forum acusticum 2026
8-12 September · Graz, Austria

AUDIO–VISUAL REDUNDANCY AND COMPLEMENTARITY IN SOUND AND MUSIC CONTEXTS

David Adlberger, Katharina Groß-Vogt

Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz

Abstract

This paper explores the relationship between audio and visuals across three core contexts: academic visualization and sonification, creative tools for synthesizing sound from images, and artistic practice in live performances and installations. We apply a framework that characterizes redundancy and complementarity in audio-visual displays.This analysis reveals distinct patterns: the academic domain predominantly relies on rule-based mappings and aims for redundancy, yet often fails to achieve perceptual equivalence, particularly for non-experts. Creative tools, while occasionally borrowing systematic conventions, largely employ unsystematic, artist-driven encodings that exploit complementarity as a generative principle. Artistic practice takes this furthest, treating redundancy and complementarity as compositional parameters. These findings reveal a gap in standardized methods and tools to support more comprehensible yet creative audio-visual practices.

Keywords: *audio-visual display, multimodality, visualization, sonification*

1. Introduction

In various areas of human-computer interaction (HCI), audio is frequently paired with visuals, and sometimes vice versa, for didactic, exploratory, and engagement purposes. This paper systematically reviews such practices across three main usage contexts: (a) the academic domain, including standard visualizations of sound and didactic examples of sonification and auralization; (b) creative tools that generate sound from visual input, including dedicated tools generating sound from static images; and (c) selected artistic works such as live performances and installations. Our corpus of examples was assembled through a combination of literature and web searches. The approach supports a hands-on assessment of redundancy and complementarity between the visual and auditory components of combined displays in current sound and music contexts.

To keep within scope, this paper does not cover HCI in the form of earcons, music education content (e.g., circular musical geometries, polyrhythm visualization, musical math), historical aspects such as Chladni figures or photographic soundtracks, or outdated methods like visuals in audio players. In the following, we cite and discuss examples from each category, assessing them according to their level of redundancy and complementarity, as described in Sec. 2.

This paper is part of an ongoing Master's thesis in Sound Design, at the University of Music and Performing Arts Graz, Austria, which aims to fill existing gaps in converting pictures to sound by using systematic approaches in creative tools.

***Corresponding author**: david.adlberger@student.kug.ac.at, vogt@iem.at.*

2. Redundancy and Complementarity Along the Perceptualization Path

In audio-visual displays, across contexts, one modality (audio, visuals) is more or less complementary to the other. To allow for a systematic assessment of this relationship, we refer to a framework for audio-visual (AV) information displays [1]. This approach looks at different stages of the perceptualization path in a design process: first, (1) in Tab. 1, is the choice of the content or dataset. Next comes (2a), the encoding of data into visuals and audio, which can qualitatively be described as encoding paradigm (e.g., abstract or symbolic, embodied, affective); and quantitatively (2b) by a transfer function, which is based on psychometric functions or standard mappings. Finally (3a,b), any AV display has a real-world representation. In a standard workplace, this refers to the visual monitor and headphones, but across use cases more immersive environments come into play (e.g., AR, VR). An illustrative overview of what redundancy and complementarity means for these aspects is shown in Tab. 1.

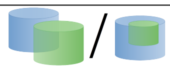
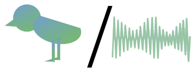
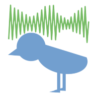
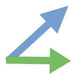
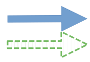
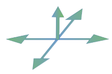
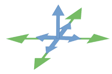
Aspect	Redundant	Complementary
1 - Content: what?		
		
2a - Encoding paradigm: how mapped?		
		
2b - Transfer function: in which way?		
		
3a - Temporal representation: when?		
		
3b - Spatial representation: where?		
		

Table 1: Basic aspects for the fully redundant and complementary case. Blue and green colours stand for either visualization and sonification or v.v. In the redundant case, the colours overlap as a gradient, meaning that both sonification and visualization employ the same or are equivalent in that regard. For comp. content and red. encoding paradigm, two versions are illustrated.

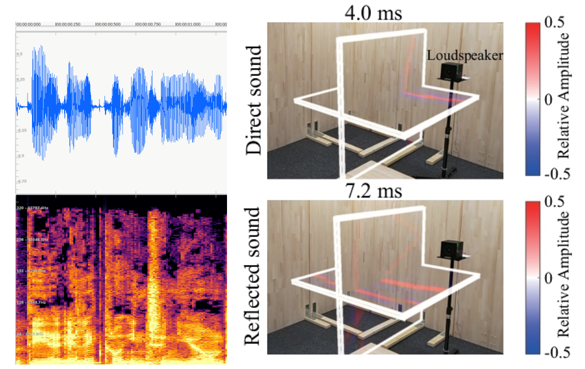
Redundant means over all aspects that both visualization and sonification employ the same or equivalent content, encoding or representation, while complementary means they differ and complement each other. Note that the relation between visuals and sounds can be assessed on a continuum between redundancy and complementarity. Full redundancy or complementarity serve as extreme cases, that can only be reached in an equivalent manor, because hearing and seeing work differently.

3. The academic domain

The academic domain contains both the visualization of acoustics for didactic and analytical purposes, as well as the sonification of data from other disciplines.

3.1. Standard Visualizations: Waveform and Spectrogram

The most common visualizations of audio signals, used for teaching acoustics and as standard representations in software tools, are waveforms and spectrograms (see Fig. 1a). Both are rule-based, directly representing physical properties through well-established transfer functions [3], [4]. Waveforms display the acoustic pressure signal in the time domain, revealing transients, pauses, periodicity, and amplitude envelopes. Spectrograms extend this by decomposing the signal into frequency components over time, enabling inspection of harmonic structures, formants, and broadband noise. Standardized amplitude scaling, logarithmic frequency



a) waveform and spectrogram
b) MR visualization of a 3D Room Impulse Response [2]

Figure 1: Examples from the academic domain

axes, and colour mappings make both representations highly reproducible across software environments.

3.2. Visualization in Room Acoustics

Beyond standard signal representations, the academic domain includes visualizations of acoustic phenomena derived from physical simulations. Sound pressure level distributions in enclosed spaces, commonly used in room acoustics education, employ color mapping to represent spatial variations corresponding to standing wave patterns and room modes, providing intuitive access to complex acoustic phenomena. Extending this into immersive technologies, Uchida et al. [2] proposed an MR visualization of room impulse response maps combining room geometry with a physical model of sound propagation (see Fig. 1b), allowing users to observe how sound waves propagate and reflect within a space. More recently, Onuma et al. [5] developed interactive sound field visualization using mixed reality, enabling real-time manipulation of acoustic simulations based on room geometry.

Another usecase is Auralization, which generates audible sound from computer data representing acoustic environments and source characteristics modified by propagation and transmission in systems such as rooms, buildings, or vehicles [6], [7]. Developing auditory simulations of concepts such as reverberation, flutter echoes, reflections, room modal effects, and sound transmission through barriers [8]. More recent implementations include walkable auralizations for experiential learning in immersive classrooms and open-source frameworks enabling real-time auralization of 3D room models [9], [10], [11].

3.3. Sonification

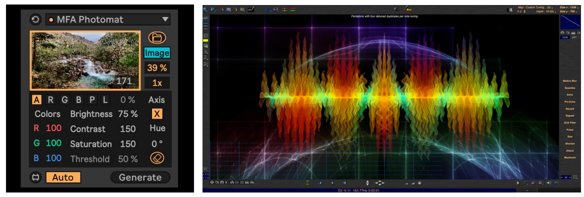
A related process is the sonification of scientific data [12]. The most direct form, audification, converts measured signals into audible waveforms without intermediate symbolic mapping—the inverse of waveform visualization (Sec. 3.1). Another common form is the auditory graph, where data values are mapped to pitch over time. More generally, parameter mapping maps any data variables to sound variables and is applied case-specifically across domains, most frequently in physical sciences, social sciences, and applied sciences, followed by life, health sciences, and arts and humanities [13]. Various toolkits support academic sonification, including the HSV Sonifier for image-content analysis [14], the WebAudioXML Sonification Toolkit (WAST) for browser-based applications [15], and the Sonification Application Research Toolbox (SonART) for platform-independent parameter mapping [16].

3.4. Discussion: redundancy in academic contexts

Applying the framework discussed in Sec. 2 reveals distinct patterns across academic contexts. Standard visualizations such as waveforms and spectrograms achieve near-complete redundancy: identical content, systematic abstract encoding, rule-based transfer functions, and synchronous temporal representation. Yet perceptual equivalence often fails for non-experts lacking training in decoding these abstract forms.
Sonification, as the inverse process, shares the aim for redundancy but introduces subtle departures. Parameter-mapping typically employs overlapping yet non-identical datasets, while transfer functions range from psychometrically grounded mappings (e.g., pitch–height) to embodied or representative sounds. Encoding paradigms shift toward complementary pairings, abstract visuals with symbolic or affective sounds, and spatial and temporal representation vary by use case: some sonifications achieve redundancy in both dimensions, while others remain monophonic or temporally offset.
Auralization occupies an intermediate position, combining physically derived content with immersive spatial representations that introduce complementarity between the visual environment and the auditory simulation.
Overall, audio-visual displays in this category are predominantly designed with redundancy as their primary objective: the visual representation aims to faithfully reflect the acoustic signal, or vice versa, to enable direct analytical or perceptual comparison. Yet they partly fail to achieve perceptual equivalence across modalities, underscoring the need for standardized mappings and evaluation methods that account for listener competencies.

4. Creative Tools for sound synthesis from visual input

In the following, we discuss creative tools that synthesize audio from visual input—excluding conventional graphical interfaces used merely to control synthesis parameters. These tools fall into two categories: customizable tools that derive audio from full images or selected regions via image processing (e.g., raster scanning, averaged features, or object-based analysis), and interactive tools where users co-create or co-manipulate visuals and sound. These tools are creation-oriented: the audio constitutes the final output, while the image functions as input and is typically not presented alongside or perceptibly connected to the resulting sound.



a) Photomat [17] b) Metasynth [18]

Figure 2: Examples from the creative tools domain

4.1. Customizable tools based on full or selected parts of an image: creating sounds from visuals

Creative tools for image sonification employ a spectrum of mapping strategies, from direct pixel-to-parameter translation to semantically informed sound assignment. At the direct end, Phosphor [19], a free software instrument for macOS, uses spectral synthesis to convert images into audio by mapping brightness to loudness, vertical position to pitch, and horizontal position to time, with color mode additionally shaping timbre. Similarly, VOSIS [20] combines audification of greyscale pixel data with image filters and ADSR envelope control, allowing non-linear pitch sequences with varying timbre. Both tools treat the image as a direct data source, achieving partial redundancy in their brightness-to-loudness or position-to-pitch mappings while introducing complementarity through user-adjustable parameters.
Moving toward more abstracted encodings, Manifest Audio's "Photomat"plug-in for Ableton Live [17], [21] (see Fig. 2a) transforms images into MIDI data using averaged pixel values, mapping RGB channels to chord voices, pixel brightness to velocity and rhythm, and luminance to percussive noise patterns. Here, the image functions as generative source material rather than a signal to be faithfully reproduced, deliberately employing unsystematic, creative mappings.
A further new strategy is represented by the Image Extender [22], which uses object-based AI image detection. Identified objects, scenes, and locations are matched to semantically appropriate sound descriptions retrieved via the freesound.org API and placed in the stereo field according to their spatial positions, creating a soundscape that represents the semantic content of the image.

4.2. Interactive tools for sound and music generation: co-creating or co-manipulating visuals and sound

Beyond tools that process existing images, a second category enables users to co-create or co-manipulate visuals and sound in real time. MetaSynth [18], available for Macintosh since the late 1990s, is perhaps the most established tool for "painting with sound". Its

Image Synth function uses drawn or imported images as sequencer sources (see Fig. 2b), where pixel areas trigger oscillators with brightness dictating amplitude and color setting stereo position. It was notably used by artists such as Richard David James (Aphex Twin) and Thomas Banghalter (Daft Punk), or composers such as Hans Zimmer and Dane Davies for music and sound effects. MetaSynth exemplifies highly unsystematic, creative encoding where visual manipulation directly yields auditory results, prioritising complementarity over redundancy.

This gesture-based paradigm extends to mobile platforms. SoundBow [23], an iOS application, enables users to generate sound by drawing curves directly on the screen, with visual gestures mapped to sonic outcomes. Built using a data sonification library, it aligns conceptually with academic principles while favoring playful, expressive mapping over systematic data-to-sound conversion. Similarly, Mazetools Soniface [24] offers a modular environment where geometry and motion generate audio through multitouch interfaces and camera-based body tracking. Its incorporation of ML-based body recognition signals a broader trend toward dynamic, real-time visual input beyond static images.

The reciprocal direction, visuals generated from audio, is exemplified by Greyland Audio's Visual Lab Pro [25], an audio-reactive plugin that translates musical parameters into over 70 visual modes. While primarily audio-to-visual, its support for custom images and extensive parameter automation also enables visual control over the auditory experience, highlighting the growing convergence of both modalities in creative software.

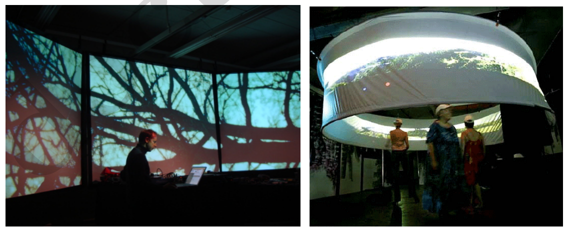
4.3. Discussion: complementarity in creative tools

Applying the framework reveals a clear contrast to the academic domain: creative tools prioritize complementarity, deliberately employing unsystematic mappings to generate novel aesthetic outcomes. Regarding content, most tools aim for redundancy—Phosphor and VOSIS treat the entire image as data, Manifest Audio's Photomat reduces images to features like mean RGB, and the Image Extender goes furthest by using AI object detection to translate visuals into semantic tags for accurate representation. The clearest departure occurs in the encoding paradigm, where visual input is paired with abstract sound. MetaSynth links pixel color and position arbitrarily to oscillator parameters, a complementary, symbolic mapping, while the Image Extender, matching objects to representational sounds, achieves partial redundancy. Transfer functions occasionally borrow conventions like pitch–height mapping, but remain predominantly intuitive and artist-driven, serving aesthetic exploration over systematic representation. Temporal and spatial representation are inherently complementary: static images mapped to dynamic soundscapes create fundamental temporal complementarity, and except for the Image Extender's spatialized soundscapes, spatial output is generally monophonic or unrelated to image features.

In sum, creative tools operationalize complementarity as a generative principle. They do not fail to achieve perceptual equivalence. They deliberately avoid it, exploiting the gap between modalities as the engine for artistic exploration. However, it can be discussed whether the approach of visual and audio sharing attributes can be achieved better by actively using redundant and complementary aspects on purpose in a more systematic approach.

5. Artistic Practice

In contrast to the creative tools discussed in Sec. 4, the works discussed in the following are generally not designed as instruments or focused on the creational context, but rather on the presentation as combined multimedia works. They can be categorized into artistic practice in live performances, where visuals are mostly added to sounds and installative context, where sounds are mostly added to visuals. The examples effectively comprises just a few subjectively selected pieces.



a) Olga Mink VJ performance 2008 [26] b) Sonic Panoramas [27]

Figure 3: Examples from the artistic works domain

5.1. Live Performances

VJing emerged from club and festival culture as real-time visual performance, where visuals are generated or mixed live using software such as TouchDesigner or Resolume, with the image–sound relationship remaining intentionally loose, improvisational, and context-dependent [26]. A related format, Live Cinema, presents audiovisual performance in a concert-like setting with the composition executed by a single artist or tightly coordinated team, as seen in the work of 242.pilots and Olga Mink (see Fig. 3a) [26]. Multimedia performances, where visuals are fixed in advance, share similar principles of temporal connection and combined-modality experience, commonly encountered on tour and at festivals.

Across these formats, several shared principles emerge. First, media equality: image and sound are treated as independent artistic layers coexisting on equal footing. Second, contextualization: concrete imagery paired with abstract sound (or v.v.) opens semantic spaces neither modality could generate alone. Third, synchronization relies on improvisation or high-level aesthetic agreements rather than fixed mappings, allowing performances to remain fluid and site-specific. While academic applications often fall short when pursuing redundancy, it can be argued that these artistic works succeed precisely through complementarity, leveraging unsystematic, creative encodings between the senses to achieve a coherent dramaturgical structure.

5.2. Installations

A second category of artistic practice mostly starts from visual material and extends it through the addition of sound.

Many installation works employ systematic image-to-sound mappings reinterpreted through artistic framing and compositional decisions, often coupling physical movement, visual exploration, and auditory response. Sonic Panoramas [27] (see Fig. 3b), for instance, uses a large-scale 360-degree projection where the visitor's position determines both the visual perspective and a spatialized soundscape generated from image data. Similarly, Photone [28] combines touchscreen interaction with musical sonification of photographs, encouraging sustained engagement through the synergy between modalities.

A special case comprises works where sound and visuals are conceived as integrated components from the outset, developed together and contributing equally to the final experience. "Voice of Sisyphus" [29] combines projected imagery with four-channel sound generated from selected image regions through raster scanning, creating an ongoing cycle between visual manipulation and sonic output. Ryo Ikeshiro employs recursive algorithms generating both visual patterns and rhythmic structures from a common computational source [30], while Marty Quinn's "Walk on the Sun" [31] uses participant movement across projected images to determine real-time sonic responses. In these works, interactivity and shared generative origins introduce an additional layer of complementarity, combining visual representation, auditory response, and physical engagement into an integrated experience irreducible to a single mapping relationship.

5.3. Discussion: complementarity in artistic works

Applying the framework reveals that artistic works predominantly seek complementarity across all aspects of the perceptualization path. In live performance, visuals and sound are treated as independent artistic layers rather than representations of one another. Installations like Sonic Panoramas or "Voice of Sisyphus" may share a common data source yet select different aspects for each modality, while Ikeshiro's recursive pieces employ shared generative algorithms producing parallel but non-identical outputs. Encoding strategies freely combine representational, abstract, and affective approaches across modalities: live performances rely on intentionally loose, improvisational mappings where synchronization follows aesthetic agreements rather than fixed rules, and contextualization, pairing concrete imagery with abstract sound, exploits semantic tension between encoding types. While the narrative layer may approach redundancy, the focus remains firmly on complementarity, with mappings subordinated to dramaturgical considerations. Temporal relationships range from improvisational synchronization in VJing to interactive, variable timing in installations like "Walk on the Sun". Spatial complementarity is pronounced in installation contexts, where multichannel sound and physical environments create distributed perceptions that complement rather than replicate the visual field. Some works introduce partial spatial redundancy, mapping position to both visual perspective and auditory spatialization, as perceptual anchors while exploring complementary relationships elsewhere.

In sum, artistic practice treats redundancy and complementarity as compositional parameters modulated to shape experience.

6. Conclusion and Outlook

This paper systematically reviewed audio-visual relationships across the academic domain, creative tools, and artistic practice, using a framework that assesses redundancy and complementarity along the perceptualization path. Our analysis reveals that the domain dictates the desired relationship: academia pursues redundancy through rule-based mappings, yet sometimes fails to achieve perceptual equivalence, highlighting a need for standardized methods. Creative tools deliberately exploit complementarity as a generative principle, using unsystematic mappings for aesthetic novelty. Artistic practice takes this furthest, treating redundancy and complementarity as compositional parameters to create semantic experiences impossible in a single modality. The framework proved apt, though this study excludes adjacent fields like HCI and music education, which future work should address. Looking ahead, this analysis reveals that creative tools and artistic practice largely forego the systematic mappings found in academic contexts. This reveals untapped potential for systematic approaches in creative contexts. The ongoing Master's thesis will therefore focus on bridging this gap by developing a systematic approach for converting pictures to sound for expressive, yet comprehensible, outcomes.

References

[1] K. Gross-Vogt, W. Aigner, S. Barrass, E. Elmquist, S. Pauletto, and K. Snook, "Redundancy and complementarity in audio-visual information displays," in *What you hear is what you see*, Forthcoming chapter, Routledge, 2027.

[2] A. Uchida, I. Tsunokuni, Y. Ikeda, and Y. Oikawa, *Mixed reality visualization of room impulse response map using room geometry and physical model of sound propagation*, Jul. 2023. DOI: 10.1145/3588028.3603693.

[3] J. Blauert, *Spatial hearing: the psychophysics of human sound localization.* MIT press, 1997.

[4] T. Rossing, *Springer handbook of acoustics*. Springer Science & Business Media, 2007.

[5] S. Onuma, A. Uchida, Y. Ikeda, A. Inoue, and Y. Oikawa, "Interactive sound field visualization with mixed reality based on geometric acoustics simulation," eng, *The Journal of the Acoustical Society of America*, vol. 158, no. 4*Supplement*, A142–A142, 2025, ISSN: 1520-8524.

[6] R. R. Torres and M. Kleiner, "Considerations for including surface scattering in room-acoustics auralization," *The Journal of the Acoustical Society of America*, vol. 105, no. 2*Supplement*, p. 1198, 1999.

[7] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, 2nd ed. Springer, 2020.

[8] L. Ronsse, D. J. Cheenne, and S. Kaddatz, "Using your ears: A novel way to teach acoustics," *Proceedings of Meetings on Acoustics*, vol. 19, p. 025 005, 2013.

[9] S. Chabot and J. Braasch, "Walkable auralizations for experiential learning in an immersive classroom," *The Journal of the Acoustical Society of America*, vol. 152, no. 2, pp. 899–910, 2022.

[10] P. Palenda, P. Schäfer, L. Aspoeck, and M. Vorländer, "Open-source real-time auralizations with virtual acoustics," *The Journal of the Acoustical Society of America*, vol. 156, no. 4*Supplement*, A64, 2024.

[11] O. Warusfel and N. Misdariis, "Evertims: Open source framework for real-time auralization in architectural acoustics and virtual reality," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2017, pp. 17–24.

[12] T. Hermann, A. Hunt, and J. G. Neuhoff, Eds., *The Sonification Handbook*. Logos Verlag Berlin, 2011.

[13] P. Lindborg, V. Caiola, M. Chen, P. Ciuccarelli, and S. Lenzi, "Re (de) fining sonification: Project classification strategies in the data sonification archive," in *AES*, vol. 72, 2024, pp. 585–602.

[14] S. Steinmetz and E. Gethner, *On musical information geometry with applications to sonified image analysis*, Mar. 2021. DOI: 10.13140/RG.2.2.27243.46882.

[15] H. Lindetorp and K. Falkenberg, *Sonification for everyone everywhere - evaluating the webaudioxml sonification toolkit for browsers*, Jun. 2021. DOI: 10.21785/icad2021.009.

[16] O. Ben-Tal, J. Berger, B. Cook, M. Daniels, G. Scavone, and P. Cook, *Sonart : The sonification application research toolbox*, Jan. 2002.

[17] Manifest Audio and N. Pred, *Sonification tools for ableton live*, Max for Live device bundle, 2025. [Online]. Available: https://manifest.audio/sonification-tools-legacy.

[18] E. Wenger, *Metasynth*, version CTX 2.x, 1997-2026. [Online]. Available: https://uisoftware.com/de/metasynth/.

[19] R. Audio, *Phosphor - a free spectral synthesizer*, 2026. [Online]. Available: https://www.roexaudio.com/phosphor.

[20] Life Orange LLC, *Vosis / vosis pro: Image sonification and visual music composition*, iOS application, 2013–2022. [Online]. Available: https://vosis.app/.

[21] Isotonik Studios and N. Pred, *Sonifying data: Image-based sound with noah pred*, Blog post, Sep. 2025. [Online]. Available: https://isotonikstudios.com/sonifying-data-image-based-sound-with-noah-pred/.

[22] D. Adlberger, *Image extender: Image sonification tool for immersive perception of sounds from images and new creation possibilities*, GitHub repository and development blog, 2026. [Online]. Available: https://github.com/addedvalue-dev/image_extender.

[23] Binaura, *Soundbow*, version Mobile App (Android 2012, iOS 2016) & Web-Based Application (HTML5 2013), 2026. [Online]. Available: https://www.binaura.net/apps/soundbow/.

[24] S. Kloß, *Mazetools soniface*, version 4.x, 2025. [Online]. Available: https://www.mazetools.com/soniface/.

[25] Greyland Audio, *Visual lab pro*, 2026. [Online]. Available: https://www.greylandaudio.com/products/p/visual-lab-pro.

[26] A. Alexander, "Audiovisual live performance," *See This Sound: Audiovisuology*, 2010.

[27] E. Kabisch, F. Kuester, and S. Penny, *Sonic panoramas: Experiments with interactive landscape image sonification*, Dec. 2005. DOI: 10.1145/1152399.1152428.

[28] N. Rönnberg and J. Löwgren, "Traces of modal synergy: Studying interactive musical sonification of images in general-audience use," *International Journal of Human-Computer Interaction*, vol. 35, no. 7, pp. 567–582, 2019.

[29] C. McGee, J. Dickinson, and G. Legrady, "Voice of sisyphus: An image sonification multimedia installation," *Leonardo Music Journal*, vol. 22, pp. 55–56, 2012.

[30] R. Ikeshiro, *Pika! ppeonjjeog, pika! ppeonjjeog / ppiikkaa!! ppppeeoonnjjjjeeoogg*, Immersive audiovisual installation, 2020.

[31] M. Quinn, "Walk on the sun: An interactive image sonification exhibit," *AI & Society*, vol. 27, no. 2, pp. 303–305, May 2012.