Affecting Audience Valence and Arousal in 360 Immersive Environments: How Powerful Neural Style Transfer Is?

Yanheng Li[1], Long Bai[2], Yaxuan Mao[1], Xuening Peng[3], Zehao Zhang[4], Antoni B. Chan[1], Jixing Li[1], Xin Tong[5(✉)], and RAY LC[1(✉)]

[1] City University of Hong Kong, Hong Kong, China
{lydia.yh-li,yaxuanmao2-c}@my.cityu.edu.hk,
{abchan,jixingli}@cityu.edu.hk, lc@raylc.org

[2] The Chinese University of Hong Kong, Hong Kong, China
b.long@link.cuhk.edu.hk

[3] University of Florida, Florida, USA
xuening.peng@ufl.edu

[4] The University of Waterloo, Ontario, Canada
z783zhang@uwaterloo.ca

[5] Duke Kunshan University, Suzhou, China
xin.tong@dukekunshan.edu.cn

Abstract. Immersive experiences in Virtual Reality (VR) platforms often require customized content that can be adapted to in-game situations and specific player actions that lead to individualized effects. Controlling the positive-negative emotional valence and the level of arousal of the immersive content allows VR systems to provide situation-specific and action-specific affective influence on players, giving them an experience that is tailor-made for the narrative and interaction espoused by the system. To generate different emotional influences for the same content, we created a system that uses Neural Style Transfer (NST) along with a set of style images with known affective ratings to procedurally generate different versions of the same 360 environments in VR with differing affective influences for players. To explore how the NST-generated VR affects participants' affective perception, we conducted two user studies (N=30 and N=28). Users experienced four separate VR environments with different affective ratings. After each experience, we performed a survey to evaluate their affection, including Emotional Matching Tasks and interviews. Findings suggested that users are more likely to be aware of arousal differences than valence differences, which are mainly perceived by the degree of contrast between color and content of the environment. The stylized features gained from NST that affect the perception of valence are the color tone, the clarity of the texture, and the familiarity of the content for the user. Our work contributes novel insight into how users respond to generated VR environments and provides a machine-learning-based strategy for constructing an immersive environment to influence the affective experience of users, without altering any content and the game mechanism.

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2024, LNCS 14706, pp. 224–243, 2024.
https://doi.org/10.1007/978-3-031-61041-7_15

Keywords: Neural Style Transfer · 360 Image Generation · Affective VR · Affective Experience · Human-computer Interaction

1 Introduction

The immersive Virtual Reality (VR) environment can elicit affective responses through simulating different sensory experiences; for example, horrible lighting and sound effects could be used in VR game environments to stimulate users' negative affection [46,55]. In a narrative VR environment, especially VR games, visual design is a significant component of drawing users into specific affective states [59]. When we engage in VR games, observing and evaluating the aesthetic elements is an important means for us to perceive emotions in the game [18]. However, it is challenging to construct an immersive environment by adjusting precise details for particular emotional influence [62]. Some VR designers usually alter the affective states by arranging different game objects in the computer-graphic engine [9], while others import pre-recorded 360 images/video sequences into the engine to build a 360 virtual environment [42]. These approaches may take much time for the designer to prepare the properties.

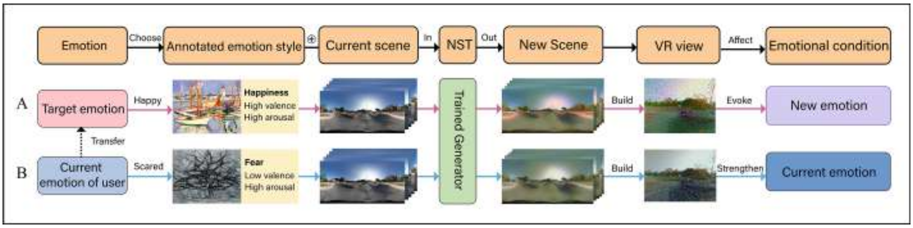

Fig. 1. Emotion-driven affective VR generation system. The *Path A* describes an approach to generate an affective VR environment based on the intention of the designer. For example, the designers want the user to be happy in a future environment; they could choose a series of style images from the datasets annotated with "happy" and then use them to perform Neural style transfer. The output 360 images could be used to construct a VR environment and evoke target emotion after the user experience each time. The emotion resource in *Path B* comes from the current affective condition of users, which may be detected with affective computing or self-report. So the selection of style images is based on the user. After generation, the user will experience the same but a more robust emotional environment to enhance the original emotion.

In non-VR video games, Artificial Intelligence (AI) has provided many benefits for reducing work in game development. Procedural Content Generation (PCG) techniques have been used in the video games industry to generate specific types of content, e.g., sounds, images, game levels, characters, textures, and dialogue [34]. Many Generative AI (GAI) techniques are employed to make the content creation process more effective and accessible. For example, ChatGPT [40] is helping with brainstorming stories [36] and facilitating dialogue [57];

DALL-E-2 [45] produces images properties allowing users using natural language prompts; Dream Fields is another model using the free-text prompt to achieve 3D object generation [30]. For controlling users' affective experiences, some other studies propose Emotion-Driven Adaptivity, an approach to adjusting the game mechanics to help users improve their in-game behavior after recognizing their affective state [51]. AI acts as an agent in the system to adjust pre-set contents to create new narratives and experiences. Though these GAI techniques are helpful for generating game content more quickly than before, designers may need to take much effort into trying different prompts to figure out the appropriate emotion-specific outcomes. They may also need to assemble all the components in the game engine manually. Thus, in terms of the core problem of emotion-specific VR environment development, an ML-based generation strategy still needs further exploration.

With all these opportunities for creating the adaptive emotional environment procedurally, there is limited research on investigating what exact properties should be adaptively changed for eliciting specific affective experiences, and how these NST-generated VR environments affect users' perception during the experience. Thus, we set the following **research questions**:

– What affective experience can be affected by the ML-based VR environment generation?
– What visual parameters of the generated VR environment can be adjusted to obtain the corresponding affective experience?
– How to control the changes of parameters for eliciting expected affective experience?

In our study, we employed Neural Style Transfer (NST) to generate affective 360 images and created a procedural VR-generating system. To get an environment with a definite emotion, we chose images with different emotional annotations as style images to let the machine learn and transfer the affective features to the 360 content images. We then assigned the generated 360 images to the skybox to build an affective VR environment. After the VR was established, we invited several participants to experience and measure their affective responses using multiple methods, including Emotional Matching tasks (EMT) and interviews. We conducted a formative study in static 360 environments and a formal study in navigated based VR environments. In the pre-study, we built four virtual environments with different valence and arousal values based on a single 360 image, in which the participants can only experience a static environment at a time. We measured the affective states of the participants after each experience. However, their emotional responses in the valence test groups differed from what was expected. Therefore, we modified the VR environments based on the comments provided by the participants in the interviews, and conducted our formal study, where we built the VR environment based on multiple 360 images and a navigation system to make the participant more immersive and thus enhance the affective stimulation.

Using AI to generate a 360-images-based affective VR environment with images of known emotions has many potential applications, which we discuss at the end of this paper. Our **contributions** are:

– We proposed an affective VR generating system using the NST approach (Fig. 1), which could help create affective VR environments without altering content.
– We contributed novel insights about how users responded to generative VR environments by reporting the survey and interview results of formative study (N=30) and formal study (N=28) on how people perceived the affective state of our generated VR environments and what specific visual elements influenced their perception and affective response. The results suggested that our method with discrete emotional word style images dataset could effectively elicit the corresponding affective responses. Transferred color and texture are the key visual elements that took effect. Moreover, generated color and texture have interactive effects when affecting user's affective state in the generated VR environments.
– We ended by discussing the potential applications of our generation system in real VR affective narrative space and game scenarios that automatically alter the affective state of the current environment. We also discussed how our findings of the visual components that affected the emotional expressions of AI-generated images could contribute to the prompt engineering in Artificial Intelligence Generated Content (AIGC) field.

2 Related Work

2.1 Emotion Measurement

A classic example of emotion measurement tools is "six basic emotions" proposed by Ekman: anger, disgust, fear, joy, sadness, and surprise [13]. Dimensional models, on the other hand, attempt to conceptualize human emotions by defining where they lie in multidimensional spaces [43]. For instance, the "circumplex model of affect" developed by James Russell suggests a two-dimensional circular space, containing valence and arousal values [48]. Valence measures the degree of an emotion being perceived as positive and negative; while arousal evaluates how strong the emotion is being felt. Psychologists also use the Positive Activation – Negative Activation (PANA) model [60] to measure self-reported positive and negative affect when testing whether emotions are influenced [25]. However, studies indicate that this model may be limited only to activating emotion states. Positive Activation scores have been found to increase during anger, suggesting that it does not always measure positivity but may be sensitive to approach motivation [26].

Our study applied the EMT instead to deal with the drawbacks of previous methods [37]. This method aims to examine people in emotion perception with multiple experiment components [56]. Different from other evaluations, the EMT considers multi-modal information of emotions, including emotional facial expressions (more than four basic emotional labels) [39], emotional words and emotional situations to compensate for the inaccuracy of positivity measuring in PANA model [2]. Therefore, our study adopted EMT to evaluate human affective states for more accurate and reliable results.

2.2 ML/AI Method for Affective Environment Generation

The use of NST leads to rapid generation and iteration of artwork production. People can use Convolutional Neural Networks (CNNs) to render a content image in different styles [21,22,32]. Based on the new method of image style transfer, the researchers use neural representations learned by CNNs to separate and recombine the content and style of arbitrary images, generating a similar image with the content image, and the style has also been reconstructed simultaneously. They work with the 19-layer VGGNet [52] pre-trained on ImageNet [11], achieved a separation of image content from style, then allowed synthesizing an image that combines the content of one image with style from another [21]. Style Transfer is not only applied in 2D images. For spherical images, researchers address the emerging demand for 3D movies or VR/AR [8]. A feed-forward network is proposed for stereoscopic style transfer [8]. The researchers addresses the challenges of style transfers applied to 360 images by employing a cubic projection to remap the equirectangular projection to a set of six cube faces [47]. The style transfer algorithm is applied to each face of the cube separately, and this methods are time-consuming. Other researches proposes a method directly processing with the stereoscopic images [23], which inspire us to use 360 images to generate VR.

2.3 Affective States Elicitation via VR and Customized Affective Game Experience

To date, numerous studies have already examined applications of VR that have been used to construct real-life scenes with different settings to arouse different human affective status [5,6,16,24,58]. For example, five virtual parks with different scenarios in previous research could elicit a specific affective state respectively (i.e., joy, sadness, boredom, anger, and anxiety) [16]. In contrast, others attempt to induce specific affective status by adjusting the lighting conditions and sounds in real-life virtual scenarios [58].

From those researches, we observed that adjusting the visual elements of a VR environment could effectively alter the perception of the affective state for different scenarios [4,12]. Therefore, inspired by the VR-driven applications of human emotion elicitation, our study aims to verify whether creating a VR environment via 2D panorama images with annotated valence and arousal values could arouse people's diverse affective states, and how the effects differ with respect to altered valence and arousal indexes.

In order to make it easier to develop game environments with different emotional states, customized affective games with smart agents are needed to help quickly and automatically optimize the emotional quality of content for experience when the player enters the environment. In some earlier studies, researchers tended to use AI to automatically adjust the mechanics and dynamics of the game according to the performance of players, thereby changing the gameplay experience to affect the in-game mood [53,54]. In addition to changing the decision tree, recent research on experience-driven procedural content generation framework also focused on the player emotion modelling [61]. Others also change the environment, e.g., text and NPCs. They stimulated alien emotional experiences by producing resemble but novel narratives according to the real-time emotions of players [28].

2.4 Content Generation in Serious Game Application

Serious games, as multimedia learning systems, usually provide players with a variety of games and learning paths to help the growth and advancement of knowledge and experience [3]. Therefore, PCG, as a convenient creative tool, is often used to create personalized game experiences in terms of individual-related background. In the Orange Care, a PCG-serious game for conveying educational content about skin lesions to primary care physicians, the character generation is based on the text describing some physical aspects and the illness condition [41]. The rendering of the appearance of the characters could influence the in-game experience of the users. Shader and texture generation can help the game influence our experience automatically.

Friendly Adaptive Technological Tools Against Cyberbullying (ATTAC) project shows an approach to generating the whole game environment via graphical domain-specific modeling language, which communicates with a computer in an alike natural language [31]. That means semantic emotion words may be used for the affective game space generation in the sense of our project. Meanwhile, Hiramon combines player self-emotional reports and machine learning methods allowing game systems to learn the relationship between player behavior and emotional experience. The game is able to know when the player is experiencing a particular emotion that is undesirable. It will adapt accordingly to make relevant responses and optimize in-game emotional experience [19].

3 Technical Implementation Methodology

We applied the data-driven Arbitrary-Style-Per-Model Fast NSF [22] as a primary method to generate the images with different valence and arousal values. Prior to this, the transformation parameters were needed when changing the style from the original photo to the stylized photo. We employed the style prediction network in [22] to predict the transformation parameters of each style. Therefore, we can change the style of the content image using a single content image and another unlinked style image. In this standard NST model, $Frobenius$ norm was used to minimize the difference of Gram matrix associated with the layer activations in style between the generated image and the content image, and $\mathcal{L}_2$-norm was used to minimize the difference of layer activations in style between the generated image and the reference style image.

In our experiment, we employed the pre-trained arbitrary image stylization model[1] [22] to perform fast artistic style transfer that works on arbitrary images. After performing NST and getting a new 360 output image, the 360 images was imported into Unity. Next, the texture shape of this 360 image was set as a cube and used to create a new skybox material. Finally, the material was assigned to the environment's material, thus rendering a panoramic view of the whole Unity scene. The VR camera was placed in the center of the whole scene. Once the scene was running in the VR headset, the users could experience the 360 scenes just like they were standing in a real 3D world.

4 Formative Study: Static 360 Scene Evaluation

4.1 Participants and Procedure

As reported in [33], we recruited 30 participants in this study (10 male, 20 female, from 18 to 38 years old ($M = 22.1$, $SD = 2.28$)). They were labeled with IDs from S1 to S30. The research procedures included a within-subject experiment in which the participants were assigned to four conditions with two different 360 content images (high valence, low valence, high arousal and low arousal VR environments) in random orders to avoid ordering effects. The first group, called the valence group, uses a 360 live performance image as the content image and two style images with inverse valence dimensions (high and low) as the input materials. The output images were imported into the game engine Unity to establish high and low-valence test environments. The other group is called the arousal group, whose content image is a 360 beach site image. Two style images with different arousal values (high and low) were used to influence the arousal condition of the test environment image. Every Participant was asked to watch four 360 outcome images in VR, as shown in Fig. 2. The VR conditions of different test groups were alternatively experienced (for example, high valence first, subsequent low arousal, then low valence, and last high arousal) to avoid decreased sensory perception after seeing two similar images continuously. In each experience, the participants were requested to accomplish an object-searching task in the environment within 4 min, ensuring to look through the whole environment. The objects the participants needed to search for were decided based on the content of the environment. Moreover, we also hoped the searching task would not distract the attention of the emotional experience as much as possible. Precisely, the participants were asked to count the number of people wearing the masks in the valence group because there were a lot of people standing around the participants in the VR environment. However, the arousal test environments contain very little human content but have a beautiful sunset beach view. Therefore, the participants were asked to find a favorite landscape of the environment. After each experience, the participants could have 5 min rest period, in which

[1] https://tfhub.dev/google/magenta/arbitrary-image-stylization-v1-256/2.

we conducted a short survey and interview to measure participants' affective states, and understand how NST-generated environments affected their affective experience. The Institutional Human Ethics Committee followed the whole procedures throughout the test.

4.2 Design of Test Materials

In order to test how the images generated by NST affect valence or arousal separately, we set two test groups (valence group and arousal group as shown in Fig. 2). Each group has a high and low-value environment for testing. All four style images were chosen from the Geneva affective picture database (GAPED)[2] [10]. We tried a series of style images with extremely distinct rating to make the outcomes show a striking contrast. For example, for the high-valence environment, we used a style image with a pleasant landscape and a high value (95 out of 100); for the low-valence environment, we used an image of scary spiders and a low valence value (12 out of 100). In our pilot test, we found that the landscape images in the P group (a group of pictures include human and animal babies as well as nature scenes) and the non-sense object images in the N group ((a group of pictures include inanimate objects) in the dataset are more suitable for our style images, because the pictures of other groups in the dataset are related to violence of moral and legal norms.

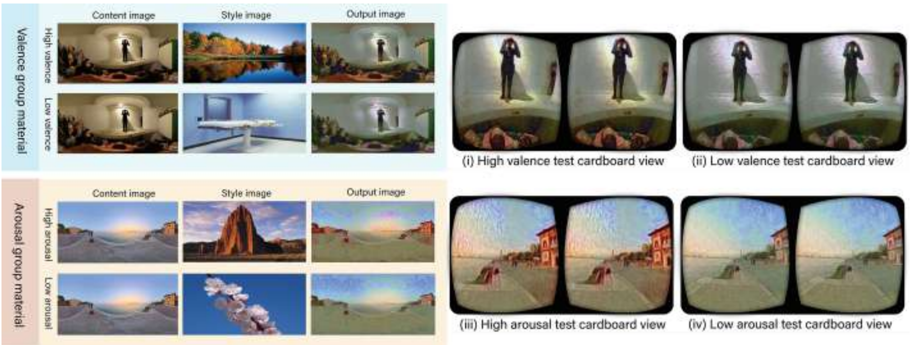

Fig. 2. Image materials for valence test group and arousal test group.

4.3 Formative Results

Discussion on Valence. Those images processed by NST are more likely to evoke negative emotions in the participants because NST introduces **textures that tend to reduce the familiarity and increase the negativity** of these

[2] http://www4.ujaen.es/~erpadial/GAPED.html.

content in the images to users. “I feel like the faces of these people in the environment are quite distorted, unlike what I would see in reality”. Most participants mentioned this in the interview (take S1, S7, S8, and S9 as cases). People who stay in an unfamiliar environment may have negative emotions such as fear. Brighter colors are considered positive in psychology [27]. However, **the perception of emotions for the environment may be influenced holistically by the combination of visual style (the color and the texture) with the content**. Therefore, the negative content, like scary human faces, makes the brighter color create tenser feelings in this case rather than comfortable. Thus, when using NST to create positive emotions in images, it appears necessary to deal with the visual style of style image that affects the texture to make the content more familiar and positive, which could be related to individual imagination based on emotional memory, the memory of experiences that evoked an emotional reaction. For example, some participants said “The texture looks like Van Gogh’s painting, which reminds me of the night he killed himself in the field”. Meanwhile, the color also needs to match the content. The emotional responses are also interactively affected by both color and content.

Discussion on Arousal. Unlike the results from the valence test group, which were opposite to our expectations, participants responded high-arousal affective state in the high-arousal environment, while they felt calmer (low-arousal) in the low-arousal environment. **The content of the 360 environments** used in the different tests (arousal vs. valence) caused systematic differences in participants’ responses. The content image of the valence test is a crowded indoor view, and the content of the arousal test is an open outdoor sightseeing scenery. The beach landscape, with the sunset and fewer people, makes it easier to recall people with some pleasant memories, so the content will not make them feel gross and scared (the high-arousal affective states). Moreover, the high-arousal condition had a **brighter color** than the low-arousal condition, which made participants have improved arousal levels, according to the interviews.

5 Formal Study: Navigation Based Evaluation

In the formative study, the perception of the valence differences from the participants was not statistically strong due to the blurred human content. We wanted to use the other content images with fewer human and outdoor views, the same as in the previous arousal test. The participants suggested in our formative study that their affective responses were weakened because of the sense of familiarity after looking at repeated content in a short period, which resulted in bias for the test analysis. Therefore, we tried to add more 360 content images as input to build a 360 navigation system like Google Maps, which allows the participants to explore more in a more extended period through rotating and changing their position by interacting with UI button using gaze cursor with the headset [49].

Moreover, navigation benefits the participants in clear recognition of the environment [17], which could help eliminate the unfamiliarity with the textured environment and enhance positive emotional responses.

5.1 Experiment Materials

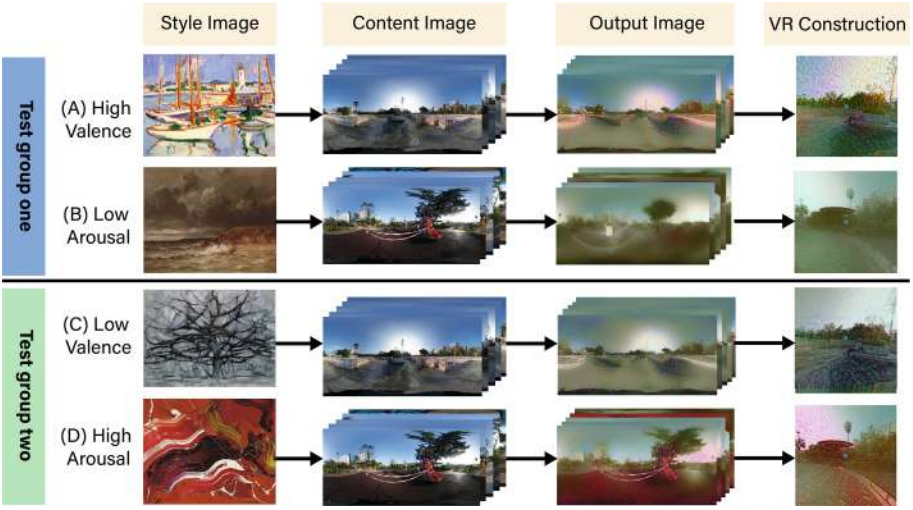

Fig. 3. Four VR environment construction and two test groups. The raw content images are used to build the control environments.

Design of Test Materials. In our formative study, we selected some annotated images from the GAPED dataset [10] as style images. They are mostly real-world landscape images. It is hard to justify what kind of visual elements in the style image have the affective effect that transferred to the 360 images. Therefore, in the formal study, the style images were selected from the WikiArt Emotion dataset[3] [38], which contains paintings from 195 artists. Artistic paintings could evoke an affective response in their viewers, and painters often express their personal emotions through every stroke, just as they imbue affective textures into their paintings [44]. Moreover, the emotion tags in the WikiArt dataset are more specific emotional words (e.g., happiness). As Fig. 3 shows, we randomly select four landscape paintings from the dataset to avoid getting strange textures from specific objects like human faces in some portraits. The emotional labels of these images represent high valence (happiness), low valence (fear), high arousal (fear), and low arousal (sadness). In addition to changing the style images, we add more content images to build a map for navigation in the VR environment.

[3] http://saifmohammad.com/WebPages/wikiartemotions.html.

We want to explore how it can influence emotional responses and improve the result by allowing them to experience in longer period and think about the situation.

Measures. Reliable and valid measurements are necessary to assess emotional knowledge with precision [39]. However, due to the complexity and invisibility of human emotions, the subjective choices made by individuals when asked to describe their affective state using a single word may be over-generalized [7]. The Emotional Matching Task (EMT) is a tool designed to facilitate effective communication and adaptive use of emotional expression. It comprises four parts: matching expressions, expression-situation matching, expression labeling, and expression label matching [1,2]. EMT can help participants reflect on their emotions and aid in assessing affective states.

In order to adapt the tasks for adults, we replaced the test materials with validated materials from the Geneva Emotion Recognition Test (GERT) and Geneva Emotion Knowledge Test (GEMOK), which were designed for adults but also assess emotional knowledge in similar task [50], [?]. The simple facial expression images in the first task were replaced with videos in the short version of GERT (GERT-S), which measures the ability to recognize emotions in the face, voice, and body of another person in a performance-based test [50]. Additionally, we provided participants with emotional word packages for the labeling task. Unlike children, adults are able to self-report their experiences verbally without the assistance of others, such as parents or teachers. Therefore, we required participants to self-report in the survey.

Participants and Procedure. We recruited 30 university students from social media interested in this topic (7 males and 23 females between 18 to 25 years old (M = 21.5, SD = 1.23)) as our participants. Two participants in the second test groups stopped the test due to VR sickness, so their data were not included in the results. We labeled the participants with IDs from P1 to P28. To avoid familiarity with the same content, we separated the participants into 2 test groups and conducted a between-subject experiment to make sure the contents of the environment were different in each group. Therefore, in Study 2, each test group of participants experienced 2 affective environments (Group 1: high valence*content 1 + low arousal*content 2, Group 2: low valence*content 1 + high arousal*content 2), and two control environments (raw 360 environments). The order of environment for each participant were randomized following a Latin square design. We engaged the participants in searching tasks as in formative study, allowing them to experience detailedly. After experiencing one VR environment, they are invited to fill out a questionnaire of EMT and some short answer questions about their related affective experience.

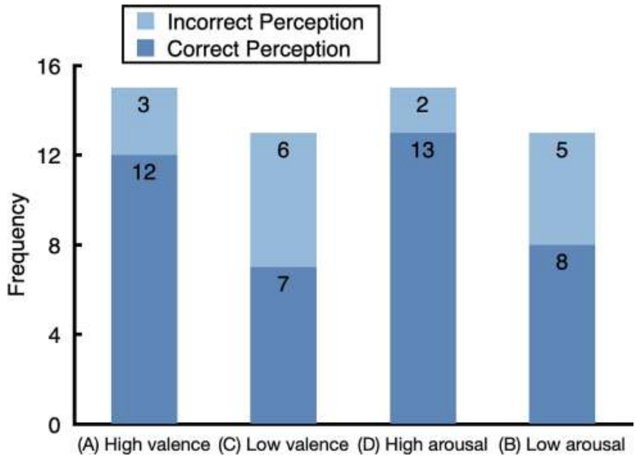

Fig. 4. Bar plot to show the ratio of correct perception in the four test groups. The bar is in order of high and low value for comparison, which is different from the group setting. This figure shows that in formal study, participants can correctly perceive specific emotion conditions, which means the NST approach can generate different emotional VR environments using different emotion-style images.

6 Formal Study Results

6.1 EMT Results Analysis

Each correct answer in the test is awarded one point, while incorrect answers receive zero points. The total score for each participant is calculated by summing up the scores for all the answers in each of the four parts of the test. A score of three or more indicates the successful perception of the expected emotion conveyed by the environment. As shown in Fig. 4, 80.0% of the participants in the high valence (A) group correctly perceived the expected affective condition, while 53.8% of the participants in the low valence (C) group correctly perceived the target emotion. 86.7% of the participants in the low arousal (B) group reported low-arousal affective responses, while those (61.5%) in the high arousal (D) group reported highly intense emotions. These results demonstrate that our generation strategy could effectively alter the participants' affective perception of the environment with annotated emotional style images.

6.2 Affective Stimuli

We list the factors that may influence the emotional responses of the participants for them to choose to help them investigate the concrete type of affective stimuli, including content in the environment, visual element, visual style, and spatial properties. We conducted a chi-square test to identify the differences in affective stimuli among different groups. The chi-square test reveals $x_2 = 3.327$, $p = 0.344$

($p > 0.05$), indicating that there are no significant differences among the types of stimuli in each group. It means there may not be one particular type (e.g., visual style) to affect the valence or arousal condition.

However, participants gave us feedback on the specific elements in the environment that evoked accurate affective responses, which can inspire us to perform the future generation.

Transferred texture can reshape the perception of particular game objects, thereby changing the emotional response. P1 mentioned "*water and waves*", which do not actually exist in the content. However, the transferred chaotic texture changed a blue wall in the environment into a swimming pool full of water. Because of the misinterpretation of the object content, which directly influences the recall of specific emotional memories, the participant felt relief other than negative emotions. This is why some participants reported incorrect perceptions of the low-valence environment. Potentially, we may be able to segment some specific items in the environment images, on which we can perform style transfer separately, letting the texture change the perception of the content, in other words, create new items. We can also recombine the patterns together at last to produce the whole environment.

Dense texture can result in blurred vision and raise the sense of unfamiliarity and high arousal. P3, P4, and P7 said "*I cannot see clearly and get scared*" Clear vision may reduce emotional arousal and increase valence. If we want to generate a scary scene, we can adjust the style parameters to produce more texture.

Complicated texture will narrow the space. More textures will make the space properties more prominent. People may feel the space is blocked and cramped. P8 mentioned "*oil painting, boundary and small space*". The complicated oil-painting style texture gained from the style image makes the VR environment seem to have an outer frame. The boundary narrows the whole space. The P12 said, "*When I was in B, it felt like being covered... which instantly reduced my positive feeling*". It inspires us to reduce the clarity of the texture when making negative environments.

Mismatch of the color and content can raise arousal and negativity. P17 told us "*In the D environment, I felt bloody and terrified, and my emotions were intense*". She also mentioned "*leisure park, contrast color, bright red and incompatible with the content*". The style image may greatly contrast with the content image for high arousal and negative environments. On the contrary, low arousal and positive environments can use a style image with a more matched and similar color property.

7 Contribution and Future Applications

We employed NST to transform real-world panoramas into VR environments with varying emotional properties. Our study also explored the potential of NST to generate VR environments with affective qualities and identified parameters that impact affective perception. These findings can be valuable references for VR designers and psychologists interested in creating affective VR experiences.

7.1 Manipulate Affective State on VR Psychological Study

In the realm of psychological research, VR has become a popular material used by researchers [20]. VR environments have been developed with the purpose of helping users confront various situations in the real world. For instance, VR has been employed to aid autistic children with affective disorders in learning how to recognize real-world emotions [29]. Additionally, exposure therapy combined with real-world scenarios has been used to assist patients in reducing fear and recovering from anxiety [15]. The primary materials used to create such VR environments include real-life panoramic photographs and specific emotional targets. Thus, the method proposed in this paper could be applied in emotion-based experiments to help psychologists swiftly iterate and identify an appropriate experimental VR space, replacing the conventional method of manual shooting on location. This method could also provide multiple alternatives. Unlike the prior process of constructing a VR environment with a single design goal, the proposed method permits the testing of several design objectives in a short period, while also saving additional environments in case the utilized one fails to meet the expectations.

7.2 Interactive Customization in Emotion-Driven Games

As depicted in Fig. 5, a game development plug-in can be designed. The plug-in can be integrated into the development engine and game system. With this plug-in, game designers can generate a game environment more easily. After pre-selecting the style and content images, designers can input the datasets to the designer side system. The AI agent then quickly generates a library of numerous scenes that game designers can choose to build the storyline and game level. Designers only need to select relevant images based on their narratives through the UI panel. Once designers set the emotion by rating the valence and arousal value using the slider, the system can automatically combine images according to the affective condition and fill in the corresponding position in the storyline and game level. Consequently, game designers can concentrate on creating a good story without spending too much time on technical implementation. Moreover, players can also arrange the stories as their expectations from the UI panel on the player side by adjusting the valence/arousal values. The storylines and scenes are then automatically matched from the library to generate new levels, providing players with more freedom. This marks a significant improvement as compared to the previous PCG games, where the emotional experience could only be continuously enhanced according to the designer's wishes.

7.3 References for Text Prompts

Using a single emotional word in the generation system may not generate the images with appropriate affective visual features, because the AI needs more detailed information to tailor the generated outcomes. We tried to add some

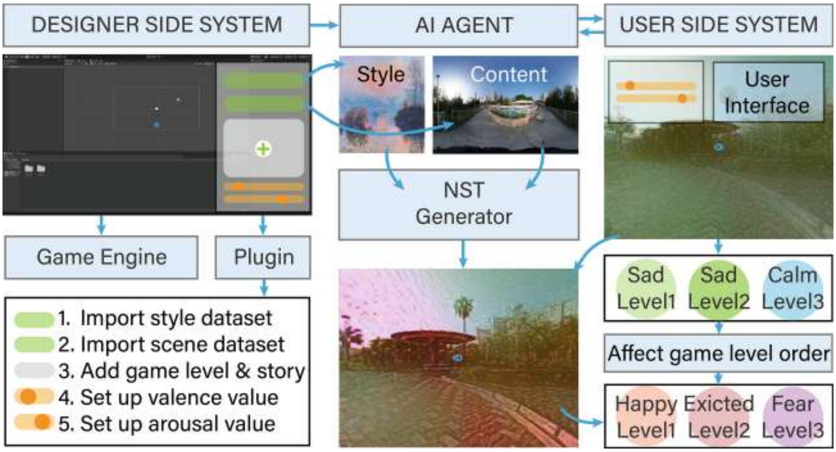

Fig. 5. Game application system hypothesis.

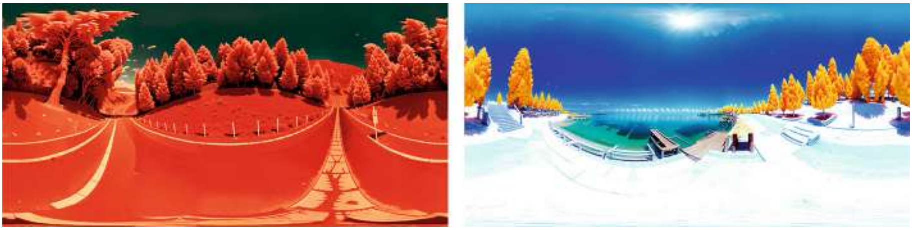

Fig. 6. Left is angry outcome, right is happy outcome.

descriptions of the affective feature we summarized in the prompt in the text-to-image generation system. For example, we prompt "low valence, high intense beach landscape, warm and red color tone, full of texture, blurred people." to generate a scary environment and prompt "high valence high-intensity beach land-scape, bright and relaxed color tone, clear texture, few people." to generate a happy environment. The outcomes are with quite distinct affective conditions, as shown in Fig. 6.

8 Conclusion, Discussion and Limitations

Our formative study justified the particular NST-generated visual elements that affect user's affection, and study two indicated how these elements take effect. The visual parameters that enable people to perceive the affective state of their environment are revealed through self-reports from participants. Typically, upon entering a new environment, people first pay attention to the content within it, such as seeing a park or observing individuals engaged in leisure activities. This content helps individuals quickly establish their understanding and familiarity with the environment. When utilizing NST to generate VR environments, it is important to prioritize the recognizability of the content. Transferred textures can then be used to blur or emphasize specific objects, eliciting either a pleasant or terrifying emotional experience within the environment. The color of an environment can significantly impact its perceived affective state. According to color psychology, red and orange are associated with warmth and cheerfulness, but they may also induce tension in some people [14]. Perception of color in combination with content varies significantly among different individuals. Color tone can directly impact arousal, with warm and strong colors typically resulting in higher levels of arousal [51]. Compared with a game called NEVERMIND [35], which assesses the emotions of the users by bio-sensors and needs to change the visual environment and gameplay mechanics to obtain particular emotional responses, our strategy Allows us only to generate a simple 360 environment with different content and objects to reach the same expectation.

Our study has some limitations. Firstly, the population of participants is not balanced and diverse enough, which may limit the generalizability of the results. Although we included tasks for participants, the VR experience may lack immersion because it is not a real game system that provides adequate interactions. Additionally, our measurement was not an in-game survey, which required participants to take off their headsets to complete the survey outside the VR environment. This interruption might weaken the affective response during the survey and interview. Bio-technologies could greatly help assess the objective and accurate affective state in real-time, providing a more accurate reflection of participants' affective state. Furthermore, our NST model is pre-trained with arbitrary images, which may not be the most effective method for transferring the affective features of annotated emotional style images, although it is effective for affective generation, as indicated by our studies. Therefore, a fine-tuned model with emotional style images may help improve the generation outcomes and affective influence on users.

References

1. Alonso-Alberca, N., Vergara, A.I., Fernández-Berrocal, P., Johnson, S.R., Izard, C.E.: The adaptation and validation of the emotion matching task for preschool children in Spain. Int. J. Behav. Dev. **36**(6), 489–494 (2012). https://doi.org/10.1177/0165025412462154
2. Alonso-Alberca, N., Vergara, A.I., Zappulla, C., Di Maggio, R., Pace, U., Sheffler, K.F.: Cross-cultural validity of the emotion matching task. J. Child Fam. Stud. **29**(4), 1159–1172 (2020)
3. Alvarez, J., Djaouti, D., et al.: An introduction to serious game definitions and concepts. Serious Games simul. Risks Manage. **11**(1), 11–15 (2011)
4. i Badia, S.B., et al.: Toward emotionally adaptive virtual reality for mental health applications. IEEE J. Biomed. Health Inform. **23**(5), 1877–1887 (2018)
5. Baños, R.M., Botella, C., Rubió, I., Quero, S., García-Palacios, A., Alcañiz, M.: Presence and emotions in virtual environments: the influence of stereoscopy. Cyberpsychol. Behav. **11**(1), 1–8 (2008)

6. Baños, R.M., Etchemendy, E., Castilla, D., García-Palacios, A., Quero, S., Botella, C.: Positive mood induction procedures for virtual environments designed for elderly people. Interact. Comput. **24**(3), 131–138 (2012)
7. Berrios, R.: What is complex/emotional about emotional complexity? Front. Psychol. **10**, 1606 (2019)
8. Chen, D., Yuan, L., Liao, J., Yu, N., Hua, G.: Stereoscopic neural style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6654–6663 (2018)
9. Craig, A.B., Sherman, W.R., Will, J.D.: Developing virtual reality applications: foundations of effective design. Morgan Kaufmann (2009)
10. Dan-Glauser, E.S., Scherer, K.R.: The Geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance. Behav. Res. Methods **43**(2), 468–477 (2011)
11. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
12. Dinis, S., Duarte, E., Noriega, P., Teixeira, L., Vilar, E., Rebelo, F.: Evaluating emotional responses to the interior design of a hospital room: a study using virtual reality. In: Marcus, A. (ed.) DUXU 2013. LNCS, vol. 8014, pp. 475–483. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39238-2_52
13. Ekman, P.: Basic emotions. In: Handbook of Cognition and Emotion, pp. 45–60 (1999). https://doi.org/10.1002/0470013494.ch3, https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013494.ch3, section: 3 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470013494.ch3
14. Elliot, A.J., Maier, M.A.: Color psychology: effects of perceiving color on psychological functioning in humans. Annu. Rev. Psychol. **65**(1), 95–120 (2014)
15. Emmelkamp, P.M., Bruynzeel, M., Drost, L., van der Mast, C.A.G.: Virtual reality treatment in acrophobia: a comparison with exposure in vivo. Cyberpsychol. Behav. **4**(3), 335–339 (2001)
16. Felnhofer, A., et al.: Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. Int. J. Hum Comput Stud. **82**, 48–56 (2015)
17. Ferguson, C., Van den Broek, E.L., Van Oostendorp, H.: On the role of interaction mode and story structure in virtual reality serious games. Comput. Educ. **143**, 103671 (2020)
18. Frome, J.: Eight ways videogames generate emotion. In: DiGRA Conference, pp. 831–835 (2007)
19. Frommel, J., Schrader, C., Weber, M.: Towards emotion-based adaptive games: Emotion recognition via input and performance features. In: Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play, pp. 173–185. CHI PLAY '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3242671.3242672
20. Gaggioli, A.: Using virtual reality in experimental psychology. Towards Cyberpsychol., 157–174 (2001)
21. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015)
22. Ghiasi, G., Lee, H., Kudlur, M., Dumoulin, V., Shlens, J.: Exploring the structure of a real-time, arbitrary neural artistic stylization network. arXiv preprint arXiv:1705.06830 (2017)
23. Gong, X., Huang, H., Ma, L., Shen, F., Liu, W., Zhang, T.: Neural stereoscopic image style transfer. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 54–69 (2018)

24. Guixeres, J., et al.: Effects of virtual reality during exercise in children (2013)
25. Harmon-Jones, C., Bastian, B., Harmon-Jones, E.: The Discrete Emotions Questionnaire: a new tool for measuring state self-reported emotions. PLoS ONE **11**(8), e0159915 (2016). https://doi.org/10.1371/journal.pone.0159915, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4976910/
26. Harmon-Jones, E., Harmon-Jones, C., Abramson, L., Peterson, C.K.: PANAS positive activation is associated with anger. Emotion **9**(2), 183–196 (2009). https://doi.org/10.1037/a0014959, place: US
27. Hemphill, M.: A note on adults' color-emotion associations. J. Genet. Psychol. **157**(3), 275–280 (1996)
28. Hendrikx, M., Meijer, S., Van Der Velden, J., Iosup, A.: Procedural content generation for games: a survey. ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) **9**(1), 1–22 (2013)
29. Ip, H.H., et al.: Enhance emotional and social adaptation skills for children with autism spectrum disorder: a virtual reality enabled approach. Comput. Educ. **117**, 1–15 (2018)
30. Jain, A., Mildenhall, B., Barron, J.T., Abbeel, P., Poole, B.: Zero-shot text-guided object generation with dream fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 867–876 (2022)
31. Janssens, O., Samyny, K., Van de Walle, R., Van Hoecke, S.: Educational virtual game scenario generation for serious games. In: 2014 IEEE 3nd International Conference on Serious Games and Applications for Health (SeGAH), pp. 1–8 (2014). https://doi.org/10.1109/SeGAH.2014.7067106
32. Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M.: Neural style transfer: a review. IEEE Trans. Vis. Comput. Graph. **26**(11), 3365–3385 (2019)
33. Li, Y., Bai, L., Mao, Y., Peng, X., Zhang, Z., Tong, X., Ray, L.: The exploration and evaluation of generating affective 360 panoramic VR environments through neural style transfer. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 759–760. IEEE (2023)
34. Liu, J., Snodgrass, S., Khalifa, A., Risi, S., Yannakakis, G.N., Togelius, J.: Deep learning for procedural content generation. Neural Comput. Appl. **33**(1), 19–37 (2021)
35. Lobel, A., Gotsis, M., Reynolds, E., Annetta, M., Engels, R.C., Granic, I.: Designing and utilizing biofeedback games for emotion regulation: The case of nevermind. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 1945–1951 (2016)
36. McGee, R.W.: Annie Chan: Three short stories written with chat GPT. SSRN 4359403 (2023)
37. Mienaltowski, A., Lemerise, E.A., Greer, K., Burke, L.: Age-related differences in emotion matching are limited to low intensity expressions. Aging Neuropsychol. Cogn. **26**(3), 348–366 (2019)
38. Mohammad, S., Kiritchenko, S.: Wikiart emotions: an annotated dataset of emotions evoked by art. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (2018)
39. Morgan, J.K., Izard, C.E., King, K.A.: Construct validity of the emotion matching task: preliminary evidence for convergent and criterion validity of a new emotion knowledge measure for young children. Soc. Dev. **19**(1), 52–70 (2010)
40. OpenAI: ChatGPT: a generative model for conversation. OpenAI Blog (2021). https://openai.com/blog/chat-gpt/

41. Pereira, Y.H., Ueda, R., Galhardi, L.B., Brancher, J.D.: Using procedural content generation for storytelling in a serious game called orange care. In: 2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), pp. 192–197. IEEE (2019)
42. Pirker, J., Dengel, A.: The potential of 360° virtual reality videos and real VR for education–a literature review. IEEE Comput. Graph. Appl. **41**(4), 76–89 (2021)
43. Posner, J., Russell, J.A., Peterson, B.S.: The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and psychopathology **17**(3), 715–734 (2005). https://doi.org/10.1017/S0954579405050340, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2367156/
44. Prinz, J.: Emotion and aesthetic value. In: American Philosophical Association Pacific Meeting, vol. 15. Kluwer Dordrecht (2007)
45. Ramesh, A., et al.: Zero-shot text-to-image generation. arXiv preprint arXiv:2102.12092 (2021)
46. Riva, G., et al.: Affective interactions using virtual reality: the link between presence and emotions. Cyberpsychol. Behav. **10**(1), 45–56 (2007)
47. Ruder, M., Dosovitskiy, A., Brox, T.: Artistic style transfer for videos and spherical images. Int. J. Comput. Vis. **126**(11), 1199–1219 (2018)
48. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6), 1161–1178 (1980). https://doi.org/10.1037/h0077714
49. Sauzéon, H., N'Kaoua, B., Arvind Pala, P., Taillade, M., Guitton, P.: Age and active navigation effects on episodic memory: a virtual reality study. Br. J. Psychol. **107**(1), 72–94 (2016)
50. Schlegel, K., Scherer, K.R.: Introducing a short version of the Geneva emotion recognition test (GERT-S): Psychometric properties and construct validation. Behav. Res. Methods **48**(4), 1383–1392 (2016)
51. Schrader, C., Brich, J., Frommel, J., Riemer, V., Rogers, K.: Rising to the challenge: an emotion-driven approach toward adaptive serious games. In: Serious Games and Edutainment Applications : Volume II, pp. 3–28 (2017). https://doi.org/10.1007/978-3-319-51645-5_1
52. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
53. Smith, G.: Understanding procedural content generation: a design-centric analysis of the role of PCG in games. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 917–926 (2014)
54. Smith, G., Gan, E., Othenin-Girard, A., Whitehead, J.: PCG-based game design: enabling new play experiences through procedural content generation. In: Proceedings of the 2nd International Workshop on Procedural Content Generation in Games, pp. 1–4 (2011)
55. Somarathna, R., Bednarz, T., Mohammadi, G.: Virtual reality for emotion elicitation – a review. IEEE Trans. Affect. Comput., 1–21 (2022). https://doi.org/10.1109/TAFFC.2022.3181053
56. Sullivan, S., Ruffman, T.: Emotion recognition deficits in the elderly. Int. J. Neurosci. **114**(3), 403–432 (2004)
57. Sun, Y., Xu, Y., Cheng, C., Li, Y., Lee, C.H., Asadipour, A.: Travel with wander in the metaverse: an AI chatbot to visit the future earth. In: 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP), pp. 1–6. IEEE (2022)
58. Toet, A., van Welie, M., Houtkamp, J.: Is a dark virtual environment scary? Cyberpsychol. Behav. **12**(4), 363–371 (2009)

59. Troxler, M., Qurashi, S., Tjon, D., Gao, H., Rombout, L.: The virtual hero: the influence of narrative on affect and presence in a VR game. In: CEUR Workshop Proceedings (2018). http://afcai18.webs.upv.es/index.html, affective Computing and Context Awareness in Ambient Intelligence
, AfCAI ; Conference date: 19-04-2018 Through 20-04-2018
60. Watson, D., Tellegen, A.: Toward a consensual structure of mood. Psychol. Bull. **98**(2), 219–235 (1985). https://doi.org/10.1037/0033-2909.98.2.219, place: US
61. Yannakakis, G.N., Togelius, J.: Experience-driven procedural content generation. IEEE Trans. Affect. Comput. **2**(3), 147–161 (2011). https://doi.org/10.1109/T-AFFC.2011.6
62. Zyda, M.: From visual simulation to virtual reality to games. Computer **38**(9), 25–32 (2005)