Ultimart: Visual Communication Journal
Vol 17 No. 2, December 2024
ISSN: 1979 - 0716, e-ISSN 2615-8124
Faculty of Art and Design
Universitas Multimedia Nusantara

AI-ENHANCED STORYTELLING: INTEGRATING VISUAL, TEXTUAL, AND AUDITORY ELEMENTS THROUGH MULTIMODALITY

Evan Raditya Pratomo

Received October. 12, 2024; Revised December. 06, 2024; Accepted December. 17, 2024.

Abstract: *This study explores the integration of Artificial Intelligence (AI) in experience design, focusing on its role in creating a multi-sensory experience that bridges visual, textual, and auditory elements. The research centers on the album "Sky of Summer Nights," where music generated by Suno AI complements illustrations and lyrics created by the author, Evan Raditya Pratomo. The project features three original illustrations, each corresponding to a song: "The Fairy Tale of Pink Summer," "A Hug for Pink Moon," and "Freed His Heart." This research demonstrates how creative expression can be enhanced through multimodal principles, highlighting the challenges of effectively conveying messages through a combination of mediums. The findings affirm that the integration of AI can preserve the integrity of traditional design practices while offering new possibilities for innovation. By applying multimodal design theory, the study illustrates how AI is not only a tool but a catalyst for new emotional and creative experiences across disciplines. This research emphasizes the collaboration between human creativity and Suno AI, driving forward innovation in the fields of visual communication and experience design.*

Keywords: *sky of summer nights; multi-sensory; multimodal; creative expression; suno AI*

Introduction

The rise of AI in the creative industry has changed how artists work. While AI disrupts some traditional practices, it also empowers creators with tools that make it easier to express ideas (Hutchin, 2024).

Recent advancements in deep learning and neural networks have enabled machines to compose music independently, challenging conventional views of creativity and authorship (Briot et al., 2020). Platforms like Suno AI are at the forefront, creating AI-generated music according to user preferences. Systematically, it takes on the roles of music producer and sound engineer. One example of this implication is the author's album, "Sky of Summer Nights." This album combines Suno AI's music with the author's illustrations and poetry-turned-lyrics, creating a seamless visual and auditory art blend. The result is distributed via Routenote and available to stream on Spotify, Apple Music, and YouTube. Highlights how AI music can mesh with an artist's broader vision, turning sound and visuals into a unified artistic experience.

Evan Raditya Pratomo is a lecturer at the Universitas Ciputra, Surabaya.

e-mail: evan.raditya@ciputra.ac.id

The use of Suno AI as a tool for creating auditory expressions in the "Sky of Summer Nights" album underscores that AI plays a dual role, serving both as an enhancer of creativity and a collaborator in the artistic process. This study explores how AI can effectively bridge visual, textual, and auditory experiences, transforming art into multi-sensory experiences and represented through multimodality to engage audiences in new and emotionally resonant ways. AI's ability to heighten emotional engagement is crucial in this method. According to Li et al. (2023), multimodal interaction frameworks provide a helpful way to analyze the emotional impact of digital art. By incorporating various modes, such as visual, textual, and auditory elements, AI can enhance the emotional resonance of a piece, delivering multi-sensory experiences that connect with audiences on multiple levels.

This research explores how listeners emotionally resonate with and aesthetically perceive a music album's visual and auditory elements. Specifically, the study seeks to understand the relationship between the album's imagery and sound and how these elements influence the listener's overall experience and engagement with the album.

Methodology

This study adopts a qualitative methodology to explore the album "Sky of Summer Nights" through three illustrations, "The Fairy Tale of Pink Summer," "A Hug for Pink Moon," and "Freed His Heart," as the main inspiration for the song generated by Suno AI. The study delves into how integrating visual, textual, and auditory elements shapes the listening experience in nuanced, context-dependent ways by focusing on participants' emotional reactions and perceptions of aesthetic appeal. Using multimodal theory (Kress, 2009), it examines how these combined modes enhance emotional engagement and creativity while considering how AI's compositional methods and algorithms influence traditional notions of creativity and authorship (Briot et al., 2020).

The qualitative methods used included in-depth interviews (Eppich et al., 2019) and open-ended survey questions. Participants will be asked to reflect on their emotional responses to the music and album artwork and how these elements interact to create a cohesive or disjointed experience. The data will be analyzed using thematic analysis to identify recurring emotional reactions, aesthetic perceptions, and insights into how the visual and auditory elements work together in the participants' minds. By adopting a multimodal approach, this research highlights the potential for AI-generated content to complement human creativity, expanding the boundaries of artistic expression

Result

1. Multimodal Theory

Multimodal theory explores how diverse modes of communication, such as visuals, text, and sound, interact to create integrated, multi-sensory experiences (Jewitt, 2008). Rather than isolating these elements, multimodality emphasizes their collaborative potential to deliver more decadent, emotionally resonant interpretations. For instance, in video games, the interplay of visual design, color, sound effects (SFX), and music layers adds emotional depth to the player's experience.

Building on this foundation, Norris (2011) highlights that multimodal theory transcends language boundaries by focusing on mediated discourse analysis, emphasizing human actions as central to communication. This approach underscores the importance of semiotic resources such as images, gestures, and

posture as essential tools for constructing meaning in interaction. Similarly, Kress (2009) extends this perspective through his metafunctional approach, arguing that modes such as visual grammar and interpersonal metafunctions dynamically interact to establish relationships between visual elements and their viewers. For example, visual features like gaze, distance, and angle evoke emotional connections, while auditory elements enhance these dynamics by contextualizing emotions and narratives. This interplay between modes demonstrates the adaptability of multimodality across diverse contexts, where distinct forms collaboratively generate layered and nuanced interpretations.

In the context of the album "Sky of Summer Nights," multimodality is evident in the interaction between AI-generated music, poetry, and illustrations. The illustrations, such as "The Fairy Tale of Pink Summer" and "Freed His Heart," were conceptualized prior to the creation of the audio compositions. This sequence represents a reverse integration process, where visual elements inspire auditory elements rather than the traditional opposite. Using Suno AI as a compositional tool, this project explored how AI algorithms could align with pre-existing visual narratives, resulting in an emotionally immersive multimodal experience. For instance, the dreamlike quality of the illustrations directly informed the musical tones and instrumentation, illustrating a symbiotic creative process between human intention and AI collaboration.

This study applies multimodal theory to analyze how the combination of illustrations, AI-generated music, and poetic lyrics in "Sky of Summer Nights" enhances the audience's emotional engagement and aesthetic appreciation. Aligning with Hiippala's (2021) principle that multiple modes evoke meanings more nuanced than any single mode alone, this project demonstrates how multimodal integration can expand the boundaries of creative expression, blending traditional artistic processes with emerging AI technologies.

2. Creative Design and Exploration

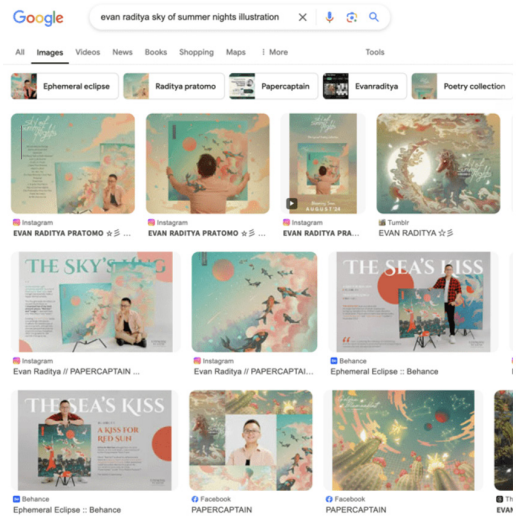

Figure 1. Google search result by inserting author's name, project, and illustration as keywords (Source: Personal research documentation)

As shown in Figure 1, the author's creative work spans illustration, design projects, and poetry collections. In 2017, the author published The Koi Fish Rhapsody, a poetry and illustration collection that established a niche approach distinct from other digital illustrators. Recognizing the parallels between poetry and lyrical songwriting, the author expanded their creative process to include music by integrating Suno AI, a platform for generating AI music.

The album "Sky of Summer Nights" in Figure 2, featuring 18 songs, exemplifies this evolution. The process began with illustrations that inspired poetic writing, which later became song lyrics. By adding an auditory dimension to the visual and textual content, the album highlights the synergy between AI-generated music and human creativity, enriching the storytell-

ing and emotional depth of the original artworks.

Figure 2. Cover digital album "Sky of Summer Nights" (Source: Personal research documentation)

During the research process, the author surveyed 34 Generation Z participants aged 17 to 22, primarily Visual Communication Design students. This demographic was selected for their familiarity with art and design principles, which provided a nuanced perspective on the album's illustrations and AI-generated music. The author shared three illustrations and the concepts behind them, each of which shares a title with one of the audio tracks in the album. The data was analyzed using thematic analysis to identify recurring emotional responses and aesthetic perceptions. The results highlight how participants perceived the interplay between visual and auditory elements, with many noting that the integration enhanced their emotional engagement and appreciation of the content. This analysis underscores the importance of a multimodal approach in fostering a deeper connection between the audience and the artwork.

3. Segmental Purpose

The album's storytelling serves as a cornerstone of the project's value, leveraging a lyricism style that resonates with Generation Z. By exploring themes of love, life, and loss, the album reflects everyday experiences familiar to its audience. This integration of textual, auditory, and visual modes enhances relatability and emotional engagement, directly contributing to the storytelling process.

The album weaves modern vocabulary with poetic narratives, authentically connecting to the experiences of its target audience. This multimodal storytelling amplifies the album's emotional depth, aligning seamlessly with its listeners' generational sensibilities.

4. Multi-Sensory Experience

As shown in Figure 3, 48% of the audience for "Sky of Summer Nights" falls within the 18–22 age range. This data aligns with the album's target demographic of Generation Z, a group known for preferring interactive, emotionally resonant content (Wright et al., 2005). This demographic data supports the relevance of the album's themes and its multimodal approach to engaging this audience.

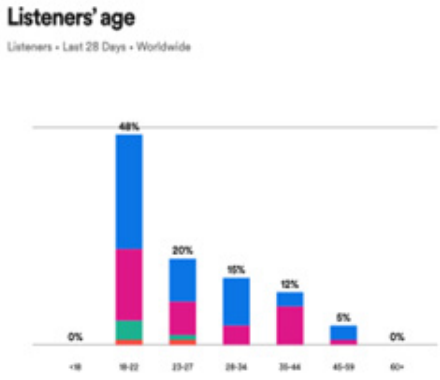

Figure 3. Demographic data for the audience's age (Source: Spotify for Artist; personal research documentation)

Integrating visual art, poetry, and music in "Sky of Summer Nights" exemplifies a multidimensional approach that unites visual and auditory art forms (Briot et al., 2020). The album employs consistent color schemes, ambiance, and terminology across its illustrations, lyrics, and music, forming a cohesive multimodal narrative. For example, in "The Fairy Tale of Pink Summer" in Figure 4 below, the recurring imagery of sunset skies, clouds, red panda, and tiger evokes a playful atmosphere. A respondent noted how the mood set by the illustrations connects seamlessly with the songs in the album, enhancing the storytelling experience.

Figure 4. Illustration of "The Fairy Tale of Pink Summer"
(Source: Personal research documentation)

Suno AI's composition process demonstrates the interplay between auditory and textual dimensions, where poetic elements such as metaphor, vivid imagery, and rhyming schemes enhance the emotional resonance of the songs. For instance, "A Hug for Pink Moon," as shown in Figure 5, illustrates a fish flying in a turquoise sky along with pink clouds. It is interpreted through the lyrics, which evoke themes of hope and longing.

This imagery creates a layered experience by juxtaposing the natural order of fish as aquatic creatures confined to water with a whimsical reality where they soar through the sky. One respondent noted that this pairing amplified the song's themes, highlighting the emotional depth achieved through this interplay of visual and lyrical elements.

Figure 5. Illustration of "A Hug for Pink Moon"
(Source: Personal research documentation)

This multimodal synergy reflects multimodal communication, where text, visuals, and sound interact to create meaning richer than any mode could achieve alone (Kress, 2009). For "Sky of Summer Nights," the interplay of these elements can be seen in "Freed His Heart" in Figure 6. A respondent noted that the illustration feels more cheerful and lighthearted when paired with its song, demonstrating how auditory and visual elements collectively evoke a dynamic narrative interpretation.

Figure 6. Illustration of "Freed His Heart"
(Source: Personal research documentation)

Discussion

1. Visual Resonance

The discussion integrates findings from a qualitative analysis of respondents' interpretations of the album "Sky of Summer Nights" and its accompanying illustrations. As described in the methodology, participants engaged with the illustrations while listening to the corresponding songs, allowing for an in-depth exploration of how visuals and lyrics combined to evoke emotional and thematic resonance. The responses were analyzed inductively, focusing on recurring patterns, symbolic interpretations, and personal connections articulated by the respondents.

1.1. The Fairy Tale of Pink Summer

As can be seen from Figure 4 in Multi-Sensory Experience above, this illustration, paired with the song The Fairy Tale of Pink Summer, explored themes of friendship and shared experiences through a multimedia lens. Respondents described the illustrations as "inviting" and "soft," with motifs such as cherry blossoms and sunset skies being interpreted as symbols of fleeting yet beautiful moments of connection. One participant noted, "The colors remind me of warm memories, like spending time with people I care about."

This qualitative insight underscores how recurring imagery of a pinkish atmosphere and delicate blossoms facilitated emotional engagement with the narrative. By closely examining participants' descriptions, it becomes evident that the multimodal design strengthens a shared understanding of friendship's ephemerality. The visual and auditory components work together to amplify and highlight the author's commitment to designing a cohesive multimedia experience, echoing contemporary art trends that increasingly employ multimedia techniques to engage audiences more deeply (Haddock, 2024). Haddock adds, "Stories are how we try and make sense of the world around us," and multimedia storytelling can offer immersive experiences that resonate on a profound level. The integration of these elements can amplify the story, fostering deeper emotional connections and enhancing learning outcomes (Haddock, 2024).

1.2. A Hug for Pink Moon

"A Hug for Pink Moon," as shown in Figure 5 above, plays a central role in shaping the album's overall aesthetic. Its prominent color scheme, featuring starry skies, pink clouds, turquoise, the pink moon, and flower petals, directly influences the album's visual elements, including the cover art shown in Figure 5. The dreamlike illustration for "A Hug for Pink Moon" prompted interpretations centered on longing and hope. As mentioned before, the depiction of fish swimming in the sky was frequently mentioned by respondents, with one interpreting it as a metaphor for transcending boundaries: "It feels surreal but comforting, like even things that seem impossible can feel close." This interpretation echoed how respondents linked the visuals to the song's lyrical repetition of "Hug," reinforcing themes of emotional proximity despite physical separation.

The qualitative responses reveal the significance of symbolism in evoking personal connections. This finding aligns with the study's focus on understanding how participants interpret and emotionally engage with multimodal. Most comment about this illustration and song version, mainly about the theme of longing and connection, interpreting the moon imagery as symbolic of emotional closeness, despite physical distance. The dreamlike atmosphere, influenced by

night scenes and anime, resonated deeply with respondents, supporting findings that immersive visual environments can enhance emotional depth and creative engagement (Chambel et al., 2013).

1.3. Freed His Heart

"Freed His Heart," as shown in Figure 6, symbolizes a figure reaching for the moon, eliciting rich and varied interpretations from participants. Many described the artwork as representing personal liberation and the courage to embrace new opportunities. One respondent stated, "The calm colors and the reaching gesture remind me of letting go of something heavy and moving forward."

These interpretations highlight how the interplay of visuals and lyrics can evoke layered meanings tailored to individual experiences. The emphasis on qualitative data allows the discussion to focus on how participants made sense of these multimodal elements in ways that resonated with their journeys. From the multimodal perspective, the consistency in color schemes, ambiance, and terminology between the illustration, lyrics, and song becomes evident. The colors used in the artwork serve as mood setters, guiding the emotional tone when the author translates the visuals into lyrics and eventually into music.

2. Future Implications

This study underscores the transformative potential of multimodal communication in crafting layered, emotionally resonant storytelling experiences. The album "Sky of Summer Nights" demonstrates how multimodal synergy enhances audience engagement by integrating visuals, music, and text. Respondents highlighted how combining colors, metaphors, and melodies contributed to a comprehensive multi-sensory experience. For instance, one participant described the illustrations as "soft and inviting." At the same time, another emphasized the emotional closeness evoked by the song A Hug for Pink Moon, showcasing the role of multimodal coherence in fostering personal connections.

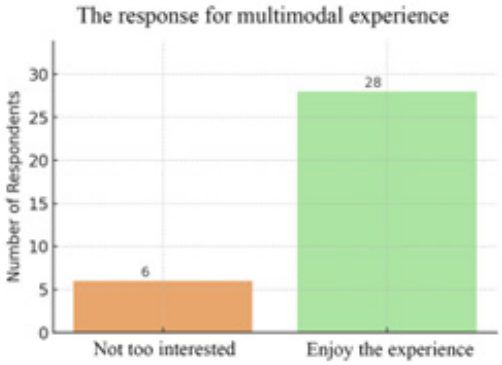

Figure 7. The response to multimodal experience (Source: Personal research documentation)

Based on Figure 7, survey respondents shared predominantly positive responses regarding the multimodal experience. One participant reflected on integrating AI music and illustrations, stating, "Some parts sound strange, especially the falsetto notes, and some melodies feel repetitive. However, even though it does not sound completely human, it does set a nice mood that matches the illustrations. I liked the lyrics too; they hold meaning and are also poetic." This highlights both the limitations and strengths of the project's multimodality, where AI-generated music and visual elements are aligned to enhance emotional resonance. Respondents' acknowledgment of the music's ability to "set a nice mood" supports the study's findings that multimodal coherence is key to audience engagement.

The illustrations played a foundational role in setting the tone and guiding the design communication process, and respondents described them as "inspirational." This observation aligns with the

research question exploring how AI-generated visuals and auditory elements contribute to emotional engagement. By integrating these elements, the project demonstrates AI's potential in expanding multimodal storytelling, even as it reveals areas for refinement, such as improving the naturalness of AI-generated sounds.

This study aligns with the theoretical concept that multimodal communication creates richer meanings by combining different modes (Kress, 2009). It highlights how visuals, text, and music interact dynamically to create cohesive storytelling, offering practical insights for industries such as branding and education. Participants' interpretations of symbolic imagery and lyrical repetition further underscore the importance of multimodal coherence in fostering emotional connections, a principle that can guide impactful multimedia campaigns.

The study points to future directions for creative industries, particularly in leveraging tools like AI for multimodal integration. Respondents noted how the album's blend of illustrations and music amplified emotional resonance, suggesting that AI-generated components could further refine these experiences. However, challenges persist, particularly in balancing technical precision with emotional authenticity. One participant noted that while specific falsetto notes felt "strange," the overall composition successfully complemented the visuals, underscoring the importance of prioritizing emotional impact in AI-driven storytelling.

Finally, multimodal communication offers a framework for bridging artistic innovation with audience engagement. Integrating text, images, and sound provides a template for cohesive narratives in branding, education, and interactive media.

Conclusion

The study confirms that the illustrations successfully stimulated cognitive and emotional engagement when reflecting on the feedback regarding illustrations and their correlation with the final output of AI-generated songs. Respondents appreciated how the visuals complemented the music, noting the evocative mood they set and the deeper connections they fostered with the lyrics. This feedback underscores the effectiveness of multimodal integration in enhancing audience experience and emotional resonance.

The findings also reveal a range of interpretations, highlighting Generation Z's exploratory nature. Their engagement with diverse artistic mediums, as evident in the survey responses, aligns with their broader characterization as "digital natives" (Grow & S, 2018) who have grown up at the intersection of technology and creativity. However, this openness is not merely a generational stereotype but was reflected in their nuanced responses to the project. For instance, participants appreciated the innovative blending of art and technology, even as they critiqued specific aspects like the repetitiveness of melodies or the strangeness of falsetto notes. This suggests a willingness to embrace emerging technologies like AI while maintaining high standards for creative outputs.

From a practical perspective, the study demonstrates the potential of tools like Suno AI to expand the creative possibilities for illustrators and musicians. By enabling seamless visual, auditory, and textual integration, such technologies open new avenues for artists to craft multisensory experiences. Respondents noted that the synergy between illustrations and music amplified the emotional impact, showcasing how multimodal storytelling can bridge gaps between traditional and digital art forms.

While the results highlight promising applications of AI in creative industries, they also underscore challenges, such as ensuring the naturalness of AI-generated sounds and addressing ethical concerns around creative ownership. Future research could delve into these challenges more deeply, exploring how multimodal coherence operates in diverse cultural contexts or emerging technologies like augmented and virtual reality. Such inquiries could further refine our understanding of how art and technology intersect to enhance storytelling.

References

Alruthaya, A., Nguyen, T. T., & Lokuge, S. (2021, November 18). The application of digital technology and the learning characteristics of Generation Z in higher education [Preprint]. arXiv. https://arxiv.org/abs/2111.05991

Briot, J.-P., & Pachet, F. (2020). Deep learning for music generation: Challenges and directions. Neural Computing and Applications, 32(981-993). 0pp[https://doi.org/10.1007/s00521-018-3813-6

Chambel, T., Bove, V. M., Strover, S., Viana, P., & Thomas, G. (2013, October). Immersive media experiences: immersiveme 2013 workshop at ACM multimedia. In Proceedings of the 21st ACM International Conference on Multimedia (pp. 1095-1096).

Clarke, M. (2010). The concise Oxford dictionary of art terms. Oxford University Press, USA. https://doi.org/10.1093/acref/9780199569922.001.0001

Eppich, W. J., Gormley, G. J., & Teunissen, P. W. (2019). In-depth interviews. Healthcare simulation research: A practical guide, 85-91.

Grow, J. M., & Yang, S. (2018). Generation-Z Enters the Advertising Workplace: Expectations Through a Gendered Lens. Journal of Advertising Education, 22(1), 7-22. doi: 10.1177/1098048218768595

Haddock, J. (2024). The unique power of storytelling in immersive learning experiences. Forbes Business Council. https://www.forbes.com/councils/forbesbusinesscouncil/2024/02/23/the-unique-power-of-storytelling-in-immersive-learning-experiences/

Herman, D. (2010). Multimodal storytelling and identity construction in graphic narratives. Narrative, 18(2), 127-153.

Hiippala, T. (2021). Distant viewing and multimodality theory: Prospects and challenges. Multimodality & Society, 1(2), 134-152

Hutchins, B. (2024). AI and the Creative Industry: A Paradox of Disruption and Empowerment. https://bobhutchins.medium.com/ai-and-the-creative-industry-a-paradox-of-disruption-and-empowerment-9a0c69b2a1e1

Jewitt, C. (2008). Multimodality and literacy in school classrooms. Review of Research in Education, 32(1), 241-267.

Kress, G. (2009). Multimodality: A social semiotic approach to contemporary communication. Routledge.

Li, X., & He, J. (2023). Exploring the emotional design of digital art under the multimodal interaction form. In Design, User Experience, and Usability: HCII 2023 (pp. 709-723). Springer. https://doi.org/10.1007/978-3-031-35699-5_40

Makhmudov, F., Kultimuratov, A., & Cho, Y. I. (2024). Enhancing Multi-

modal Emotion Recognition through Attention Mechanisms in BERT and CNN Architectures. Applied Sciences, 14(10), 4199.

Norris, S. (2011). Three hierarchical positions of deictic gesture in relation to spoken language: a multimodal interaction analysis. Visual communication, 10(2), 129-147.

Perloff, M. (1998). Poetry on & off the page: Essays for emergent occasions. Evanston: Northwestern University Press.

Silvia, P. J. (2009). Looking past pleasure: anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. Psychology of Aesthetics, Creativity, and the Arts, 3(1), 48.

Wright, P., & McCarthy, J. (2005). The value of the novel in designing for experience. In Future interaction design (pp. 9-30). London: Springer London