International Symposium on World Ecological Design
F. Ying et al. (Eds.)

doi:10.3233/FAIA240079

Rhythm Vision: Applying Audio-Visual Isomorphism to Map Anxiety Emotions in College Students

Zhihao HUANG [a,*], Yousheng YAO [a,b,*,1], Yuan XIE [a], Fengping FENG [a], Xiqin PAN [a], E TANG [a], Junpeng ZHENG [a], Junqi ZHAN [a], Jin PENG [a], and Haoteng CHEN [a]

[a] *Hexiangning College of Art and Design, Zhongkai University of Agriculture and Engineering, China*

[b] *Faculty of Humanities and Arts, Macau University of Science and Technology, China*

Abstract. This paper introduces a wearable device named Rhythm Vision designed to assist college students in accurately mapping and identifying emotions during work. To ensure the emotional mapping accuracy of Rhythm Vision, we employed the MAX3012 heart rate and blood oxygen detection sensor for physiological information collection. An audio-visual isomorphic interactive system was designed using MaxMSP software. Rhythm Vision defines emotions as calm, relaxed, excited, and anxious for feedback. Participant experiments were conducted with Rhythm Vision, and the results indicate a high level of acceptance among college students when wearing Rhythm Vision during work, providing beneficial support for emotional well-being.

Keywords. Audio-Visual Isomorphism, Affective Computing, College Student Anxiety Emotions, Sensors

1. Introduction

Anxiety disorders are prevalent psychological health issues among college students, as evidenced by data from the American College Health Association, indicating an increasing number of students reporting excessive anxiety. This trend is observed globally. While research on anxiety disorders in college students is a hot topic, primarily focusing on qualitative aspects and intervention methods, there is limited work addressing the identification of anxiety emotions in college students.

College students may struggle to recognize and comprehend their own anxiety emotions, often misinterpreting them as normal nervousness or stress, thus neglecting the anxiety emotions. Prolonged anxiety can lead to more severe mental health problems such as anxiety disorders and depression. Auditory and visual cues present a potential solution; previous studies suggest that physiological signals are more objectively accurate than subjective self-reports [1], and multimodal expressions can

---

[*] The first two authors contributed equally to the article.

[1] Corresponding author, Yousheng YAO, Hexiangning College of Art and Design, Zhongkai University of Agriculture and Engineering, 24 Dongsha Street, Fangzhi Road, Haizhu District, Guangzhou, China, 510225, E-mail: 120752037@qq.com.

convey emotional information more comprehensively[2]. However, no research has investigated the reflection of anxiety emotions in college students during work.

In this paper, we introduce Rhythm Vision, a wearable device that utilizes a heart rate and blood oxygen concentration monitoring sensor to detect the user's emotional state. It provides feedback through sound and images. Worn during work, Rhythm continuously monitors the user's emotional state and generates music and images at the end of the work session to help users understand their anxiety emotions, facilitating timely intervention.

2. Related Work

2.1. College Student Anxiety Emotions

Previous research has affirmed the significance of managing anxiety emotions among college students, as anxiety is a psychological factor that influences academic performance [3]. Instruments like the Social Anxiety Cognitive Scale for College Students (SACS-CS) have been developed, demonstrating good reliability and effectiveness [4]. Digital interventions are becoming increasingly popular for prevention and treatment [5]. Prior work lays the foundation for further studying anxiety psychology in college students, suggesting that digital means could be effective tools for providing convenient and personalized support and interventions.

2.2. Affective Computing

Affective computing is a multidisciplinary field exploring how technology can provide information for understanding human emotions, how interactions between humans and technology are influenced, and how systems can be designed to enhance capabilities through the understanding of influence. It involves endowing computers with the ability to observe, interpret, and generate features related to effect, similar to human observations, to improve the quality of human-computer communication [6]. Affective computing has practical applications in various fields, including robot systems, avatars in human-computer interaction, e-learning, game characters, and supportive devices. It holds promise in the field of psychotherapy, aiding in increasing our understanding of therapy, training future clinical professionals, and providing care alongside human therapists [7]. Exploring emotions as another model of interaction: dynamic, culturally mediated, socially constructed, and experiential. To achieve the required online assessment of these emotional states, features are extracted from users' physiological signals (blood volume pulse, skin conductance response, skin temperature, and pupil diameter), which can be processed by pattern recognition systems to classify users' emotional states [8]. Using electroencephalogram (EEG) signals to detect emotional states and identify 12 subtle emotions: anger (annoyed, angry, tense), calm (calm, peaceful, relaxed), happy (excited, happy, joyful), and sad (tired, bored, sad) [9]. Heart rate responses can provide an objective assessment of emotional processing [10]. Building on previous work, we further explore the advantages of affective computing in detecting anxiety emotions in college students to help them recognize and take further action.

2.3. Audio-Visual Isomorphism

Sound and images can be combined in various ways. One approach involves interactive media systems that synthesize sound based on the analysis of drawings made by users [11]. Another method studies the Mismatch Negativity (MMN) in audio-visual stimuli demonstrations, where visual context influences incoming auditory information, leading to a mismatch with universally expected auditory memory traces [12]. Additionally, artists strategically use sound or silent images to create awareness and challenge perception [13]. Technological inventions, such as image projection sound, utilize holographic projection technology to provide new audio-visual experiences [14]. Learning machines have also been developed to synchronously combine sound and images for educational purposes[15]. These examples illustrate various ways of integrating sound and images simultaneously.

Research using Event-Related Potentials (ERPs) measurements has revealed that music and chords exhibit emotional connotations in the early stages, approximately within 200 milliseconds. Major chords are associated with positive emotions, while minor chords are associated with negative emotions, emphasizing the importance of music and chords as the smallest musical building blocks for retaining emotional information [16]. Confirming that color stimuli affect observers' emotional states, not only determined by the hue of the color as traditionally assumed but also by all three color dimensions and their interactions [17]. Relevant studies have uncovered the perception and impact of audio-visual isomorphism on emotions.

3. Method

Based on the literature review conducted in this study, a preliminary system design was proposed, as depicted in Figure 1. Heart rate and blood oxygen concentration values were utilized as sources for affective computing, defining emotions as calm, relaxed, excited, and anxious. Audio-visual isomorphism content mapped emotions through musical scales and colors.

To gain a deeper understanding of college students' emotional responses, we recruited a group of college student volunteers for a physiological signal collection experiment. The focus was on assessing heart rate and blood oxygen concentration changes in volunteers under different emotional states, providing a reliable foundation for emotion classification. The experiment employed the MAX3012 sensor for real-time monitoring of heart rate and blood oxygen values. The sensor works by shining two beams of light onto the fingertip or earlobe and measuring the reflected light using a photodetector. This method, known as photoplethysmography, detects pulses through light. The sensor was connected to an Arduino controller, transmitting data via a Bluetooth module to a computer for recording and analysis. This non-invasive monitoring design helped ensure the comfort and natural state of the subjects during the experiment.

We recruited 8 healthy volunteers for testing, ensuring they had no cardiovascular or respiratory diseases and were not taking medications affecting heart rate and blood oxygen. The purpose and procedures of the experiment were explained, ensuring that the volunteers understood the nature of the experiment.

The experiment consisted of four stages, each representing an emotional state. To evoke different emotional states in the subjects, a series of emotion-inducing tasks were designed, including calm, relaxed, excited, and anxious scenarios. Each stage lasted approximately 10-15 minutes. A quiet, peaceful environment was provided, encouraging volunteers to meditate or view natural scenery images to induce calm emotions. Comfortable seating, soft music, and mild lighting were provided to induce relaxed emotions. Rock music or videos and light sports activities were encouraged to induce excited emotions. Tense and challenging board games, incorporating time pressure elements, were provided to induce anxious emotions. Additionally, after each emotional data test, volunteers were interviewed to check if the emotions generated during the test were accurate. If not, the test data were discarded, and the test was repeated.

Monitoring results of the volunteers showed that in a calm state, the average heart rate ranged from 60-80 beats per minute; in a relaxed state, the average heart rate ranged from 55-75 beats per minute; in an excited state, the average heart rate ranged from 70-100 beats per minute; in an anxious state, the average heart rate ranged from 80-120 beats per minute. Regarding blood oxygen concentration, it was 95-100% in calm, relaxed, and excited states, but in an anxious state, it ranged from 92-98%. Some volunteers reported experiencing rapid breathing during the anxious state. It's worth noting that some volunteers were unaware of being in an anxious state, attributing their heightened focus to normal tension.

College students may struggle to recognize and understand their anxiety emotions, often misinterpreting them as normal nervousness or stress, thereby neglecting their anxiety emotions. To address this issue, we propose Rhythm Vision, an interactive ear clip designed for continuous collection of the wearer's physiological signals to reflect their emotions. The device showcases the wearer's emotions through audio-visual isomorphism, with specific musical scales and image content, facilitating the wearer's awareness and understanding of emotions, thereby enabling timely management of negative emotions. Due to potential interference when using a single physiological metric, both heart rate and blood oxygen concentration were chosen for emotion assessment.

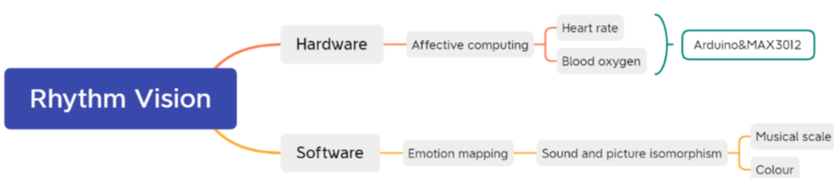
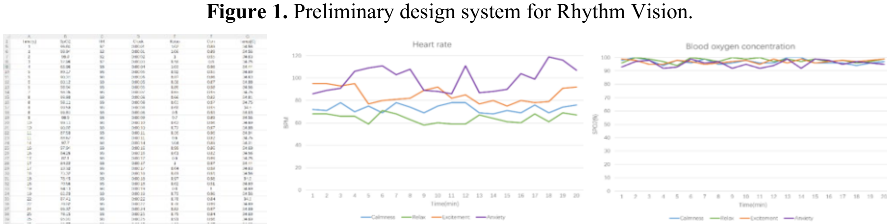


Figure 1. Preliminary design system for Rhythm Vision.



(a) (b) (c)

Figure 2. Physiological signal acquisition experiment of emotion. (a) Detection data collection. (b) The average heart rate of the four emotions. (c) Average blood oxygen concentrations for the four emotions.

Table 1. The average ranges of heart rate and blood oxygen concentration of the four emotions.

Emotion	Heart rate (beats/minute)	Blood oxygen concentration (%)
Calmness	60-80	95-100
Relax	55-75	95-100
Excitement	70-100	95-100
Anxiety	80-120	92-98

4. Rhythm Vision: A Wearable Ear Clip for Emotional Perception

Based on the literature review and the results of the physiological signal collection experiment on emotions, we developed the design system for Rhythm Vision. This system comprises the hardware component for effective computing and the software component for the backend monitoring system, with communication facilitated through Bluetooth.

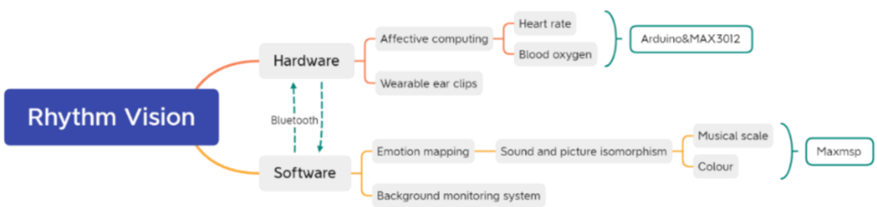


Figure 3. Rhythm Vision design system.

4.1. Hardware

Rhythm Vision is a wearable ear clip equipped with the MAX3012 heart rate and blood oxygen detection sensor. Since college students often rely on typing with their fingers during work, the MAX3012 sensor is positioned on the earlobe. To prevent data interference caused by sweat and skin contact with the sensor panel, an insulating enclosure was created using 3D printing, as shown in Figure 4a. The system utilizes an Arduino Uno as the microcontroller, a 5V 1A lithium battery as the power source, and an HC-05 Bluetooth communication module for wireless communication with the monitoring computer. The module transmits the data collected by the MAX3012 sensor, as illustrated in Figure 4c. All sensors are connected using DuPont wires, as depicted in Figure 4d. A 3D-printed collar clip is used to secure the device to the wearer's collar, with the MAX3012 sensor attached to the ear clip, as shown in Figure 4e. The actual wearing of the device is illustrated in Figure 4f.

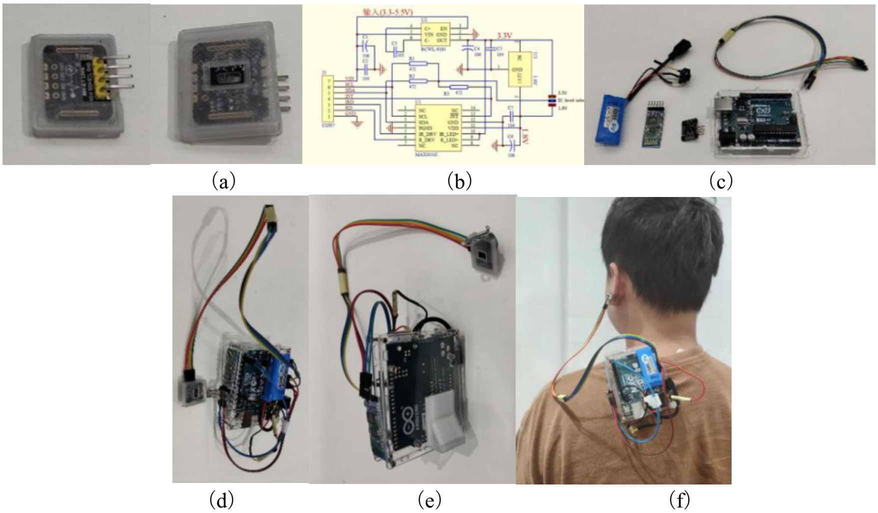


Figure 4. Hardware devices. (a) MAX3012 sensor insulation package. (b) Schematic diagram of MAX3012 sensor. (c) Hardware components. (d) Hardware front. (e) Hardware back. (f) Hardware wear display.

4.2. Software

The sound and picture isomorphic mapping part of Rhythm Vision is taken care of by the background monitoring system, which is deployed on a computer, and the sound and picture isomorphic mapping is realized using Maxmsp software, as shown in Fig. 5. Serial instruction is used in Maxmsp to read the physiological signals captured by the heart rate and blood oxygen detection sensor of the MAX3012 in the hardware part, and to categorize the emotional physiological values in Table 1. The corresponding scales are compiled, music is generated using the make note instruction, and images are generated using jit. gl. Both music and images are generated in real-time, allowing the user to observe the emotions in real-time and to subsequently view the generated content. Based on the foundation of previous research, we propose the interaction mechanism of the system, as shown in Table 2, the scale generated by the calm emotion is C minor, the instrument used is the piano, the timbre tends to be pure acoustic sound, and the generated image is a tiny blue particle spreading in the space with low gravity, at this time the mapped content is bland and graceful; the scale generated by the relaxation emotion is D minor, using the keyboard as the instrument, the timbre tends to be vintage The scale generated by relaxed mood is D minor, using keyboard as the instrument, the tone tends to be retro style, the picture is green small particles attracting each other in the low gravity space, the mapping content is relaxed and harmonious; excited mood is generated by the scale of F major, using synthesizer as the instrument, the timbre tends to be avant-garde with drums and rhythm, the picture is red large particles attracting each other in the high gravity space, the mapping content is focused and energetic; anxious mood is generated by the scale of F minor, using strings as the instrument, the tone tends to be melancholic atmosphere, the generated picture is purple. Anxiety is generated in the key of F minor, using strings as instruments, the tone tends to be melancholic, and the resulting image is a huge example of purple color attracting each other in the high gravitational space, mapping the content is depressing, tense, and dark.

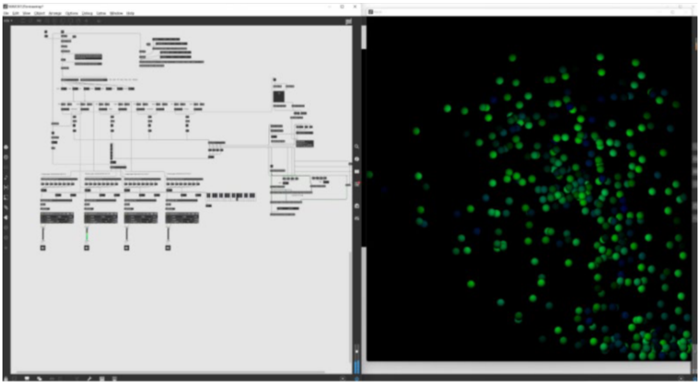
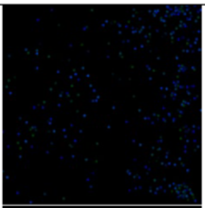
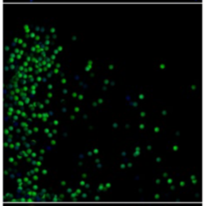
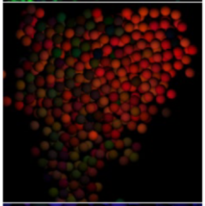
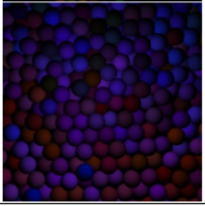


Figure 5. Maxmsp patch interface, emotion mapping visualization window.

Table 2. Audio-visual isomorphic content of emotion mapping.

Emotion	Scale	Musical instruments	Tone	Video	Video content
Calmness	C minor scale	Piano	Pure soundtrack		Blue, Tiny particles, microgravity diffusion
Relax	D minor scale	Keys	Return to the ancients		Green, small particles, microgravity aggregation
Excitement	F major scale	Sequence	Avant-garde rhythm		Red, large particles, large gravity aggregation
Anxiety	F minor scale	Strings	Melancholy		Purple, large particles, large gravity aggregation

5. User Study: Emotion Mapping and Recognition

We conducted a second experiment to investigate whether university students can accurately map and recognize emotions while working.

5.1. Experimental Procedure

We recruited 16 participants from different majors and grades. Participants were provided with detailed informed consent forms explaining the purpose and process of the experiment, as well as potential risks and their rights. Participants were shown how to properly wear the Rhythm Vision device to ensure accurate collection of heart rate and blood oxygen data. The participants were divided into two groups. The first group wore the Rhythm Vision device while completing their tasks without additional assigned tasks. The second group wore the Rhythm Vision device and completed emotion-inducing tasks, including calmness, relaxation, excitement, and anxiety, similar to the tasks in the first experiment. The experiment lasted for 20 minutes, as shown in Figure 6. During the experiment, participants could freely choose whether the emotion feedback from Rhythm Vision was in real-time or played back later. After the experiment, participants were asked to evaluate their perceived emotions using the Likert scale. Follow-up interviews or discussions were conducted to gather feedback on the participants' experience using Rhythm Vision and their thoughts on the experiment.

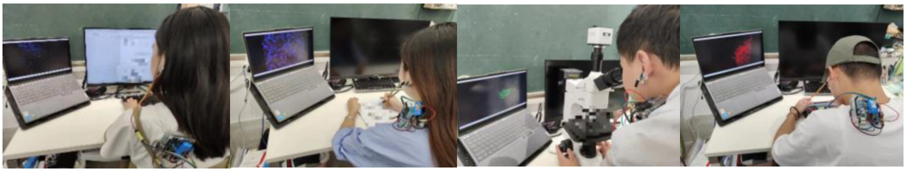


Figure 6. College student participants.

5.2. Results

As shown in Figure 7 (Likert scale), all participants demonstrated a high willingness to use Rhythm Vision. However, they expressed varying opinions on the use of emotion-mapping strategies. The majority of participants indicated that using Rhythm Vision helped them gain a more thorough understanding of their emotions, and real-time feedback during the emotion mapping process was helpful. Additionally, some participants expressed high expectations for future functionalities of Rhythm Vision and provided suggestions. Some participants hoped that Rhythm Vision could offer more diverse emotion intervention modes in the future, such as dealing with anxiety and excitement. Another participant stated, "I hope Rhythm Vision can become a comprehensive emotion management tool in the future, not just limited to calmness, relaxation, excitement, and anxiety." Participants believed that Rhythm Vision allowed them to genuinely experience emotions with meaning. They mentioned that in the past, they might not have been aware of their emotional states, and Rhythm Vision provided them with a deeper understanding.

During the experiment, we also observed some interesting behaviors. Some participants exhibited a reliance on Rhythm Vision, habitually checking it whenever they experienced a certain emotion. Additionally, some participants deliberately chose different emotional states to observe the feedback from Rhythm Vision. This indicates that participants not only used Rhythm Vision to correct accurate emotion mapping but also attempted to understand the impact of different emotions on themselves.

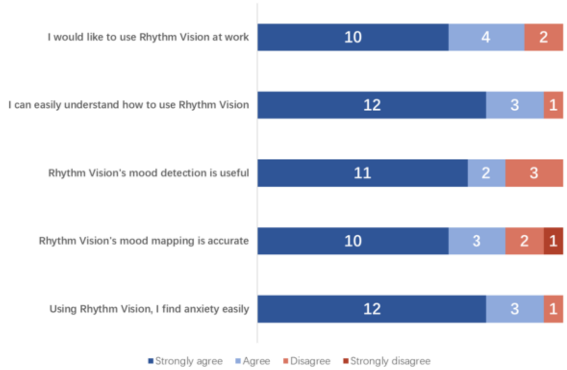


Figure 7. Questionnaire results of participants.

5.3. Discussion

The results show that the feedback of audio-visual isomorphism can accurately reflect the emotions and help people understand the emotions, Rhythm Vision is a good concept, which is useful for college students to find out the emotions in their daily work, especially for anxiety, and it is also found in the participant experiment that college students are easy to misjudge the anxiety and misinterpret it as a result of stress and nervousness, and audio-visual isomorphism can allow college students to understand emotions from multimodal perspectives instead of relying only on their feedback, by using a specific tone scale and specific images can allow college students to understand emotions from a multimodal perspective, rather than relying solely on their feedback. In addition, the experiment also found that participants provided feedback on the need for timely intervention for anxiety, and we will optimize the Rhythm Vision system by adding active parameter interventions to allow users to spontaneously create audio-visual isomorphisms.

6. Conclusion

In this study, we introduced a wearable device called Rhythm Vision, designed to help university students accurately map and recognize emotions while working through audio-visual isomorphism. The results showed that the participants had a positive attitude towards the concept of Rhythm Vision, expressing their favoritism towards its potential effects on emotion mapping. With Rhythm Vision, we conducted a study aimed at improving college students' perception and management of emotions at work. By introducing new audio-visual isomorphic cues, we expect to provide users with a more comprehensive and intuitive emotion mapping experience, which will help them better understand and cope with emotional challenges at work. We believe that this study will provide useful support for improving the quality of work life and emotional well-being of university students.

Acknowledgments

We would like to thank all the participants who took part in this project. This study was supported by the 2021 Guangdong Provincial Quality Engineering Modern Industrial College Project "Eco-design Industrial College", the Guangdong First-class Professional Construction Point: Product Design, the 2021 Guangdong Provincial Department of Science and Technology/Guangdong Provincial Bureau of Rural Revitalization "Guangdong Provincial Rural Science and Technology Specialists in Towns and Villages," stationed in Jiangdong Town, Chaoan District, Chaozhou City, Project No.: KTP20210374, 2021 Zhongkai College of Agricultural Engineering Quality Engineering Special Talent Cultivation Program Construction Project, "Excellence in Rural Revitalization Talent Design and Innovation Class" support.

References

[1] Soman, Karthik, Varghese Alex, and Chaithanya Srinivas. "Analysis of physiological signals in response to stress using ECG and respiratory signals of automobile drivers." *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing* (iMac4s). IEEE, 2013.

[2] Gerdes, Antje BM, et al. "Emotional sounds modulate early neural processing of emotional pictures." *Frontiers in psychology* 4 (2013): 741.

[3] Huang, Mao, and Jiaxin Yang. "Cognitive appraisal and coping styles of college students’ English learning anxiety." *SHS Web of Conferences.* Vol. **171**. EDP Sciences, 2023.

[4] Zha, Yuxin, et al. "Development of social anxiety cognition scale for college students: Basing on Hofmann’s model of social anxiety disorder." *Frontiers in Psychology* 14 (2023): 1080099.

[5] Liu, Xin-Qiao, Yu-Xin Guo, and Yi Xu. "Risk factors and digital interventions for anxiety disorders in college students: Stakeholder perspectives." *World Journal of Clinical Cases* 11.7 (2023): 1442.

[6] Rosalind, W, Picard. Affective Computing.1997.

[7] Jianhua, Tao. “Multimodal Information Processing for Affective Computing.” 2010.

[8] Barreto, Armando, Jing Zhai, and Malek Adjouadi. "Non-intrusive physiological monitoring for automated stress detection in human-computer interaction." *Human‑Computer Interaction: IEEE International Workshop, HCI 2007 Rio de Janeiro, Brazil, October 20, 2007 Proceedings 4*. Springer Berlin Heidelberg, 2007.

[9] Kimmatkar, Nisha Vishnupant, and B. Vijaya Babu. "Novel approach for emotion detection and stabilizing mental state by using machine learning techniques." *Computers* 10.3 (2021): 37.

[10] Nishith, Pallavi, Michael G. Griffin, and Terri L. Weaver. "Utility of the heart rate response as an index of emotional processing in a female rape victim with posttraumatic stress disorder." *Cognitive and Behavioral Practice* 9.4 (2002): 302-307.

[11] Goto, Saori, Nanako Kondo, and Shinji Mizuno. "RAKUGACKY: making sounds with drawing." *ACM SIGGRAPH 2013 Posters.* 2013. 1-1.

[12] Ullsperger, Peter, et al. "When sound and picture do not fit: Mismatch negativity and sensory interaction." *International journal of psychophysiology* 59.1 (2006): 3-7.

[13] Thoma, Andrea. "Discussing the relationship between sound and image: The scope of audio/visual strategies in a contemporary art context." *The International Journal of New Media, Technology and the Arts* 10.3 (2015): 17-28.

[14] Zhang, Xin. Picture projection sound. 2018.

[15] Wang, Yeci.Sound and picture comparison learning machine. 2005.

[16] Bakker, David Radford, and Frances Heritage Martin. "Musical chords and emotion: Major and minor triads are processed for emotion." *Cognitive, Affective, & Behavioral Neuroscience* 15 (2015): 15-31.

[17] Wilms, Lisa, and Daniel Oberfeld. "Color and emotion: effects of hue, saturation, and brightness." *Psychological research* 82.5 (2018): 896-914.