Design and Evaluation of Affective Virtual Reality System Based on Multimodal Physiological Signals and Self-Assessment Manikin

Dan Liao, Lin Shu, *Member, IEEE*, Guodong Liang, Yingxuan Li, Yue Zhang, Wenzhuo Zhang, and Xiangming Xu, *Member, IEEE*

***Abstract*—Affective materials have always been an important factor in improving the efficiency of affect related research for their ecological validity. In order to achieve high ecological validity in emotion elicitation, this paper proposed a novel Affective Virtual Reality System (AVRS) by utilizing the immersion, privacy and design flexibility of virtual reality. Design elements including subject features, sound features, motion features and color features were extracted by referring to the referenced emotion materials, art works and the existing VR scenes. Affective VR scenes were then designed by Unreal Engine 4.12 and their effectiveness was validated by Self-Assessment Manikin (SAM). Furthermore, arousal was used as an indicator to compare the difference of ecological validity between VR and video emotional materials with an intergroup experiment through Electroencephalography (EEG), Heart Rate (HR), Galvanic Skin Response (GSR), and SAM. Results proved that VR scenes could achieve the same emotion elicitation as video. Especially, there was significant difference in measures of Fearful in SAM evaluation, indicating that VR emotion materials were expected to deliver a good effect on negative emotional scenes. The proposed AVRS as well as the multimodal physiological signal database with valence, arousal and dominance (VAD) labels could help the development of emotion related studies.**

***Index Terms*—Virtual Reality (VR) Scenes design, Emotion elicitation, Arousal Evaluation, Multimodal Physiological Signals.**

Manuscript received July 31, 2019; revised September 21, 2019; accepted October 10, 2019. Date of publication October 21, 2019; date of current version August 21, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant U1801262, in part by the Natural Science Foundation of Guangdong Province of China under Grant 2018A030310407, in part by the Science and Technology Program of Guangzhou under Grant 201704020043, in part by the Fundamental Research Funds for the Central Universities under Grants 2018ZD11 and 2019MS028, in part by the Major Science and Technology Projects in Guangdong Province under Grant 2016B010108008, and in part by the Project of Research and Reform in Guangdong Provincial Education Department under Grant x2sjN5190650. This article was presented in part at the IEEE MTT-S 2019 International Microwave Biomedical Conference, Nanjing, China, May 2019, and International Conference on Virtual Reality and Visualization, Zhengzhou, China, October 2017. *(Corresponding authors: Lin Shu; Xiangmin Xu.)*

D. Liao is with the School of Electronic and Information Engineering and School of Design, South China University of Technology, Guangzhou 510641, China (e-mail: danl@scut.edu.cn).

L. Shu, G. Liang, Y. Li, Y. Zhang, W. Zhang, and X. Xu are with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: shul@scut.edu.cn; lamontagne.liang@gmail.com; 1003040356@qq.com; 475794822@qq.com; 974857377@qq.com; xmxu@scut.edu.cn).

Digital Object Identifier 10.1109/JERM.2019.2948767

I. Introduction

EMOTION elicitation [1] was crucial for emotion research in many areas which was for the study of psychology and psychiatry for the analysis of psychological and physiological indicators followed by certain emotions [2]. The main methods of eliciting emotion could be classified by different sensory channels of human beings as follows: visual induction, auditory induction, olfactory induction and multichannel induction.

Visual induction included emotional words and images which was more convenient but with less ecological validity comparing to other methods [3]. A variety of emotion elicitation datasets had established in many research, such as Affective Norms for English Words (ANEW) [4], International Affective Picture System (IAPS) [5], Pictures of Facial Affect (POFA) [6], and Chinese Affective Picture System (CAPS) [7].

Auditory induction included natural sound recordings, nonverbal syllables, and musical materials, which were more intensive and durable but less standardized. The representative datasets were International Affective Digital Sounds (IADS) [8] and Chinese Affective Digital Sounds (CADS) [9].

As for olfactory induction, participants were usually required to smell sort of odor intentionally or unconsciously to achieve the emotion elicitation. Like other emotional sensory channels, olfactory stimuli could elicit positive or negative emotions of the participants, thus affecting the individual's perceptions and behaviors. But it was also short of effective and standardized audio materials [10].

Multichannel induction was the fusion of visual and auditory which could reach higher ecological validity by its richer content and almost realistic plot comparing to pure visual and auditory stimulation. However, there were not many available materials since it needed to follow the plot of standardized film materials. There were some two-dimensional video materials, such as American Emotional Film Library [11], French Emotional Film Library [12] and Chinese Affective Video System (CAVS) [13].

In addition, there existed posture induction [14] based on embodied cognitive theory, that was, the participants' emotions were elicited by controlling their body posture. But this method required participants to perform behavior, thus leading individual difference [15]. Imagination induction elicited emotions by guiding participants to imagine or recall certain situations.

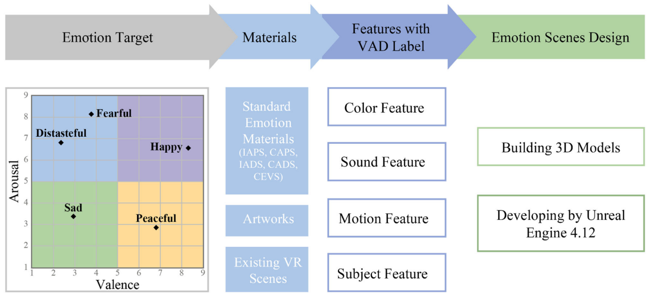

Fig. 1. The design framework of VR scenes.

But this method required the participants to concentrate and cooperate actively.

All the emotion elicitation methods mentioned above can successfully elicit certain emotions, but each of them had its own shortcomings including low emotion elicitation efficiency, high susceptibility to ambient interference, significant individual variance, and low ecological validity [16]. To address these problems, virtual reality was considered, which could bring an undisturbed and immersive experience to participants while having certain advantages in solving some of the problems mentioned above by designing dynamic 3D models and applying a 3D display with sensor technology to achieve multi-perception, multi-interaction, high-immersion, and high-privacy emotion elicitation. Therefore, a new eliciting system named The Affective Virtual Reality System (AVRS) was established in this paper to improve the ecological validity.

This paper was organized as follows. Section II introduced the methods of VR scenes design, SAM evaluation, and arousal evaluation. In Section III, the results of VR scenes evaluation and arousing level measurement were reported. The discussion and conclusion were described in Sections IV and V.

II. Methodology

A. *Affective VR System Design and SAM Validation*

1) Design Framework: The design framework of VR scenes [17] was illustrated by Fig. 1. Five representative basic targeted emotions (Happy, Peaceful, Distasteful, Fearful, and Sad) based on the three-dimensional theory of emotion and discrete theories of emotion were selected [18], [19]. Then design elements were extracted including subject features, sound features, motion features and color features by referring to the referenced emotion materials (IAPS, IADS, CADS, CAVS), art works and the existing VR scenes. The 3D models were build using Unreal Engine 4.12, and then the affective VR scenes were completed.

2) Feature Extraction: Feature extraction included four main parts: subjects feature, sound feature, motion feature, and color feature. According to the IAPS, CAPS and CAVS, the subject features were extracted with the corresponding emotion labels. According to the digital audio information in the IADS and CADS, the sound features under the corresponding emotion tags could be obtained. According to motion information in the CAVS, the motion features of objects and perspective were extracted. As to the artwork and the existing VR scenes, the art

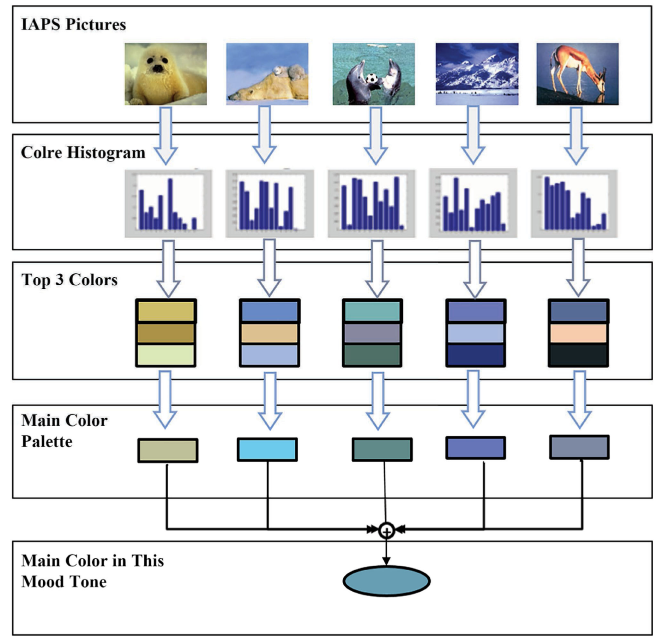

Fig. 2. The process of color feature extraction.

elements, such as light, space and stories were also extracted. Those three features were extracted by the design semiotics theory, which was used to establish an overall understanding of the design through a deep comprehension of symbolic forms and meanings. That meant to make a descriptive summary of the form, meaning and attribute of the targeted VAD numerical database to deepen the understanding of the characteristics of the materials, and also to guide the completion of the design at last [20]. Especially, since color occupied a vital position in emotion elicitation, the main color extraction module was designed of which the extraction steps were given in Fig. 2 to guide VR scenes design. The octree algorithm [21]–[23] was used to extract the three theme colors of each picture of every emotion. Then based on the RGB 3D coordinates, the proportion of RGB color channel of the three theme colors was calculated so as to merge these three colors into one color, after which the theme color of each picture was extracted. The final theme color of each emotion was got by merging every theme color of the pictures based on the proportion of RGB color channel of these colors. Based on these affective features, VR 3D models, panoramic scenes and audios of target emotion could be built.

3) Creation Process: The target emotional VR scenes were created as following steps: 1) writing an 80-secend script for each target emotion, creating materials and scene scenarios that were needed by 3D Max, 3) building a static scene which combined scenes, subjects and sound elements, 4) writing C# scripts to implement action events. The process of creating the AVRS were shown in Fig. 3.

4) SAM Validation: The evaluation model Self-Assessment Manikin (SAM) [24] shown in Fig. 4 was used to measure the scoring to obtain the ratings of valence, arousal, and dominance. Each rating range was from 1 to 9. The higher the rating was, the

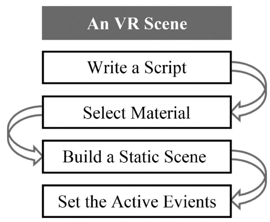

Fig. 3. The process of creating the AVRS.

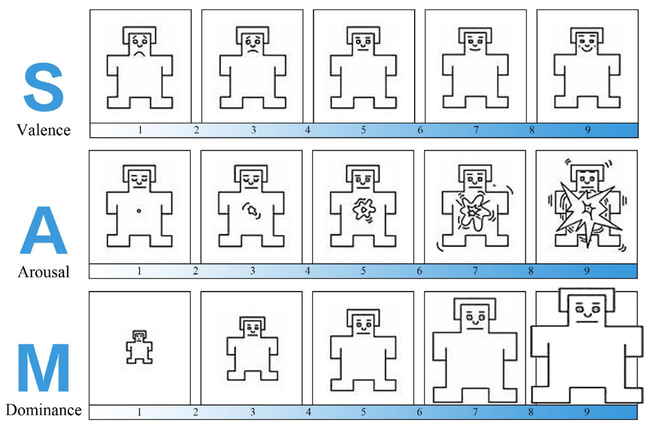

Fig. 4. Self-assessment manikin (SAM).

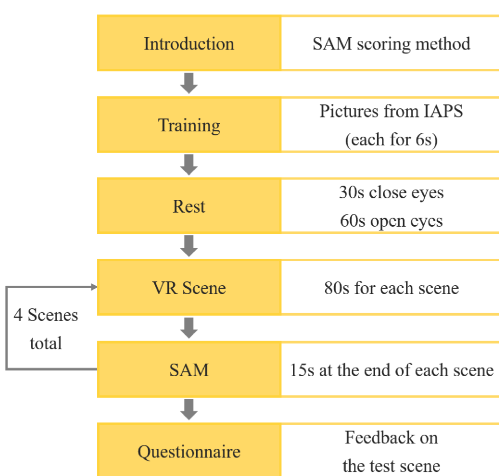

Fig. 5. The process of VR scene SAM assessment.

more intense the emotion would be (high pleasure, high arousal and high dominance).

The selected participants were balanced on the gender as follows: the ratio of male to female was 1:1, and the age was 23 $\pm$ 2.9; 30 participants (15 males and 15 females) participated in this assessment. According to the medical examination and mental health test reports, we chose the individuals without a prior history of cardiovascular disease, and whose single scores of SCL-90 test was greater than 3 points.

The process of validation was shown in Fig. 5. SAM scoring method was introduced to the participants before the participants

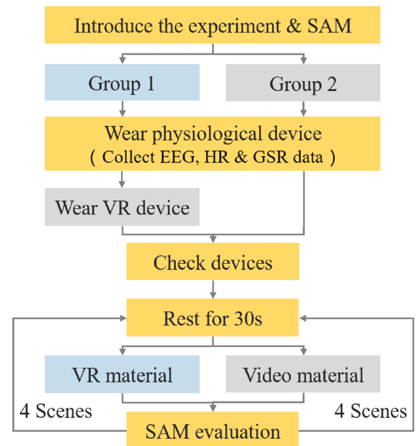

Fig. 6. The experimental procedure of arousing level measurement.

wore VR device. To begin with, participants had been trained for SAM rating with IAPS (each picture showed for 6 seconds), then they were required to close their eyes for 30 seconds to have a rest and to blink 10 seconds for adapting bright environment. After that VR scenes were played (each scene lasted for 80 seconds) and the SAM was performed for 15 seconds at the end of each scene. In particular, the assessment was accomplished through handle interaction. In the end, questionnaire was conducted.

B. Arousal Evaluation Between VR and Video Scenes by Multimodal Physiological Signals and SAM

Intensity, duration and discreteness are considered to be three indicators for evaluating emotional ecological validity [25]. Considering that physiological indicators of duration and discreteness were not clear enough in current research and the same emotional material content was used in two groups in this study, we assumed that arousal could be used as a reference index to measure intensity based on different media for comparative experiments. At the same time, emotional arousal was also the key point to induce physiological arousal and to obtain changes in physiological indicators. Based on the above reasons, we mainly focused on the arousal of the participants in this experiment.

1) Experimental Procedures: All procedures have been approved by the Research Ethics Committee of the Body Data Science Engineering Center of Guangdong Province in China (BDS18-03). Informed consent was signed before the experiment began.

This study used an intergroup experimental design. Participants were randomly divided into two groups. Experimental group contained 20 participants watching VR materials, while control group contained 20 participants watching video materials. In order to avoid the sequence effect, 4 video materials were played in a random sequence. Emotional counseling was given by a counselor for psychologically unwell responders. The experiment lasted for about 30 minutes of which procedure was as shown in Fig. 6.

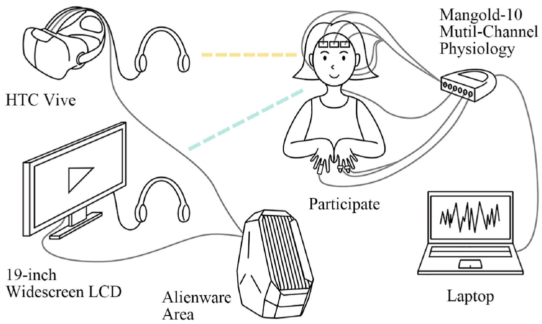

Fig. 7. The experiment system.

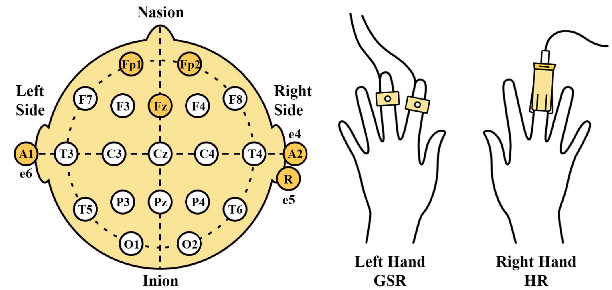

Fig. 8. The positions of the electrodes, HR and GSR devices.

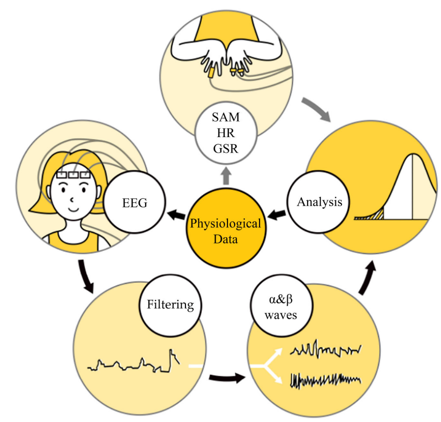

Fig. 9. The specific process of data processing and analysis.

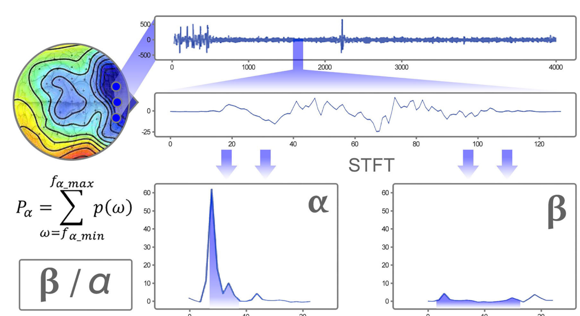

Fig. 10. The process of arousal evaluation method using EEG signals (Arouse: Beta/Alpha Ratio).

2) Participants: 9 participants' data were excluded due to the problems of physiological data acquisition equipment, environmental and artifact interferences during the experiments. Therefore only 31 participants' data were used eventually (13 males and 18 females, the age was 23 ± 2.1).

3) Experimental Materials: In the resting state, the screen and VR device both displayed black contents which were kept for 30 seconds. Then, two positive and two negative emotional VR scenes were selected as experimental materials in group 1, while the same scenes in the format of 2D videos were delivered in group 2. After watching the scenes, participants were asked to rest for 30 seconds with the black content after filling in a SAM scale.

4) Experimental Instruments: The experimental system was composed of a video playback system and a virtual reality system, which included an HTC-vive device (HTC VIVE, HTC Corporation, Taiwan), a desktop host (Alienware Area-51R2, Alienware, China), a 19-inch widescreen LCD (LG W2363DV, LG Electronics, China), a laptop (Alienware m15 R2, Alienware, China), a mangold-10 multi-channel physiological device (Mangold-10 Multil-Channel Physiology, Mangold International, German), data analysis software, and a number of K-1 (AH) disposable patch electrodes. The whole system was shown in Fig. 7.

During the entire experiment, the physiological data from the participants, including the forehead EEG, skin conductance response and heart rate, were continuously collected.

The forehead electrodes of FP1, FP2, and FPZ were chosen to obtain EEG signals, which were attached to the left, middle, and right positions of the forehead. The electrodes 4, 5 for obtaining reference signals were attached to the right posterior ear. The electrode 6 was attached to the left ear at the back, and the ground electrode 7 was attached to the occipital protuberance below. The device, acquiring the heart rate signal, was clipped the index finger of the right hand of the participants, and the devices, obtaining the skin conductance response, were wrapped around the index finger and the middle finger of the left hand of the participants. The positions of all the devices were shown in Fig. 8.

5) Data Processing and Analysis: All the physiological data including EEG, HR, and GSR were derived from a software named Biotrace+. Particularly, EEG data was filtered by a band-pass filter to extract the alpha and beta wave for the conducting of mean analysis and t-text between groups. The process was shown in Fig. 9.

6) EEG, GSR and HR: Electroencephalogram (EEG) was a method of recording brain activity by using electrophysiological indicators which consisted of five bands named δ(1–3 Hz), θ(4–7 Hz), α(8–13 Hz), β(14–30 Hz), γ(30–100 Hz). It was demonstrated that changes in alpha and beta waves contained important information of emotions. The β/α index was pointed out that could be used to estimate arousal of emotions [26] as shown in Fig. 10. The EEG signal was collected, part of which was enlarged to extract Alpha band and Beta band by Short-Time Fourier Transform (STFT). The Alpha band power and Beta band power were then calculated to measure Arousal (Arousal: Beta/Alpha Ratio).

TABLE I
VIDEO MATERIAL DESCRIPTION

NO	name	Emotion	Time length	Summary	Start of the scene	End of the scene
1		Peaceful (Rainforest Scenery)	80s	A sightseeing rail car tours in the forest.	A man sat in a sightseeing bus in the forest, surrounded by open grasslands	The sightseeing bus stopped by the lake, and there were wild ducks flying across the water.
2		Happy (Fairy tales)	80s	A man rode a roller coaster to visit the fairyland.	A man went through the fairyland.	A man looked the beauty of the fairy world.
3		Distasteful (Chaotic house)	80s	A man walked in a house full of reptiles.	There were lots of worms on the table in front of them.	There were lots of worms on the table in front of them.
4		Fearful (Abandoned House)	80s	A man explored in an abandoned old house.	A man woke up from a coma and finds himself in a gloomy room.	A man hid behind a stone ball and faced a monster.

Fig. 11. The plots of VR materials.

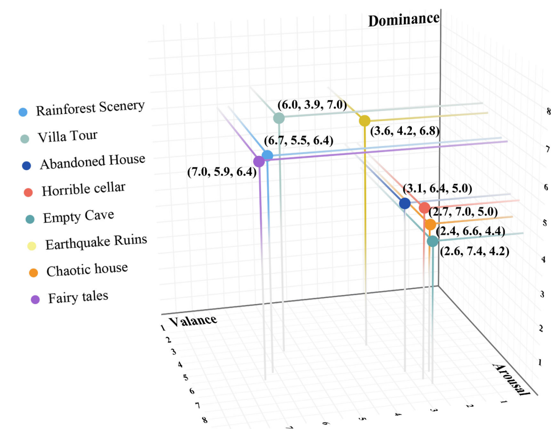

Fig. 12. The distribution of emotional space of eight VR scenes.

Galvanic Skin Response (GSR) [27] was a psychophysiological measurement which referred to the changes in electrical conductivity caused by activation of the sweat glands on the skin surface due to stress or other stimuli. In Mandryk's research, skin data could be used to explore the emotional changes when participant was playing computer games [28].

Heart rate (HR) referred to the number of heart beats per unit time, which could reflect emotional and cognitive activities [29]–[31]. It was an important physiological indicator for judging a person's emotional and cognitive activities. Ward and Marsden found that people's heart rate was much higher when browsing a rough-designed website than when browsing a well-designed website [32].

III. RESULTS

A. Scenes Design

The eight VR scenes were created with the title of Rainforest Scenery, Villa Tour, Abandoned House, Horror Palace, Empty Cave, Earthquake Ruins, Chaotic House and Fairy Tale Castle as shown in Fig. 11.The details of the four scenarios we selected for arousal evaluation were shown in Table I.

B. SAM Evaluation

Through the SAM evaluation, the mean scores of valence, arousal and dominance of those scenes were got as shown in Fig. 12. It was verified that the emotions inducted by our VR scenes were successfully distributed in the three-dimensional emotional space contained multiple emotions.

C. Arousal Evaluation

In the process of the experiment, due to the complication of the physiological data acquisition equipment, the environmental interference, and body movement affecting the accuracy of EEG data, 9 participants were excluded. 31 participates were obtained during the data processing (13 males and 18 females, the age was 23 ± 2.1).

- EEG Results

As was shown in Table II, the mean values of EEG factors under VR stimulations were all slightly higher than those under

TABLE II
EEG RESULTS (M±SD)

	Peaceful			Happy		
	A1	A2	A3	A1	A2	A3
VR	5.055±4.299	4.427±3.614	1.080±0.052	4.967±4.185	3.963±3.551	1.055±0.022
Video	4.205±3.288	3.442±2.878	1.054±0.023	3.261±3.094	2.849±3.519	1.062±0.039
p-value	0.517	0.308	0.400	0.177	0.165	0.440
	Distasteful			Fearful		
	A1	A2	A3	A1	A2	A3
VR	4.605±3.592	3.499±3.16	1.069±0.046	4.967±4.185	3.963±3.551	1.055±0.022
Video	2.786±2.300	2.628±2.286	1.067±0.048	3.261±3.094	2.849±3.519	1.062±0.039
p -value	0.103	0.085	0.783	0.177	0.165	0.440

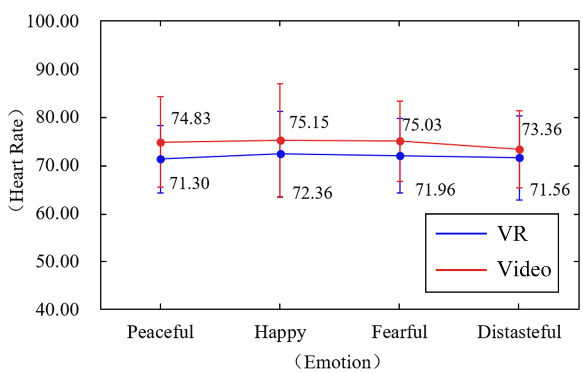

Fig. 13. The heart rate results.

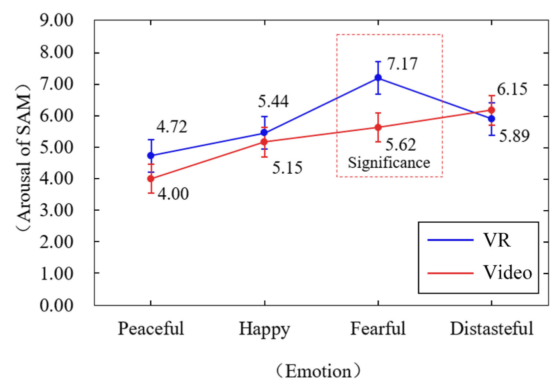

Fig. 14. The SAM results.

TABLE III
GALVANIC SKIN REACTION RESULTS (M±SD)

	Pleasure	Happy	Fearful	Distasteful
VR	3.246±0.127	3.432±0.344	3.293±0.255	3.325±0.195
Video	3.233±0.189	4.896±0.227	3.974±0.322	4.577±0.260
p-value	0.992	0.375	0.655	0.453

video stimulations, while there was no significant difference between them.

- Heart Rate Results

As was shown in Fig. 13 of HR, the mean arousal of VR for the four emotions is lower than that of video. Also, there was no significant difference between the VR and the video for each emotion arousal.

- Galvanic Skin Response Results

As was shown in Table III of GSR, the mean arousal of VR was slightly bigger than that of video for Peaceful, but smaller than that of video for the rest three emotions. Also, there was no significant difference between the VR and video for each emotion arousal.

- SAM Results

As was shown in Fig. 14 of SAM, of SAM, the mean arousal of VR was higher than that of video for Peaceful, Happy and Fearful, but lower than that of video for distasteful. It could be seen that the arousal of VR is significantly different from that of video for fearful ($p(\text{Fearful}) = 0.028 < 0.05$), while there was no significant difference between the VR and video for the rest three emotions.

The signal detection results of EEG, GSR, and HR suggested that there was no significant difference between the arousal of VR and video. According to the statistical analysis of SAM scales, in Peaceful, Happy, and Distasteful arousal experiment, there was no significant difference between the arousal of VR and video. However, in Fearful arousal experiment, there was a significant difference between the arousal of VR and video.

IV. DISCUSSIONS

The results proved that all the scenes in AVRS could achieve emotion elicitation, and also proved the hypothesis that the method of scenes design was referring to the elements of the same VA value extracted from the existing database was feasible. The result of VR Fearful and Peaceful scenes was consistent with previous findings, in which the authors used VR scenes for psychological exposure therapy [33] and emotion regulation [34] successfully. It also proved that VR pleasant and happy scenes could fulfill the function of emotion elicitation, which demonstrated that the AVRS could meet the needs of different emotion elicitation. This system could be widely applied to many fields, such as psychological research, mental illness diagnosis and virtual reality interaction research. Furthermore, 3-D visual environments made for a better fit to our perceptual system [35], and it provided a space with higher privacy and immersion at the user experience level. Therefore, the establishment of AVRS was meaningful [36].

In SAM measurement, the arousal of VR is significantly different from that of video for Fearful. This was consistent with results by researchers who used VR scenes to treat acrophobia [37], claustrophobia [38], flight phobia [39], and post-traumatic stress disorder [40] indicating that VR is more effective than traditional therapy. For the reason that people had attentional bias [41] for negative emotions such as Fearful and VR made people less likely to be disturbed by external environment with its immersion and privacy to be emotionally aroused in almost real-life environment, it was more effective to trigger fear under VR scene.

On the other hand, there was little difference in other emotion analysis utilizing SAM and psychological measurement. This result was consistent with the results of many other experiments. For example, Weidner and Hoesch investigated the different impact of non-VR and VR on physiological responses, simulation sickness, and driving performance within a single driving simulator, and found that there was no significant difference regarding physiological responses [42]. In Natalie's study of inducing fear emotion using virtual reality and 2D video, results of all the physiological assessment (i.e.,heart rate, blood pressure) showed that both VR and 2D induced Fearful significantly, but there was no significant difference between these video screen conditions [43].

However, there were still a few studies suggested that VR materials elicit stronger emotions [44], [45]. In Moseley's VR meditation and relaxation experiment [46], results showed that the VR meditation environment with biofeedback had the best meditation effect because it could help maintain concentration. And in Kosunen's study [47], a neuroadaptive virtual reality meditation system that combined virtual reality with neurofeedback was developed, which demonstrated that meditation with both head-mounted display and neurofeedback worked best.

The common feature of these studies was that during the experiment, the participants were required to have interaction with the VR scenes. But in our study, participants were required not to move their body during the experiment in order to ensure the accuracy of experimental phycological data, which resulted in an indirect interaction between the contents of VR materials and participants. Thus, the interaction of VR scenes was speculated to be the main factor that might affect the arousal efficiency of emotion elicitation.

At the same time, VR resolution and the delicacy of VR scenes design were considered that it could also affect the experimental results [48]. With the improvement of VR headset display technology, the 8k resolution can present clearer video. The high-quality rendered VR scenes could bring better sensory experience to users, thus would improve the ecological validity.

Therefore, in future work, some reasonable interaction elements will be needed in VR scenes to verify our inference. And more VR scenes with high qualities in 8k resolution should be developed. The other two dimensions of emotion, valence and dominance, should be included in the evaluation of the emotion elicitation effect, and more subjects should be recruited. In order to further verify the scientificity and effectiveness of AVRS system, a clinical comparative experiment would be conducted involving participates with depression symptoms.

V. Conclusion

In this paper, we extracted multi-features from the existing referenced affective systems, art works and existing VR scenes to complete AVRS and verified this system with SAM measurement. Furthermore, we used arousal as an indicator of ecological validity to compare the difference between virtual reality materials and video materials through objective and subjective assessment demonstrating that VR could elicit emotions as well as video and Fearful had stronger elicitation in subjective assessment, while there was no significant difference in emotional arousal in other emotional measures. There was significant difference in measures of Fearful in SAM evaluation, indicating that VR emotion materials were expected to deliver a good effect on negative emotional scenes.

Acknowledgment

The authors would like to thank L. Zhu and J. Xie for their assistance in the experiment.

References

[1] J. James, W. Robert, and Levenson, "Emotion elicitation using films," *Cogn. Emotion*, vol. 1, no. 1, pp. 87–108, Jan. 2007.

[2] M. Martin, "On the induction of mood," *Clin. Psychol. Rev.*, vol. 10, no. 6, pp. 669–697, Dec. 1990.

[3] P. Zheng *et al.*, "An overview of mood-induction methods," *Adv. Psychol. Sci.*, vol. 20, no. 1, pp. 45–55, Jan. 2012.

[4] M. M. Bradley and P. J. Lang, "Affective norms for english words (ANEW): Instruction manual and affective rating," Center Res. Psychophysiol., Univ. Florida, Gainesville, FL, USA, Tech. Rep. C-1, 1999.

[5] C. F. O. Emotion and A. Cseanimh, "The international affective picturesystem Digitized photographs," in *Encyclopedia of Personality and Individual Differences*. New York, NY, USA: Springer, 1999.

[6] P. Granato, S. Vinekar, O. Godefroy, J.-P. Vangansberghe, and R. Bruyer, "A study of visual recognition of facial emotional expressions in a normal aging population in the absence of cognitive disorders," *Open J. Psychiatry*, vol. 4, no. 3, pp. 251–260, Jul. 2014.

[7] B. Lu, M. Hui, and H. Yuxia, "The development of native chinese affective picture system–A pretest in 46 college students," (in Chinese) *Chin. Mental Health J.*, vol. 19, no. 11, pp. 719–722, Nov. 2005.

[8] M. M. Bradley and P. J. Lang, *The International Affective Digitized Sounds: Affective Ratings of Sounds and Instruction Manual*. Gainesville, FL, USA: Univ. Florida, 2007.

[9] T. S. Liu, Y. J. Luo, H. Ma, and Y. X. Huang, "The compilation and evaluation of localized emotional sound library," (in Chinese) *Psychosocial Sci.*, vol. 26, no. 2, pp. 406–408, 2006.

[10] C. Ferdenzi *et al.*, "Affective dimensions of odor perception: A comparison between Swiss, British, and Singaporean populations," *Emotion*, vol. 11, no. 5, pp. 68–81, May 2011.

[11] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cogn. Emotion*, vol. 9, no. 1, pp. 87–108, Jan. 2008.

[12] S. Alexandre *et al.*, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers," *Cogn. Emotion*, vol. 24, no. 7, pp. 1153–1172, Oct. 2010.

[13] P. F. Xu, Y. X. Huang, and Y. J. Luo, "Preliminary compilation and evaluation of Chinese emotional image materials library," (in Chinese) *Chin. J. Mental Health*, vol. 24, no. 7, pp. 551–554, Jul. 2010.

[14] P. M. Niedenthal, "Embodying emotion," *Science*, vol. 316, no. 5827, pp. 1002–1005, Jun. 2007.

[15] D. D. Dougherty *et al.*, "Ventromedial prefrontal cortex and amygdala dysfunction during an anger induction positron emission tomography study in patients with major depressive disorder with anger attacks," *Arch. Gen. Psychiatry*, vol. 61, no. 8, pp. 795–804, Sep. 2004.

[16] C. D. Hughes *et al.*, "The subscale specificity of the affective control scale: Ecological validity and predictive validity of feared emotions," *Motivation Emotion*, vol. 39, no. 6, pp. 984–992, Jun. 2015.

[17] W. Zhang, L. Shu, X. Xu, and D. Liao, "Affective virtual reality system (AVRS): Design and ratings of affective VR scenes," in *Proc. 7th Int. Conf. Virtual Reality Vis.*, Zhengzhou, China, 2017, pp. 311–314.

[18] J. Pittermann, A. Pittermann, and W. Minker, "Human emotions," in *Handling Emotions in Human-Computer Dialogues*. Dordrecht, The Netherlands: Springer, 2010, pp. 19–45.

[19] P. Ekman, R.W. Levenson, and W. V. J. S. Friesen, "Autonomic nervous system activity distinguishes among emotions," *Science*, vol. 221, no. 4616, pp. 1208–1210, Sep. 1983.

[20] D. Liao *et al.*, "Design of virtual reality scenes with variable levels of fear evocation," in *Proc. Int. Conf. Virtual Augmented Mixed Reality*, 2018, pp. 103–115.

[21] M. Gervautz and W. Purgathofer, "A simple method for color quantization: Octree quantization," *Graph. Gems*, vol. 36, no. 5, pp. 287–293, Dec. 1990.

[22] P. H. Jun, K. K. Baek, and C. Eui-Young, "An effective color quantization method using octree-based self-organizing maps," *Comput. Intell. Neurosci.*, vol. 2016, Jan. 2016, Art. no. 5302957.

[23] B. Guo and X. Fu, "A modified octree color quantization algorithm," in *Proc. 1st Int. Conf. Commun. Netw.*, 2006, pp. 1–3.

[24] M. Margaret *et al.*, "Measuring emotion: The self-assessment manikin and the semantic differential," *Behav. Therapy Exp. Psychiatry*, vol. 25, no. 11, pp. 49–59, Mar. 1994.

[25] Y. Xie and Z. Yang, "A comparative study on the effectiveness of different emotional induction methods," (in Chinese) *Stud. Psychol. Behav.*, vol. 14, no. 5, pp. 591–599, 2016.

[26] D. O. Bos, "EEG-based emotion recognition," *Influence Vis. Auditory Stimuli.*, Jan. 2006, pp. 1–17.

[27] L. Shu *et al.*, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, Jun. 2018, Art. no. E2074.

[28] R. Mandryk, K. Inkpen, and T. Calvert, "Using psychophysiological techniques to measure user experience with entertainment technologies," *Behav. Inf. Technol.*, vol. 25, no. 2, pp. 141–158, Mar. 2006.

[29] W. Boucsein, *Electrodermal Activity*. New York, NY, USA: Springer, 2012.

[30] O. Pollatos, B. M. Herbert, E. Matthias, and R. Schandry, "Heart rate response after emotional picture presentation is modulated by interoceptive awareness," *Int. J. Psychophysiol.*, vol. 63, no. 1, pp. 117–124, Feb. 2007.

[31] A. Harada, K. Watanuki, K. Kaede, and K. Muramatsu, "Evaluation of emotions induced by 2D and 3D video systems in a two-dimensional emotion space," in *Advances in Affective and Pleasurable Design*. Cham, Switzerland: Springer, 2019.

[32] R. D. Ward and P. H. Marsden, "Physiological responses to different WEB page designs," *Int. J. Human-Comput. Stud.*, vol. 59, no. 1/2, pp. 199–212, Jul. 2003.

[33] D. Liao *et al.*, "Arousal evaluation of VR affective scenes based on HR and SAM," in *Proc. IEEE MTT-S Int. Microw. Biomed. Conf.*, 2019, pp. 1–3.

[34] B. O. Rothbaum *et al.*, "Virtual reality exposure therapy for PTSD Vietnam veterans: A case study," *J. Traumatic Stress*, vol. 12, no. 2, pp. 263–271, Apr. 1999.

[35] I. Spence and F. Jing, "Video games and spatial cognition," *Rev. Gen. Psychol.*, vol. 14, no. 2, pp. 92–104, Jun. 2010.

[36] N. Kiefl, P. Figas, and C. Bichlmeier, "Effects of graphical styles on emotional states for VR-supported psychotherapy," in *Proc. 10th Int. Conf. Virtual Worlds Games Serious Appl.*, 2018, pp. 1–4.

[37] P. M. G. Emmelkamp *et al.*, "Virtual reality treatment versus exposure in vivo: A comparative evaluation in acrophobia," *Behav. Res. Therapy*, vol. 40, no. 5, pp. 509–516, Jun. 2002.

[38] C. Yeh, C.-F. Tsai, C. Yu, and T.-C. Huang, "Compare virtual reality and augmented reality systems for claustrophobia from HRV," in *Proc. 12th Int. Conf. ITS Telecommun.*, 2012, pp. 155–159.

[39] B. O. Rothbaum *et al.*, "A controlled study of virtual reality exposure therapy for the fear of flying," *J. Consulting Clin. Psychol.*, vol. 68, no. 6, pp. 1020–1026, Jan. 2001.

[40] V. Guillén *et al.*, "Users' opinion about a virtual reality system as an adjunct to psychological treatment for stress-related disorders: A quantitative and qualitative mixed-methods study," *Frontiers Psychol.*, vol. 9, Jun. 2018, Art. no. 1038.

[41] M. Dyck *et al.*, "Negative bias in fast emotion discrimination in borderline personality disorder," *Psychol. Med.*, vol. 39, no. 5, pp. 855–864, Sep. 2018.

[42] W. Florian, A. Hoesch, S. Poeschl, and W. Broll, "Comparing VR and non-VR driving simulations: An experimental user study," in *Proc. IEEE Virtual Reality*, 2017, pp. 281–282.

[43] C. Natalie van derWal *et al.*, "Inducing fear: Cardboard virtual reality and 2D video," in *Proc. Int. Conf. Human–Comput.*, 2017, pp. 711–720.

[44] H. Kenta, H. Qin, and J. Kobayashi, "Preliminary test of affective virtual reality scenes with head mount display for emotion elicitation experiment," in *Proc. 17th Int. Conf. Control, Autom. Syst.*, 2017, pp. 325–329.

[45] J. Ham, D. Cho, J. Oh, and B. Lee, "Discrimination of multiple stress levels in virtual reality environments using heart rate variability," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2017, pp. 3989–3992.

[46] R. Moseley, "Deep immersion with Kasina: An exploration of meditation and concentration within virtual reality environments," in *Proc. Comput. Conf.*, 2017, pp. 523–529.

[47] I. Kosunen *et al.*, "RelaWorld: Neuroadaptive and immersive virtual reality meditation system," in *Proc. 21st Int. Conf. Intel. User Interfaces*, 2016, pp. 208–217.

[48] P. P. Routhier, "Virtually perfect: Factors affecting the quality of a VR experience and the need for a VR content quality standard," in *Proc. SMPTE 2016 Annu. Tech. Conf. Exhib.*, 2016, pp. 1–20.

Dan Liao was born in Wuhan, Hubei, China, in 1985. She received the M.A. degree in industrial design from the Wuhan University of Technology, Wuhan, China, in 2010. She is currently working toward the Ph.D. degree on the research of human–computer interaction with the School of Electronics and Information Engineering, South China University of Technology, Guangzhou, China. Her research interests include virtual reality and human–computer interaction.

Lin Shu (S'10–M'13) received the B.E. degree in information engineering and the M.E. degree in communication and information systems from the South China University of Technology, Guangzhou, China, in 2005 and 2008, respectively, and the Ph.D. degree from the Institute of Textiles and Clothing, Hong Kong Polytechnic University, Hong Kong, in 2012. He is currently a Senior Engineer with the Guangzhou Key Lab of Body Data Science, South China University of Technology.

Guodong Liang received the M.Ed. degree in psychology from Shihezi University, Shihezi, China, in 2017. He is currently working toward the Ph.D. degree on the research of human–computer interaction with the School of Electronics and Information Engineering, South China University of Technology, Guangzhou, China. His research interests include virtual reality, machine learning, and human–computer interaction.

Yingxuan Li received the B.E. degree in industrial design from the South China University of Technology, Guangzhou, China. She is currently working toward the master's degree in electronic system and signal processing with the South China University of Technology, Guangzhou, China. Her research interests mainly include virtual reality, human–computer interaction, and machine learning.

Yue Zhang received the B.E. degree in industrial design from the South China University of Technology, Guangzhou, China. She is currently working toward the master's degree in electronic system and signal processing with the South China University of Technology, Guangzhou, China. Her research interests mainly include virtual reality, machine learning, and emotion recognition based on multi-physiological signals.

Wenzhuo Zhang is currently working toward the master's degree in electronics and communication engineering with the South China University of Technology, Guangzhou, China, where she is also a master student with Guangzhou Key Lab of Body Data Science. Her research interests include affective virtual reality system, human computer interaction, and image recognition.

Xiangmin Xu (M'13) received the Ph.D. degree from the South China University of Technology, Guangzhou, China. He is currently a Full Professor with the School of Electronic and Information Engineering, South China University of Technology. His current research interests include image/video processing, human computer interaction, computer vision, and machine learning.