Title
The Sound of Suggestion: Priming Emotional Interpretation with Non-Linguistic Audio Cues

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Han, Xizhen
Zhang, Yifei
Chen, Jie
et al.

Publication Date
2026

The Sound of Suggestion: Priming Emotional Interpretation with Non-Linguistic Audio Cues

Xizhen Han[1,2], Yifei Zhang[1,3], Jie Chen[4], Rongrong Chen[1,5]* (rainerrchen@bnbu.edu.cn)

1. Applied Psychology Programme, Beijing Normal-Hong Kong Baptist University, Zhuhai, China
2. Division of Psychology and Language Sciences, University College London, London, United Kingdom
3. Applied Social Sciences, Hong Kong Polytechnic University, Hong Kong SAR, China
4. Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China
5. Guangdong Provincial/Zhuhai Key Laboratory IRADS, Zhuhai, China

Abstract

Recent research in affective computing has shown growing interest in how multisensory integration and metaphorical expression shape emotional experience. However, the specific role of non-linguistic sounds in shaping affective information processing remains underexamined. Here we engaged thirty participants in an adapted Thematic Apperception Test featuring affectively ambiguous black-and-white line drawings. We exposed them to 300 Hz or 800 Hz quasi-white noise while selecting a color for a central figure and an emotional interpretation for the scene. The 800 Hz sound elicited significantly more positive interpretations and warm color choices, whereas the 300 Hz sound elicited negative interpretations and cool color choices. These findings show that frequency-specific auditory cues act as affective signals biasing cross-modal meaning construction. Consistent with evolutionary accounts of frequency-based emotional coding, they support using non-linguistic sound as an implicit emotional prime, offering insights for art therapy and multisensory design.

Keywords: Cross-modal integration, Auditory affective priming, Conceptual metaphor, Frequency code theory

Introduction

Hearing a sudden floorboard creak while seeing a shadow shift can instantly create a feeling of dread, whereas hearing wind chimes with the same shadow creates a sense of calm, demonstrating how the auditory cue disambiguates an ambiguous sight and dictates its emotional meaning. Recent work in affective computing has become increasingly captivated by how the multisensory integration and metaphorical expression shape emotional experiences. For instance, researchers demonstrated that actively exploring metaphorical associations across color, shape, and semantics can promote creative association and emotional expression (Kang et al., 2021). Co-creating metaphorical visual narratives with generative AI supports users' self-reflection and meaning making from emotional experiences (Wan et al., 2024). These interdisciplinary studies highlight the potential of cross-modal integration for eliciting emotion and stimulating creativity.

According to the *Conceptual Metaphor Theory*, abstract emotional concepts are often grounded in embodied sensory experiences (Lakoff & Johnson, 1980). Empirical studies suggest that perceiving different colors triggers distinct physiological reactions, which in turn shape emotional experience and cognitive appraisal of the external environment (Goldstein, 1942). These effects appear to transcend specific cultural contexts: for example, *red* evokes *high arousal*, whereas *blue* conveys *calmness* (Elliot & Maier, 2014). Moreover, among Westerners in a musical context, *high pitch* is more strongly associated with *high spatial elevation* and *brightness*, whereas *low pitch* is more strongly associated with *darkness* and *winter* (Eitan & Timmers, 2010). It is worth noting, however, that despite the pitch–elevation mappings, the word *wise* and *pleasant* are significantly associated with high spatial elevation but not with high pitch.

In alignment with the *Conceptual Metaphor Theory*, evolutionary psychologists suggest that the human emotional perception of sound frequencies may originate from an evolutionary legacy of “frequency codes”: high-frequency sounds, such as a baby’s cry, signal vulnerability and elicit protective instincts, while low-frequency sounds, such as the growl of a predator, convey threat and trigger defensive responses (Morton, 1977). These hardwired associations may unconsciously extend to and shape an individual’s emotional interpretation of visual scenes.

While previous empirical research has investigated the conceptual metaphorical framework using complex musical excerpts, how meaningless, non-linguistic auditory features (e.g., pure frequency) exert latent effects on cognitive operations involved in interpreting ambiguous visual narratives remains unclear. To address this gap, we adapted a Thematic Apperception Test paradigm and asked the following research questions:

RQ1. How does a non-linguistic auditory frequency influence viewers' color assignments for a character in an ambiguous visual narrative?

RQ2. How does this auditory frequency shape viewers' affective interpretation of the narrative, and do these affective interpretations follow systematic patterns in color assignment?

Conceptual Metaphor Theory

In 1980, linguists George Lakoff and philosopher Mark Johnson proposed the *Conceptual Metaphor Theory*, a framework suggesting that human understanding of abstract concepts, such as emotion, time, and morality, relies on systematic mappings from specific spatial and

* Xizhen Han and Yifei Zhang contributed equally to this work. Rongrong Chen is the corresponding author.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

sensory-motor experience domains to the abstract concept domain (1980). This theory emphasizes the embodied nature of cognition, positing that conceptual thought is deeply rooted in bodily interactions with the environment. For instance, Meier and Robinson examined how linguistic expressions of abstract ideas are structured using spatial terms, e.g., positive states as *up* or *high*, and negative states as *down* or *low* (2004). As the field evolved, conceptual metaphor theory expanded beyond language into multimodal and cross-sensory research, now serving as a fundamental framework for exploring how perception, emotion, and reasoning interact across sensory domains (Pitt & Casasanto, 2022). For example, low-frequency music is often associated with negative or neutral emotions such as *sadness* or *calmness* among Japanese listeners, whereas high-frequency music tends to elicit positive emotions such as *excitement* or *happiness* (Balkwill et al., 2004). Moreover, ascending pitch induced an illusory positive bias in interpreting subsequent facial emotion, whereas descending pitch induced a negative bias (Zhang et al., 2025). Color also shares a metaphorical link with emotion, evoking analogous perceptual and affective responses (Shi, 2021; Wilms & Oberfeld, 2018). Green, for instance, calls to mind lush forests which are often associated with *vitality, harmony, resilience* and fresh sprouts which typically symbolize *hope* and *new beginnings*. These associations convey latent emotional meanings such as *energy*, *safety*, *growth*, and *optimism*.

Regarding the sound-color correspondence, early literature shows that children tend to associate mid-to-high frequency sounds with warm hues such as *orange* and *yellow* and low-frequency sounds to cool hues such as *blue* and *purple* (Simpson et al., 1956). Studies using complex auditory stimuli (musical excerpts) found that *high-pitched*, *happy* music led participants to select *yellow* and *red* more often, while *low-pitched*, *sad* music elicited a preference for *dark blue* (Barbiere et al., 2007; Lindborg & Friberg, 2015). Moreover, music-color correspondences are found to be emotionally mediated, reinforcing the role of affective experience in abstract cross-modal mapping (Lin et al., 2024; Palmer et al., 2013; Valdés-Alemán et al., 2022; Whiteford et al., 2018).

Frequency Code Theory

Rooted in evolutionary theory, the Frequency Code Hypothesis suggests that animals communicate information about body size and intention through vocal pitch (Morton, 1977). By analyzing animal vocalizations, Morton observed that high-frequency sounds (e.g., birdsong) are typically produced by smaller animals and often signal non-threatening intentions such as *courtship* or *submission*. In contrast, low-frequency sounds (e.g., the growls of large predators) commonly serve to *intimidate* or *assert dominance*. This innate association between frequency, physical size, and intention appears to persist in humans, forming a biological basis for emotional bias in response to pitch. Similarly, cross-linguistic research demonstrates that high pitch is frequently associated with *compliance* or *friendliness*, while low pitch conveys *authority* or *threat* (Ohala, 1984). These evolutionary legacies may continue to shape modern humans' emotional interpretation of non-linguistic sounds. For instance, exposure to low-frequency sounds may make individuals more inclined to perceive visual stimuli as powerful or dangerous. Nevertheless, empirical validation of this hypothesis in cross-modal contexts remains limited, especially regarding how auditory cues and visual information interact to prime emotional interpretation.

Methodology

Participants

Thirty college students (11 males and 19 females, mean age 21.4 years, SD = 0.75) responded to our online recruitment post and participated in this study. All participants had normal hearing and color vision (confirmed via two questions from the Color Blindness Test Map, 5th Edition)(Yu, 2012). This study adheres to the Institutional Ethical Review Guidelines, and all participants provided written informed consent prior to participation.

Experimental Design

We adapted a Thematic Apperception Test paradigm using emotionally ambiguous black-and-white line drawings designed to trigger unconscious projection of internal emotional, motivational, and cognitive states (Frank, 1939). To investigate the auditory affective priming effect, we instructed participants to complete two tasks: a *Figure Coloring Task* and an *Emotion Interpretation Task*, while being exposed to 300 Hz or 800 Hz quasi-white noise at 80 dB. The experimental sessions were conducted on two separate days, counterbalanced and spaced at least five days apart, to mitigate potential carryover or primacy effects from initial stimulus exposure (see Figure 1).

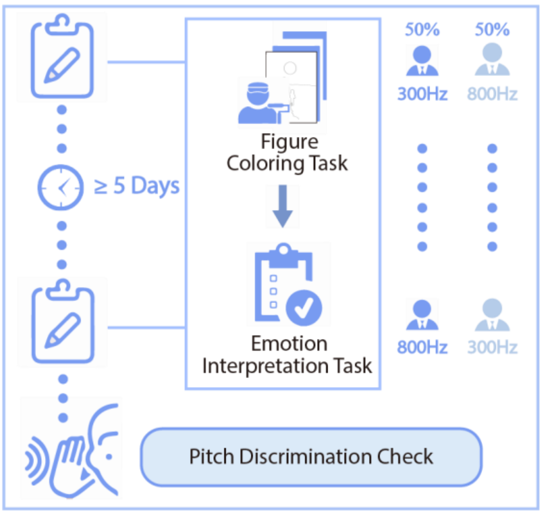


Figure 1. Illustration of experimental design.

To accommodate substantial individual differences in color and emotional perception, we implemented a within-subject design. At the end of the experiment, we administered a *Pitch Discrimination Check* to ensure that our participants had normal pitch perception.

Figure Coloring Task

In this task, we asked participants to color the characters based on their intuitive interpretation of the situation depicted in the visual scenarios and their imagination of the characters' emotional states (see Figure 2). Unlike the fully abstract patterns used in the classic Rorschach inkblot test, which often yield highly divergent interpretations that are difficult to anchor to specific situational or emotional contexts (Frank, 1939), our black-and-white line drawings depicted one or two characters with integrated narrative cues. They retained basic contextual cues while minimizing explicit emotional content, thereby ensuring comprehensibility and providing ecologically valid stimuli for assessing unconscious emotional projection. The stimulus generation process comprised three stages:

Stage 1. Stimulus Selection: We sourced initial images from public databases under Creative Commons licenses, including both AI-generated and hand-drawn ink works. Selection followed strict criteria: neutrally expressed characters, minimal background elements, and moderate narrative tension.

Stage 2. De-emotionalization Processing: We processed the selected images using Adobe Illustrator and Photoshop to minimize inherent emotional bias. This involved desaturating the images, vectorizing their contours, and removing or attenuating emotionally suggestive decorative elements, such as deleting the pipe held by a character in Scenario 1.

Stage 3. Comprehensibility and Neutrality Validation: Six college students who did not participate in the main experiment evaluated the processed images for situational comprehensibility and emotional neutrality. Any image that received two or more negative ratings was replaced with a new candidate, ensuring that the final stimulus set remained both interpretable and emotionally ambiguous.

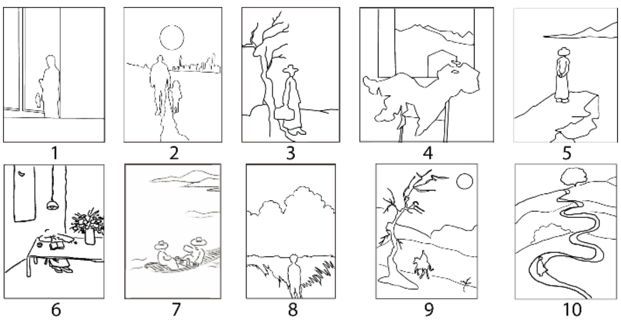


Figure 2. Ten scenarios in the Figure Coloring Task.

To ensure participants had adequate time to interpret each scenario and to encourage deliberate rather than random selection, they were required to hand-color the characters using six Kuelox Super Soft Oil Pastels (see Figure 3): purple, blue, green, yellow, orange, and red.

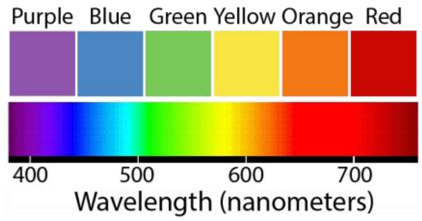


Figure 3. Depiction of the six color options selected from the human visible spectrum.

Pilot testing showed that pure sinusoidal waves, especially at higher frequencies, induced dizziness and discomfort. Quasi-white noise at 300 Hz or 800 Hz mitigated these issues, providing a more comfortable and acoustically stable environment (Awada et al., 2022; Genc et al., 2024). To ensure ecological validity and minimize bias in emotional responses, we therefore used quasi-white noise (centered at either 300 Hz or 800 Hz, 80 dB) instead of pure tones.

Emotion Interpretation Task

For each of the ten scenarios, participants chose which of the two emotional-state descriptions (see Appendix I) best matched their interpretation of the character from the Figure Coloring Task. If neither fit, they could provide their own interpretation.

To ensure objectivity and consistency in the emotional valence of the description options, we employed *DeepSeek* to generate two sentences for each image: one conveying a positive emotion and the other a negative emotion. Using standardized prompts, we instructed the AI to adhere to the following requirements:

1) Both sentences should be narrative in style, incorporating imaginative story elements centered on the characters in the images.
2) Each sentence should reflect a distinct emotional tone, one positive and one negative.
3) Each description should contain no more than 30 words and be expressed in clear layman language.
4) Sentences should be simple declarative statements, without emphasis on literary quality.

Following the initial generation of descriptive content for each scenario, two authors independently evaluated all AI-generated sentences for emotional clarity, linguistic coherence, and contextual relevance. Disagreements or unsatisfactory ratings were resolved through discussion or by regenerating the descriptions. A pilot study (n = 6) confirmed that layperson could reliably distinguish between the intended positive and negative emotional options.

Pitch Discrimination Check

At the end of the experiment, participants were re-exposed to both auditory stimuli for ten seconds and asked to identify which had a higher pitch. Correct identification confirmed basic conscious frequency discrimination ability. This step was essential to verify explicit perception of the frequency difference, given the fundamental distinction between cognitive pathways for explicit versus implicit processing (Dehaene et al., 2006; Mizrahi & Axelrod, 2023).

Data Analysis

To standardize data, we developed dedicated coding systems for both the Figure Coloring and the Emotion Interpretation Task.

Color selection coding: The six colors available for the Figure Coloring Task correspond to spectral positions within the visible light range (380–750 nm; see Figure 3). We assigned a numerical value to each color based on increasing wavelength: Purple = 1 (380–450 nm), Blue = 2 (450–485 nm), Green = 3 (492–565 nm), Yellow = 4 (565–590 nm), Orange = 5 (590–625 nm), and Red = 6 (625–750 nm) (Best, 2012). We classified red, orange, and yellow as warm colors, and green, blue, and purple as cool colors, following conventional color theory.

Emotional valence coding: Drawing on Russell's classic Emotion Model, we implemented a three-level quantitative coding system to evaluate emotional valence in participants' interpretations (Russell, 1980). We coded positive emotional statements (e.g., pleasure, relaxation) as 1, neutral statements (e.g., no feeling, no clear tendency) as 0, and negative emotional statements (e.g., anxiety, disgust) as –1. For open-text responses, we applied a double-blind coding strategy: two independent coders annotated the data, and a third researcher arbitrated any discrepancies

Apparatus

We presented sound stimuli using Edifier W820NB over-ear Bluetooth headphones. These closed-back headphones feature a frequency response range of 20 Hz to 20 kHz. We enabled the hybrid active noise cancellation (ANC) function, which reduces ambient noise by up to 38 dB, to minimize acoustic interference in non-soundproof environments.

Results

After the main experiment, all participants correctly distinguished between the 300 Hz and 800 Hz sound signals in the Pitch Discrimination Check (100% accuracy), confirming normal pitch discrimination ability. Crucially, however, post-experiment debriefing revealed that none of the participants had consciously noticed the frequency differences during the Figure Coloring Task or the Emotion Interpretation Task. Thus, although the frequency difference was readily detectable, it operated implicitly under incidental listening conditions. We therefore retained all 30 datasets for analysis.

For each scenario, we aggregated responses from the 30 participants into a composite profile of mean color assignment and emotional rating. We then conducted paired sample t-tests for each scenario to compare responses between the two frequency conditions, thereby preserving stimulus-specific variability while effectively examining sound-induced cognitive and affective effects.

Color Selection for the Character

When presented with the same scenario, 65.3% of participants in the 800 Hz condition assigned warm colors (red, orange, yellow) to the character in the visual narrative, compared to 45.7% in the 300 Hz condition (see Figure 4). The color selection scores were also significantly higher in the 800 Hz condition (Mean ± SD: 3.59 ± 0.57) than in the 300 Hz condition (3.03 ± 0.43, $t(29) = 3.44$, $p = .002$, 95% CI [-2.13, -.54]). This indicates that auditory frequencies around 800 Hz can induce a preference for warm colors. These results address **RQ1**, demonstrating that non-linguistic sound signals can affect color assignment for the main characters in an ambiguous scene, implying potentially different interpretations of the scene.

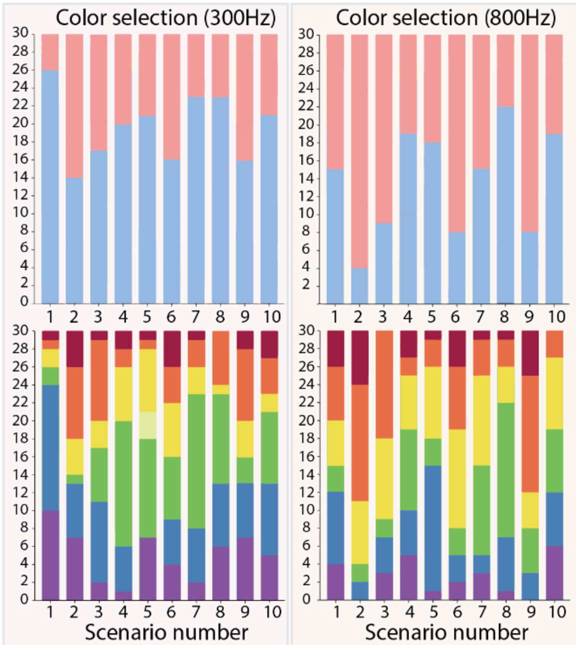


Figure 4. Illustration of colour selection for main characters across two frequency conditions. Upper panel: aggregated warm (red, orange, yellow) and cool (green, blue, purple) color families. Lower panel: distribution of individual color assignment.

Emotional Interpretation of the Scenario

Under the 300 Hz condition, 70% of pictures were interpreted negatively, whereas under the 800 Hz condition, 90% of pictures were interpreted positively (see Figure 5a). This differential effect was further reflected in the affective interpretation ratings. For the same visual scenarios, ratings were significantly higher in the 800 Hz condition (Mean ± SD: 0.36 ± 0.37) than in the 300 Hz condition (-0.1 ± 0.38; $t(29) = 4.42$, $p < .001$, 95% CI [-1.07, -0.40]), indicating a more positive affective bias under the higher frequency. This auditory affective priming effect was significant in six out of the ten scenarios (60%, see Figure 5b). These results suggest that medium-to-high frequencies (e.g., 800 Hz) play a crucial role in shaping emotional interpretation, promoting more positive comprehensions of visual narratives. This pattern addresses RQ2, confirming that non-linguistic auditory frequencies can steer a viewer's affective interpretation. Furthermore, this positive bias under 800 Hz was linked with the preference for warm colors assigned to characters under the same condition.

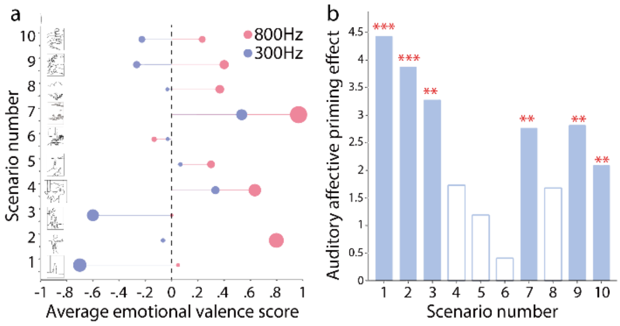


Figure 5. (a) Average emotional valence ratings for the ten visual scenarios. Data points (red: 800 Hz; blue: 300 Hz) represent means averaged across 30 participants. (b) Non-linguistic frequency signals led to significant discrepancies in affective interpretation in six out of ten scenarios. **: p <.01, ***: p<.001.

Discussion

Effect of Sound on Color Coding

The study found that audio frequency significantly influenced color selection. At 800 Hz, participants tended to choose warm colors such as yellow and red for the character(s) in the ambiguous scenarios, whereas at 300 Hz, cool colors like blue and purple were preferred. This finding is consistent with prior music research suggesting that emotion associations mediate the cross-modal correspondence between pitch and color (Lindborg & Friberg, 2015; Palmer et al., 2013). For instance, higher-pitched music conveying joy has been linked to preferences for *yellow* and *red*, whereas lower-pitched music conveying sadness has been associated with preferences for *dark blue* (Barbiere et al., 2007; Lindborg & Friberg, 2015). It is worth noting that most previous studies have documented sound-color correspondence effects using complex musical stimuli and explicit color naming or pointing tasks. Our paradigm demonstrates that non-linguistic quasi-white noise alone can elicit such responses within a projective coloring framework. These results indicate that basic sound frequency provides a sufficient cue to implicitly bias color choices, suggesting that neither melody nor harmony is a prerequisite to initiate cross-modal affective priming.

According to Conceptual Metaphor Theory (Lakoff & Johnson, 1980; Wilms & Oberfeld, 2018), and later extended by Color-in-context Theory (Elliot & Maier, 2012) and Concretization Simulation Theory (Gallese, 2012), participants unconsciously simulate character actions and situations using abstract meanings conveyed through color (e.g., green symbolizing hope). This process activates emotional memories and facilitates affective projection onto the characters in the drawings. In other words, under different sound-frequency conditions, it is quite plausible that participants resonate with the narrative using color as an implicit emotional cue. Specifically, high-frequency sounds evoke warm color tones that elicit positive emotions, whereas low-frequency sounds lead to cool color tones that trigger negative emotions.

Effect of Sound on Emotional Interpretation

Our findings indicate that individuals exposed to an 800 Hz sound tended to interpret the ambiguous visual scenarios more positively, whereas those exposed to a 300 Hz sound were more inclined to perceive negative contents. This observation aligns with Conceptual Metaphor Theory (Lakoff & Johnson, 1980), which posits that higher frequencies are associated with positive emotions and lower frequencies with negative ones. The results are also supported by evolutionary psychology perspectives. Humans appear to have an innate tendency to interpret high-pitched sounds as non-threatening and low-pitched sounds as potential threats (Morton, 1977). Additionally, neuroimaging evidence reinforces this connection: functional magnetic resonance imaging (fMRI) research has shown that non-verbal high-frequency sounds activate brain regions linked to positive affect, such as the ventral striatum and orbitofrontal cortex, whereas non-verbal low-frequency sounds engage areas associated with negative emotions, such as the amygdala and insula (Symons et al., 2016). These neural mechanisms provide a biological basis for the emotional responses elicited by different sound frequencies.

Although our findings align with Conceptual Metaphor Theory and evolutionary frequency code accounts, alternative explanations merit consideration. First, the observed effects may partly reflect differences in arousal and dominance rather than valence alone, and future studies should include more comprehensive measurements of emotion (Cowen & Keltner, 2021). Second, while strong synesthetic experiences are rare, weak synesthetic associations are common in the general population. Synesthesia, where stimulation of one sensory modality leads to involuntary experiences in another (e.g., sound triggering visual associations), can enhance artistic expressiveness and emotional depth (Simpson et al., 1956; Ward, 2013). Such cross-sensory integration may facilitate richer aesthetic experiences and more nuanced emotional interpretations through the blending of multimodal cues.

Exceptions to the Priming Effect

Despite the overall consistency, several visual scenarios did not exhibit the auditory affective priming effect. For example, Scenario 6, which depicts a student studying late at night with their head on a desk, yielded no significant difference in content interpretation across the two frequency conditions. This scenario is strongly associated with negative experiences such as fatigue, anxiety, and sleep deprivation, particularly among university students. Visual elements like the desk lamp, books, and moon likely trigger potent situational memories, promoting participants to recall their own stressful experiences of compulsory late-night studying. In such cases, the latent emotional associations evoked by the image itself appear to override the influence of sound frequency as a secondary stimulus.

Age may also contribute to these exceptions. Previous studies suggest that, due to the ongoing development of the nervous and endocrine systems in younger individuals, neural coordination between cortical and subcortical regions tends to be less balanced (Bailen et al., 2019; Silvers, 2022). Consequently, adolescents and young adults may exhibit stronger, more immediate emotional responses than middle-aged and older adults, who often display more context-independent and restrained emotional associations. Moreover, the typically heightened sensitivity to sensory stimulation in younger individuals may contribute to more intense emotional connections to visual and auditory stimuli, potentially making their responses more difficult to modulate in highly relevant scenarios.

Implications

Our findings offer actionable insights for leveraging cross-modal resonance in applied settings. First, this study advances our understanding of how non-linguistic auditory cues shape unconscious emotional projection and cognitive appraisal during engagement with ambiguous visual narratives, offering actionable strategies for multimodal art therapy interventions. For example, by integrating the Expressive Therapies Continuum (ETC) framework (Hinz, 2019), therapists can incorporate high-frequency sounds (e.g., 800 Hz) during symbolic painting or collage activities to facilitate self-exploration in a positive direction. This auditory priming encourages clients to express positive emotions through warmer color choices, such as selecting golden yellow during moments of emotional safety, while deepening engagement with internal affective states.

Furthermore, this study provides practical guidelines for industrial and interaction design, grounded in the Multisensory Design (MSD) approach, which emphasizes aligning multisensory inputs with innate human perceptual tendencies through structured sensory exploration and analysis to create cohesive, emotionally resonant user experiences (Schifferstein, 2011). Specifically, pairing specific frequencies with colors can evoke targeted emotional states and enhance user satisfaction. For example, combining high-frequency sounds (800 Hz) with green hues (wavelengths ≈ 500–570 nm) may amplify positive cognitive information processing. This suggests tangible applications in interface design, such as assigning a high-frequency alert tone alongside a green "Send" button in messaging apps to reinforce positive user feedback.

These insights also extend to existing affective interaction systems. Our findings position non-linguistic sound frequency as a potentially significant design parameter that can work alongside linguistic modalities such as poetic generation (Rajcic & McCormack, 2020), enabling more nuanced and universally accessible emotional modulation. Integrating these findings could facilitate the development of closed-loop, emotion-responsive interfaces capable of dynamically adapting color schemes and auditory feedback in response to real-time user affect detected through facial expression or vocal analysis. For instance, during episodes of negative affect, the system could transition the interface to warm colors (e.g., orange or red), introduce high-frequency (800 Hz) auditory cues, and present encouraging messages; during states of overstimulation or high arousal, it could shift to cool colors (e.g., blue or green) coupled with low-frequency (300 Hz) tones to promote calmness. Such a system would establish a deeply integrated emotion–sound–color loop, offering new possibilities for adaptive emotional regulation in interactive digital environments.

Limitations and Future Directions

This study has several limitations. First, given that age, educational background, and cultural context may significantly influence cross-sensory emotional associations, the homogeneous participant group in this study limits the generalizability of our findings. For instance, prior research has documented cultural variations in color-emotion mappings: *red* symbolizes *luck* in Chinese cultures but *danger* in many Western contexts (Xu, 2024). Second, the use of a three-point coding system (positive/neutral/negative) to classify emotional valence may have oversimplified the nuanced spectrum of human emotions. Subtle or mixed affective states, such as nostalgia, anticipation, or bittersweetness, are not captured under this restrictive framework, and their potential interactions with sound-color associations remain unexplored (Cowen et al., 2019; Cowen & Keltner, 2021; Keltner et al., 2023). These limitations underscore the need for future studies to incorporate more diverse, cross-cultural samples and employ finer-grained emotional assessment tools. Such efforts would help advance multisensory emotion processing research as a method for investigating complex, culturally situated affective experiences.

Conclusion

This study demonstrates that non-linguistic auditory signals influence how people perceive ambiguous visual scenes, as reflected in their color assignments to the main characters and emotional interpretations. Specifically, 800 Hz sounds prompted a preference for warm colors and more positive emotional interpretations, whereas 300 Hz sounds elicited cooler color choices and more negative interpretations. The observed cross-modal associations align with Conceptual Metaphor Theory and evolutionary accounts of frequency-based emotional coding. The findings underscore the role of non-linguistic sound as an implicit emotional prime and offer practical insights for art therapy and multisensory design.

Acknowledgement

Our work is supported by Guangdong Provincial Key Laboratory IRADS (2022B1212010006) and Guangdong Basic and Applied Basic Research Foundation (UICR0600025, UICR0300017). We also extend our gratitude to four anonymous reviewers for their insightful comments and useful suggestions.

References

Awada, M., Becerik-Gerber, B., Lucas, G., & Roll, S. (2022). Cognitive performance, creativity and stress levels of neurotypical young adults under different white noise levels. *Scientific Reports*, *12*(1), 14566. https://doi.org/10.1038/s41598-022-18862-w

Bailen, N. H., Green, L. M., & Thompson, R. J. (2019). Understanding emotion in adolescents: A review of emotional frequency, intensity, instability, and clarity. *Emotion Review*, *11*(1), 63–73. https://doi.org/10.1177/1754073918768878

Balkwill, L. L., Thompson, W. F., & Matsunaga, R. I. E. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research*, *46*(4), 337–349. https://doi.org/10.1111/j.1468-5584.2004.00265.x

Barbiere, J. M., Vidal, A., & Zellner, D. A. J. E. S. o. t. A. (2007). The color of music: Correspondence through emotion. *Empirical Studies of the Arts*, *25*, 193 – 208.

Best, J. (2012). *Colour Design: Theories and Application* (J. Best, Ed.). Woodhead Publishing Ltd.

Cowen, A., Sauter, D., Tracy, J. L., & Keltner, D. (2019). Mapping the passions: Toward a high-dimensional taxonomy of emotional experience and expression. *Psychological Science in the Public Interest*, *20*(1), 69–90. https://doi.org/10.1177/1529100619850176

Cowen, A. S., & Keltner, D. (2021). Semantic space theory: A computational approach to emotion. *Trends in Cognitive Sciences*, *25*(2), 124–136. https://doi.org/10.1016/j.tics.2020.11.004

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211. https://doi.org/10.1016/j.tics.2006.03.007

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: cross-domain mappings of auditory pitch in a musical context. *Cognition*, *114*(3), 405–422. https://doi.org/10.1016/j.cognition.2009.10.013

Elliot, A., & Maier, M. (2012). Color-in-context theory. *Advances in Experimental Social Psychology*, *45*, 61–125. https://doi.org/10.1016/B978-0-12-394286-9.00002-0

Elliot, A. J., & Maier, M. A. (2014). Color psychology: Effects of perceiving color on psychological functioning in humans. *Annual Review of Psychology*, *65*(1), 95–120. https://doi.org/10.1146/annurev-psych-010213-115035

Frank, L. K. (1939). Projective methods for the study of personality. *Journal of Psychology: Interdisciplinary and Applied*, *8*, 389–413. https://doi.org/10.1080/00223980.1939.9917671

Gallese, V. a. G., M. . (2012). Embodying movies: Embodied simulation and film studies. *Cinema: Journal of Philosophy and the Moving Image*, *3*, 183–210.

Genc, H., Tanriverdi, S., & Sait Akar, M. (2024). The effect of virtual rainforest and a white noise mobile application on satisfaction, tolerance, comfort, and vital signs during arthroscopy. *Complementary Therapies in Clinical Practice*, *55*, 101848. https://doi.org/10.1016/j.ctcp.2024.101848

Goldstein, K. (1942). Some experimental observations concerning the influence of colors on the function of colors on the function of the organism. *Occupational Therapy*, *21*, 147–151.

Hinz, L. D. (2019). *Expressive Therapies Continuum: A Framework for Using Art in Therapy* (2nd Edition ed.). Routledge. https://doi.org/10.4324/9780429299339

Kang, Y., Sun, Z., Wang, S., Huang, Z., Wu, Z., & Ma, X. (2021). MetaMap: Supporting visual metaphor ideation through multi-dimensional example-based exploration. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Article 427, 421–415. https://doi.org/10.1145/3411764.3445325

Keltner, D., Brooks, J. A., & Cowen, A. (2023). Semantic space theory: Data-driven insights into basic emotions. *Current Directions in Psychological Science*, *32*(3), 242–249. https://doi.org/10.1177/09637214221150511

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, *4*(2), 195–208. https://doi.org/10.1207/s15516709cog0402_4

Lin, K., Zhang, D., & Chen, R. (2024). The emotional bridge: Exploring the association between color and western and Chinese music. *Proceedings of the 17th International Symposium on Visual Information Communication and Interaction (VINCI'24)*, 1–5. https://doi.org/10.1145/3678698.3687189

Lindborg, P., & Friberg, A. K. (2015). Colour association with music is mediated by emotion: Evidence from an experiment using a CIE Lab interface and interviews. *PLoS One*, *10*(12), e0144013.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science*, *15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Mizrahi, T., & Axelrod, V. (2023). Naturalistic auditory stimuli with fNIRS prefrontal cortex imaging: A potential paradigm for disorder of consciousness diagnostics (a study with healthy participants). *Neuropsychologia*, *187*, 108604. https://doi.org/10.1016/j.neuropsychologia.2023.108604

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, *111*, 855 – 869. https://doi.org/10.1086/283219

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F_0 of voice. *Phonetica*, *41*(1), 1–16. https://doi.org/doi:10.1159/000261706

Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-Leon, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences USA*, *110*(22), 8836–8841. https://doi.org/10.1073/pnas.1212562110

Pitt, B., & Casasanto, D. (2022). Spatial metaphors and the design of everyday things. *Frontiers in Psychology*, *13*, 1019957. https://doi.org/10.3389/fpsyg.2022.1019957

Rajcic, N., & McCormack, J. (2020). Mirror ritual: An affective interface for emotional self-reflection. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376625

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178. https://doi.org/10.1037/h0077714

Schifferstein, R. (2011). Multi sensory design. *Proceedings of the 2nd Conference on Creativity and Innovation in Design (DESIRE '11)*, 361–362. https://doi.org/10.1145/2079216.2079270

Shi, S. Y. (2021). Review and prospect of research on the relationship between color and emotion. *Advances in Social Sciences*, *10*(05), 1224–1230. https://doi.org/10.12677/ass.2021.105166

Silvers, J. A. (2022). Adolescence as a pivotal period for emotion regulation development. *Current Opinion in Psychology*, *44*, 258–263. https://doi.org/10.1016/j.copsyc.2021.09.023

Simpson, R. H., Quinn, M., & Ausubel, D. P. (1956). Synesthesia in children: association of colors with pure tone frequencies. *Journal of Genetic Psychology*, *89*(1), 95–103.

Symons, A. E., El-Deredy, W., Schwartze, M., & Kotz, S. A. (2016). The functional role of neural oscillations in non-verbal emotional communication. *Frontiers in Human Neuroscience*, *10*, 239. https://doi.org/10.3389/fnhum.2016.00239

Valdés-Alemán, P., Zamudio-Gurrola, A., & Téllez-Alanís, B. (2022). The emotional link between color and music: What happens with atonal music? *Psychomusicology: Music, Mind, and Brain*, *32*(1-2), 46–57. https://doi.org/10.1037/pmu0000288

Wan, Q., Feng, X., Bei, Y., Gao, Z., & Lu, Z. (2024). Metamorpheus: Interactive, affective, and creative dream narration through metaphorical visual storytelling. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, *Article 166*, 1–16. https://doi.org/10.1145/3613904.3642410

Ward, J. (2013). Synesthesia. *Annual Review of Psychology*, *64*(1), 49–75. https://doi.org/10.1146/annurev-psych-113011-143840

Whiteford, K. L., Schloss, K. B., Helwig, N. E., & Palmer, S. E. (2018). Color, music, and emotion: Bach to the Blues. *i-Perception*, *9*(6), 1–27. https://doi.org/10.1177/2041669518808535

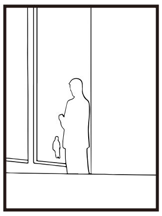
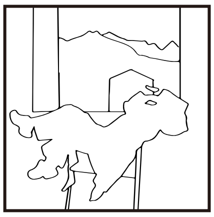
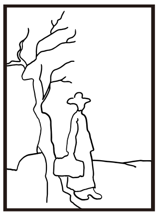
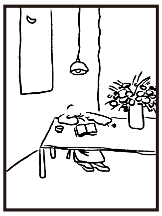
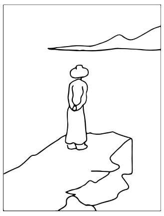
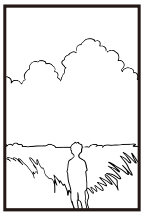
Wilms, L., & Oberfeld, D. (2018). Color and emotion: effects of hue, saturation, and brightness. *Psychological Research*, *82*(5), 896–914. https://doi.org/10.1007/s00426-017-0880-8

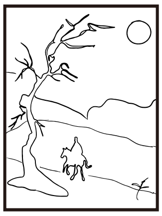
Xu, M. (2024). Analysis of the cultural significance and social rationale of "Red" in Chinese culture. *Proceedings of the 2nd International Conference on Social Psychology and Humanity Studies*, *42*, 69–75. https://doi.org/10.54254/2753-7048/42/20240827

Yu, Z. P., Cao, Y., Cao, K. (2012). *Colorblind Map (Version 5)* (5 ed.). People's Health Publishing House.

Zhang, Y., Han, X., & Chen, R. (2025). Non-linguistic pitch contextually modulates facial emotion recognition. *Proceedings of the 18th International Symposium on Visual Information Communication and Interaction (VINCI'25)*, *Article No.: 16*, 1–5. https://doi.org/10.1145/3769534.3769539

Appendix I: Descriptive Options for the Ten Scenarios

No.	Scenarios	Positive emotion description options (Emotional coding: 1)	Negative emotion description options (Emotional coding: -1)
1		A father is walking home happily with his child who just got out of school in sunset. 爸爸牵着刚放学的孩子，在夕阳下快乐地散步回家。	Under the setting sun, the divorced father took his children to another home in loneliness. 夕阳下，离婚的爸爸带着孩子落寞的去往另一个家。
2		A man is waiting at the window expectantly to welcome his wife back home from a business trip. 男人在窗前期待的迎接出差回家的妻子。	The man reluctantly watched his wife leave for work in another place. 男人不舍地目送去外地工作的妻子离开。
3		On a sunny afternoon, she lay comfortably on a chair, enjoying the quiet time. 在阳光充足的午后, 她惬意的躺在椅子上，享受恬静的光。	In the dead of night, she helplessly lay on a chair, escaping from the frustrations of work. 在夜深人静的夜晚，她无奈的躺在椅子上逃避工作的不顺。
4		After he became successful, he went home to visit his parents, and he was filled with joy at the prospect of reuniting with his family. 他在功成名就后回家探望父母,心中是将与家人团聚的悦。	He lonely left his hometown to work, and he was reluctant and unwilling to leave his hometown. 他独自离开家乡独自外出打拼，心中是即将背井离乡的不舍。
5		With the books spread out on the table and the moonlight outside the window, he was swimming in the ocean of knowledge, his heart was full of determination and persistence in pursuing his dreams. 桌上摊开的书本，窗外的月色，他在知识的海洋中遨游，内心满是追求梦想的坚定与执着。	It was midnight, he was still writing under the desk lamp, in this quiet night, his heart was full of anxiety about preparing for the exam alone. 夜已深，他还在台灯下奋笔疾书，在这寂静的夜晚，内心满是独自备考的焦虑。
6		He stood on the top of the mountain, facing the breeze, his heart full of infinite longing for the future. 他站在山顶，迎着微风，心中充满对未来的无限憧憬。	He stood on the edge of the cliff, bowed his head and thought, his heart full of heavy worries and uneasiness. 他站在悬崖边，低头沉思，心中满是沉重的忧虑不安。
7		He stood on the boundless grassland, looking at the sky in the distance, his heart full of hope, as if he saw the infinite possibilities of the future, and every step was in the rhythm of freedom. 他站在无垠的草地之上，望着远方的天空，心中充满希望，仿佛看到了未来的无限可能，每一步都踏着自由的节奏。	He stood alone on the vast grassland, surrounded by empty space, the grass swaying in the wind, his heart was full of confusion and loneliness, as if he was abandoned by the world and didn't know where to go. 他独自站在广袤的草原上，四周空旷无人，草在风中摇曳，他心中满是迷茫和孤独，仿佛被世界遗弃，不知该何去何从。

8		The two of them were boating on the lake, looking at the greenery, and their hearts were filled with leisurely and comfortable contentment. 两人泛舟湖上，看着这满眼绿意，内心满是悠然自得的意。	He and his old friend sat opposite each other on the boat, about to part, and his heart was full of silence and reluctance before parting. 他和老友在船上对坐，即将分别，心中满是离别前沉默和不舍。
9		The sun was high in the sky, he galloped on the mountain trail, the sun shone on his body, and his heart was filled with the joy of running towards freedom and the expectation of the road ahead. 太阳高悬，他策马奔腾在山间小道，阳光洒在身上，心中是奔向自由的轻快愉悦和对前路的期待。	As the sun set, he rode alone on a winding mountain road, with mountains in a row, his heart was full of desire for home and confusion about the unknown. 夕阳西下，他独自骑马行进在蜿蜒的山路上，群山连绵，心中满是对归途的渴望和对未知的迷茫。
10		He walked along the winding path, the green trees beside him swaying in the breeze, his heart full of hope, looking forward to the infinite possibilities ahead. 他沿着曲折的小道前行，身旁的绿树在微风中摇曳，他心中满怀希望，期待着前方的无限可能。	He walked alone on the winding mountain road. There was no one around him and the trees beside him stood quietly. He felt lonely and helpless, as if he was abandoned by the world. 他独自走在蜿蜒的山路上，四周空旷无人，身旁的树木静静伫立，他心中满是孤独无助，仿佛被世界遗弃。