stimulus such as films [7], images [13], or sounds [5]. With a passive stimulus, the user is typically given no task other than to absorb the stimulus and be aware of their emotions. This paradigm comes with a number of advantages. Passive stimuli are typically thought to be well-standardized (all users experience the same thing and thus the same emotion), they are typically easy to administer, and can be performed using standard software.

The use of active stimuli is also becoming more popular. This has been done more commonly in the driving context for works that wish to detect negative states during driving [10, 12], but has recently been expanded to other computerized tasks, such as simple games [3, 17]. The use of active stimuli for collecting physiological data meant to help build a model to recognize affect may benefit from the elicitation source being closer to the final task. They also potentially eliminate the need for the cultural or linguistic understanding needed for films that present a narrative story, possibly making them more universally applicable. Thus we consider it potentially useful to further investigate their effects and benefits.

At the same time, virtual reality is becoming more widely known as an immersive medium able to emotionally engage the user through the user's ability to interact more directly with the virtual world [18]. The use of VR as an elicitation tool has been gaining popularity as well, though primarily still in the context of passive stimuli. For example, viewing a 360 degree video [14, 21] or viewing moving or static environments [15, 16], which we consider immersive analogues of standard films and images. While these works achieve noteworthy results and in some instances are able to elicit emotions difficult to obtain in non-VR environments [1], they do not use the medium to the full potential we believe it to have. While passive stimuli often must rely on characters or narratives presenting an emotion that the viewer must empathize with [14], VR presents an opportunity to have the the participant experience the emotion more directly through their interaction with the environment.

We have developed and begun testing the use of affective tasks for eliciting emotion with virtual reality. Our goal is to detect negative states in students using educational VR applications by building a model from physiological data mapped to emotions we expect to see in educational settings. To that end, our tasks target frustration, confusion, boredom, and pleasure. We discuss in this paper their design and challenges behind it, results of an initial feedback study demonstrating their ability to elicit the targeted emotions, and plans for future experiments to compare the effectiveness of active and passive stimuli in and out of virtual reality.

2 TASK DESIGN

2.1 Task Descriptions

We describe all tasks by the named emotion they intend to elicit and their desired location on the Russell circumplex model [19]. This commonly-used model proposes that affective states arise from two neurophysiological systems, one related to valence (the positive/negative continuum), and the other to arousal (the calm/excited continuum). This creates four quadrants in which all emotions are described as having low/high valence and low/high arousal.

All tasks are performed within a similar environment: an open space with a large black floor and a skybox with colors picked to match the intended emotion (similar to a strategy done by Liao et al. [15]). All tasks have a score, with correct interactions gaining

MEEC '21, May 09, 2021, Virtual

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

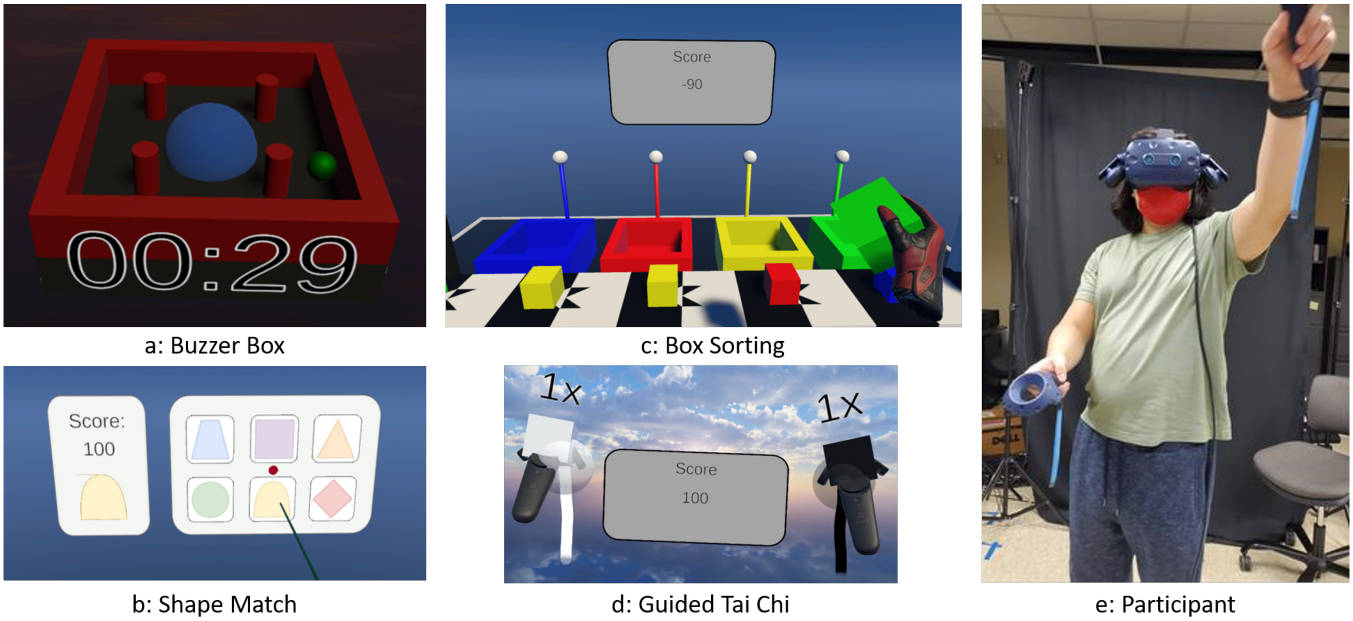


Figure 1: Affective Tasks. Buzzer Box asks the user to keep the green ball in the blue dome center while an invisible force pushes it away. Shape Match asks the user to select the shape shown on the left from the selection on the right. Box Sorting asks the user to sort the box into the correctly colored bin while occasionally giving incorrect feedback for correct sorts. Guided Tai Chi asks the user to follow the colored spheres with their tracked hands while providing positive feedback. E shows a participant performing Guided Tai Chi.

points to encourage engagement with the task. All users are told that there is a point threshold that must be reached to "win", though tasks are always ended after a certain amount of time so as to avoid unwanted frustration due to poor performance. Tasks are pictured in Figure 1.

Frustration: Buzzer Box. Based loosely on an analogous 2D task [17], this task aims to elicit frustration by presenting a deceptively simple task with a difficult goal and high amounts of negative feedback. Frustration is considered to have high arousal with moderately low valence. We consider that difficult tasks [3] and frequent unpleasant noises [8] elicit stress, a similar emotion, and thus incorporate those into our task design. We hypothesize that frustration can be felt more strongly through active stimulus than passive as the frustrating stimulus is being experienced directly by the user instead of indirectly.

In Buzzer Box, the user is presented with a box containing four pegs and a ball that follows their tracked hand's position and rotation. At the start of the task, they are told that they must use their hand to balance the box and keep the ball within the marked center of the box in order to score points, and that their score will decrease when the ball is outside of that spot. Once the task starts, an invisible force that the user is not told about pushes the ball from the center and the user must quickly rotate their hand to counteract the force. When the ball hits a peg or side of the box, a loud buzzer is played. Due to the difficulty of the task this is heard frequently.

Confusion: Box Sorting. This task aims to elicit confusion by presenting a simple task and occasionally breaking its rules without informing the user. Confusion is among the lesser-studied emotions, and while it was not always recognized as one [11], it is recently been deemed important to consider, especially in the education context [2, 6] where it is important to recognize and address for optimal educational results. Confusion is generally considered to have moderately high arousal and low valence, thus placing it in a different position on the same low valence / high arousal quadrant as frustration.

In Box Sorting, the user is stationed in front of a conveyor belt and four colored bins. Once the task starts, boxes colored to match one of the bins appear on the conveyor belt at a rate of one per second. The user is asked to physically sort the boxes (grabbing by reaching out and squeezing the trigger, throwing by releasing the trigger during a throw motion) into the appropriate bin. If the user performs an incorrect sort, a loud buzzer noise is played. Once every five to ten correct sorts, the rule is broken and it is treated as an incorrect sort, thus confusing the user.

Boredom: Shape Match. This task aims to elicit boredom by presenting a simple task for a long period of time. Boredom is considered to have low arousal and valence. We consider a boring task to be one which is easy but requires frequent interaction [3].

In Shape Match, the user is shown a shape on a screen in front of them and must select the matching shape from a selection on a separate screen. Upon a correct match, points are awarded and the user is given a two second break before the next target shape is shown.

Upon an incorrect match, points are deducted and the loud buzzer is played. Standard wand selection techniques are used to avoid the interaction method causing any unnecessary entertainment. The Box Sorting activity was originally planned as the boring task, but upon initial pilots it was found that the act of throwing the boxes alone made the task moderately enjoyable to pilot participants.

Pleasure: Guided Tai Chi. This task aims to elicit pleasure by mimicking a low-intensity analogue activity (basic tai chi movements) typically done for relaxation and pleasure [20] combined with game design elements meant to enhance satisfaction. Pleasure is typically considered to have a high valence and neutral arousal.

In Guided Tai Chi, the user is asked to move their hands along a pre-recorded path illustrated by guiding spheres with a leading trail telling them where to move. Positive game design elements include motivational background music, a score multiplier that increases when the user keeps their hands within the guiding spheres, and basic animations and positive sound effects for performing well.

Neutral. A neutral room was designed to be placed in between tasks so that users could recover from the emotion elicited from the previous task, as is commonly recommended for films [7]. The room contains a sky-blue skybox with basic rain sounds and a timer showing time remaining until the next task.

2.2 Design Challenges

One of the primary challenges for affective tasks is the need for standardization between participants. While all participants viewing a passive presentation (e.g. movie) may be expected to receive the stimulus in the same way [7], this may be less true for an active stimulus. For example, while viewers of a movie may differ in their internal interpretation of the film, they will outwardly experience it from similar viewing angles, room conditions, etc..., as there are no ways to interact with the stimulus. Participants using an active stimulus such as driving or playing a game may engage with the tasks in ways that give them different experiences, and thus potentially elicit different emotions. In our attempt to standardize tasks, we made them as simple as possible, removing all possible manners of interaction aside from those deemed essential to the desired emotion.

Confusion was the most difficult emotion to design a task for. Initial planning considered a lecture on a difficult topic or a lecture with occasional incorrect information with a quiz at the end, similar to other works eliciting confusion by presenting a passage with incongruent information [9]. While this would possibly create a stimulus more similar to the feeling of confusion encountered in an actual classroom, it presented several problems. If the user happened to already be familiar with the material, they could experience no confusion at all, while a user with no familiarity with even the base concepts could disengage and experience boredom. In order to get across enough material to present a meaningful quiz, the task would have to last a larger amount of time that might result in obfuscating the moments of confusion in physiological data. For these reasons we pivoted to a simpler task with a more universal confusing stimulus.

Table 1: Summary of self-assessment results. Emotion refers to the label selected by the most participants. Strength of the elicited emotion (out of 7) and valence and arousal (out of 9) are given as averages.

Task	Emotion	Strength	Valence	Arousal
Buzzer Box	Frustrated	4.2	3	6
Shape Match	Bored	5.2	4.5	3.5
Guided Tai Chi	Pleased	4.4	6.5	6.5
Box Sorting	Frustrated	5.2	3.5	6

3 USER EVALUATION

We conducted an informal user evaluation with the intent of verifying if tasks elicited their targeted emotions. Results were obtained with end-of-task self-assessments and discussion questionnaires given after all tasks were completed. Five participants[1] were recruited for the study.

After an initial discussion of the study, participants donned an HTC Vive Pro Eye headset and Empatica E4 wristband. Heart rate, electrodermal activity (EDA), eye openness, and pupil diameter were recorded throughout the experience for future analysis. Participants were told that they would be asked to complete a series of tasks meant to elicit certain emotions and an assessment of how they felt during the task right after. Participants remained standing throughout the entire study, lasting approximately 15 minutes. The order of the tasks was randomized.

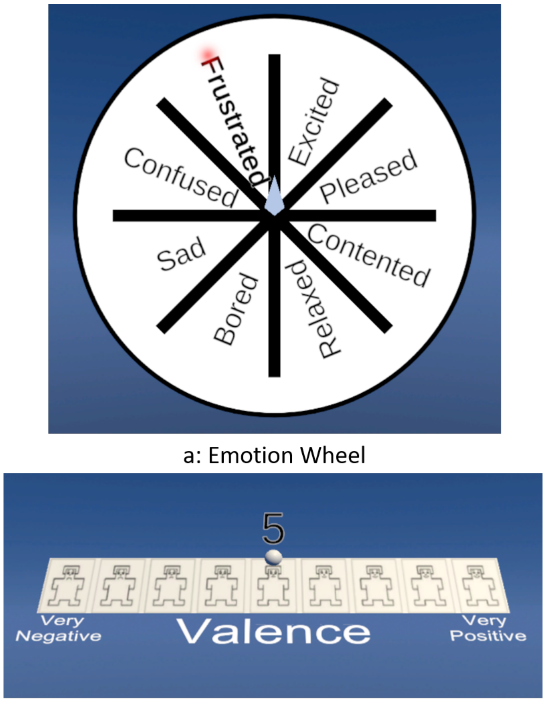
Each task lasted a set amount of time based on preliminary testing. Buzzer box lasted 90 seconds, as testing indicated that the difficult task would get more frustrating over time, but a user may give up if it lasts too long. Box Sorting lasted 60 seconds, so that the broken rule would happen a small number of times and the participant would not catch on to the rule breaking. Shape Match lasted 180 seconds, as testing showed that the simple task would get more boring over time. Guided Tai Chi lasted 120 seconds, as testing suggested this gave users enough time to enjoy the experience before the novelty wore off. Participants rested in the neutral room for 30 seconds between tasks. After the final task was complete, participants were interviewed about their responses.

After each task, users completed self-assessments using a simple in-world interface picture in Figure 2. To obtain a descriptive emotion label, a wheel showing 8 emotions (frustrated, confused, sad, bored, excited, pleased, contented, relaxed) was displayed. Two emotions were picked from each quadrant of the circumplex model so as to display a variety of options while keeping the interface simple enough to meet VR constraints. Four were picked from those we desired to elicit, and four were picked from other common low-number emotion models. Once a label is selected, users are asked to rate how strongly that emotion was felt on a scale from 1 to 7. Finally, they rated their valence and arousal on a SAM-like scale [4] from 1 to 9.

3.1 Results

Three of the four tasks primarily elicited the targeted emotion, with all falling into their expected quadrants in the circumplex model.

[1]Participant numbers are low in part because of COVID-19 related restrictions



a: Emotion Wheel

b: SAM Slider

Figure 2: Interface used for self-assessment after each task.

Results are summarized in Table 1. An initial statistical analysis on physiological features supports a difference between tasks for eye-related features.

3.1.1 Self-Assessment and Feedback. For Buzzer Box, three participants indicated they were frustrated while two selected excited, with the average strength of the elicited emotion at 4.2 out of 7. Average valence-arousal ratings were 3 and 6 out of 9, respectively. During freeform feedback, those who were frustrated explained that it was more difficult than they expected and that the buzzer noise itself was frustrating, stating that they stopped caring about points and only wanted to avoid the buzzer. Those who were excited stated that the invisible force on the ball made it fun, and once they were able to keep the ball in the right spot they felt like the achieved a difficult goal.

For Shape Match, four participants selected that they were bored while one selected they were frustrated, with an average strength of 5.2. Average valence-arousal ratings were 4.5 and 3.5. Those who were bored stated it was because the task was easy, repetitive, and long, with the score goal providing incentive to continue despite boredom. The participant who selected frustrated stated the same reasons, plus distraction from uncontrolled noises outside of the lab environment.

For Guided Tai Chi, three participants selected pleased while two selected contented, with an average strength of 4.4. Average valence-arousal ratings were 6.5 and 6.5. All participants noted that the vibration was relaxing, the task was simple yet fun, and that they enjoyed the novelty of the experience, with no apparent distinction in the feedback from the two emotion-label groups. Three noted that the experienced pleasure was hampered by an inability to see the targets for both hands at one time, a unique challenged posed by the nature of VR.

Table 2: Mean pupil size (in mm) and blink rate (in blinks per minute) reported per task.

Task	Pupil Size	Blink Rate
Buzzer Box	4.49	12.8
Shape Match	2.78	24.3
Guided Tai Chi	3.71	27.5
Box Sorting	3.49	43.6

Box Sorting was the least agreed-upon, with three selecting frustrated, one selecting excited, and one selecting confused, with an average strength of 5.2. Average valence-arousal ratings were 3.5 and 6. Those who were frustrated stated that they were confused for the first two to three forced rule breaks, but then realized the game was "tricking" them and felt that it was unfair. The participant who selected excited stated they enjoyed just being able to throw the cubes, and the negative feedback did not bother them much. The one who selected confused cited the forced rule breaks, but also the fact that, due to a function of the development toolkit used, each box was highlighted in a light blue color when it became grabbable, thus momentarily confusing them on which bin to sort it into.

All participants stated that the neutral area allowed enough time and appropriate environment to transition out of the emotion elicited by the previous task.

3.1.2 Initial Physiological Analysis. Heart rate, EDA, and pupil diameter (per eye) were recorded throughout each task. Averages were taken per user across the middle 50% of each task to remove initial physiology changes and any rolloff from users getting used to the task. Blink rate was extracted (per eye) from eye openness values by counting any blink that passed a certain amplitude threshold and dividing by the length of the task. Values for eye-related features were not found to differ substantially between eyes, thus both eyes were averaged together for analysis.

Repeated measures ANOVA tests were performed on averages and blink rates. A significant difference between tasks was detected for all eye-related features: $F(4, 12) = 14.522, p < 0.001$ for pupil size and $F(4, 12) = 9.334, p = 0.001$ for blink rate, indicating good potential for its use as a feature for differentiating emotional states. Followup pairwise tests on pupil size indicate significant differences ($p < 0.05$) between all tasks except Box Sorting and Tai Chi. Followup tests on blink rate indicate significant differences between all tasks except Shape Match and Tai Chi. No significant difference between tasks was detected for heart rate and EDA: $F(3, 12) = 2.642, p = 0.097$ for heart rate and $F(1.08, 4.32) = 0.353, p = 0.598$ for EDA. A low participant count gives minimal statistical power, and we are investigating sensor reliability.

Means for eye-related features can be seen in Table 2. Initial impressions suggest a possible correlation between pupil size and arousal value, as Shape Match produces the lowest diameter. A complex relationship is suggested between blink rate and emotion; its highest rate is during Box Sorting, intended to create confusion and cognitive load, while its lowest rate is during Buzzer Box, intended to just create frustration, though both tasks were primarily reported as being frustrating. This relationship will be investigated more thoroughly in future works.

4 DISCUSSION AND FUTURE WORKS

Results, though interpreted lightly, suggest the validity of using affective tasks to elicit emotion. All tasks except for Box Sorting successfully elicited their targeted emotions, with each receiving useful feedback that participants believed would enhance the desired effects of each task. We intend to use this feedback to improve the tasks and conduct a larger study comparing them to passive stimuli. For example, based on feedback we believe Box Sorting will be made more confusing by reducing the time, including fewer rule breaks, and introducing more ambiguous rules, such as boxes that are colored to not sort correctly into any bin. Buzzer box will be made more frustrating by increasing the invisible force on the ball, thus making it more difficult and less enjoyable. Guided Tai Chi will be made more enjoyable by creating a new flow that allows the user to look at both hands at all times.

We also consider the possibility that by creating tasks involving movement, there may be a conflation of physiological reactions elicited by movement alone and by other factors. For example, for the Guided Tai Chi task, it is unknown how much of the physiological reaction is caused by the necessary movements vs. the rumble effect, score multipliers, and other effects put in to help elicit its targeted emotion. Future studies could consider task variations that replicate the same physical motions but change or omit emotive backgrounds, scores, or other aspects. This would allow us to identify more clearly what components of each task elicit certain physiological reactions and to better understand physiological effects of motion.

In a future study, we intend to run a two-phase experiment in which one phase sees participants view film clips from the American Emotional Film Library [7] and 360 degree videos from [14] in VR, and the other phase sees participants completing either our affective tasks or their movement-only variants. Comparisons can then be made on the strength of their elicitation and, perhaps more importantly, the ability to distinguish between SAM ratings and physiological readings for the emotions they elicit.

Once they have been verified, we intend to use our affective tasks, along with other research into the physiology of attention and distraction, to build models to recognize inattentive or otherwise negatively affected students in VR classrooms. These models would be used to notify teachers of students who may need extra help in the classroom. We believe this would address a critical gap in the educational VR space in which users are less able to use and read non-verbal communication due to the use of avatars with limited expression.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1815976 and by the Louisiana Board of Regents Support Fund under contract LEQSF(2019-20)-ENH-DE-22.

REFERENCES

[1] Mohammed Ali Anwary and Siphumelele Sigwebela. 2018. *Emotion Elicitation in the Laboratory: Virtual reality environments can make you feel fear.* Ph.D. Dissertation. University of Cape Town.

[2] Amaël Arguel, Lori Lockyer, Gregor Kennedy, Jason M. Lodge, and Mariya Pachman. 2019. Seeking optimal confusion: a review on epistemic emotion management in interactive digital learning environments. *Interactive Learning Environments* 27, 2 (2019), 200–210. https://doi.org/10.1080/10494820.2018.1457544

[3] Fernando Bevilacqua, Henrik Engström, and Per Backlund. 2019. Game-calibrated and user-tailored remote detection of stress and boredom in games. *Sensors (Switzerland)* 19, 13 (2019), 1–43. https://doi.org/10.3390/s19132877

[4] 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, I (1994), 49–59.

[5] Margaret M Bradley and Peter J Lang. 1999. International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings. *Technical Report B-2, The Center for Research in Psychophysiology, University of Florida, Gainesville, FL* (1999).

[6] Maher Chaouachi, Imène Jraidi, Susanne P. Lajoie, and Claude Frasson. 2019. Enhancing the learning experience using real-time cognitive evaluation. *International Journal of Information and Education Technology* 9, 10 (2019), 678–688. https://doi.org/10.18178/ijiet.2019.9.10.1287

[7] James A Coan and John JB Allen. 2007. *Handbook of emotion elicitation and assessment.* Oxford university press.

[8] S. Cohen, G.W. Evans, D. Stokols, and D.S. Krantz. 2013. *Behavior, Health, and Environmental Stress.* Springer US. https://books.google.com/books?id=WFNDBAAAQBAJ

[9] Francis T. Durso, Kaitlin M. Geldbach, and Paul Corballis. 2012. Detecting Confusion Using Facial Electromyography. *Human Factors* 54, 1 (2012), 60–69. https://doi.org/10.1177/0018720811428450

[10] Jennifer A. Healey and Rosalind W. Picard. 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems* 6, 2 (2005), 156–166. https://doi.org/10.1109/TITS.2005.848368

[11] Ursula Hess. 2003. Now you see it, now you don't–the confusing case of confusion as an emotion: Commentary on Rozin and Cohen (2003). *Emotion* 3 (2003). Issue 1.

[12] Charlotte Jacobé de Naurois, Christophe Bourdin, Anca Stratulat, Emmanuelle Diaz, and Jean Louis Vercher. 2019. Detection and prediction of driver drowsiness using artificial neural network models. *Accident Analysis and Prevention* 126, December 2017 (2019), 95–104. https://doi.org/10.1016/j.aap.2017.11.038

[13] Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. 1997. International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention* 1 (1997), 39–58.

[14] Benjamin J Li, Jeremy N Bailenson, Adam Pines, Walter J Greenleaf, and Leanne M Williams. 2017. 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), 2116.

[15] Dan Liao, Lin Shu, Guodong Liang, Yingxuan Li, Yue Zhang, Wenzhuo Zhang, and Xiangming Xu. 2020. Design and Evaluation of Affective Virtual Reality System Based on Multimodal Physiological Signals and Self-Assessment Manikin. *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology* 4, 3 (2020), 216–224. https://doi.org/10.1109/JERM.2019.2948767

[16] Javier Marín-Morales, Juan Luis Higuera-Trujillo, Alberto Greco, Jaime Guixeres, Carmen Llinares, Enzo Pasquale Scilingo, Mariano Alcañiz, and Gaetano Valenza. 2018. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific Reports* 8, 1 (2018), 1–15. https://doi.org/10.1038/s41598-018-32063-4

[17] Daniel J. McDuff, Javier Hernandez, Sarah Gontarek, and Rosalind W. Picard. 2016. COGCAM: Contact-free measurement of cognitive stress during computer tasks with a digital camera. *Conference on Human Factors in Computing Systems - Proceedings* (2016), 4000–4004. https://doi.org/10.1145/2858036.2858247

[18] Radiah Rivu. 2020. Understanding Emotions in Virtual Reality. In *Momentary Emotion Elicitation and Capture.*

[19] James A. Russell. 1978. Evidence of convergent validity on the dimensions of affect. *Journal of Personality and Social Psychology* 36, 10 (1978), 1152–1168. https://doi.org/10.1037/0022-3514.36.10.1152

[20] Jing Sun and Nicholas Buys. 2016. Health benefits of Tai Chi. *Canadian Family Physician* 62, 11 (2016), 881–890.

[21] Tong Xue, Surjya Ghosh, Gangyi Ding, Abdallah El Ali, and Pablo Cesar. 2020. Designing real-time, continuous emotion annotation techniques for 360° VR videos. In *Momentary Emotion Elicitation and Capture.* 1–4. https://doi.org/10.1145/3334480.3382895