How can cross modal relationships between visual and auditory perception be used in generative music systems?

Dominic Goodeve

April 2026

Abstract

This research project investigates how cross-modal relationships between visual and auditory perception can be used to drive generative music systems, drawing on literature from visual art, cognitive science and music technology alongside primary research comprising a survey, focus group, practitioner interview and practice-based experimentation using Max/MSP.

The findings suggest that cross-modal associations demonstrate greater consistency than initially expected, though this requires qualification as the sample contained a disproportionately high number of participants with synaesthetic experience. Despite this, participants demonstrated a broad willingness to engage with cross-modal relationships meaningfully, providing a viable basis for generative system design. Practice-based experiments revealed that the relationship between visual input and sonic output exists on a continuum from transparent and direct to complex and mediated, and that apparent autonomy in generative systems is a product of human design rather than an inherent property of the technology.

The study concludes that visual stimuli can function effectively as generative inputs, but that the quality of the outcome is directly proportional to the designer's skill and iterative effort. As these systems move into professional contexts including live performance, installation and education, practitioners must navigate legal and ethical frameworks governing data protection, copyright and accessibility, considerations that are integral to responsible system design rather than peripheral to it.

Contents

Introduction

Relationships between sound and visual perception have long been explored across artistic practice, cognitive science and music technology, with research into cross-modal perception demonstrating that sensory modalities interact to form unified perceptual experiences (Levitin, 2006), while advances in computational music systems have enabled new approaches to musical creation through generative and algorithmic processes (Nierhaus, 2009).

Arts Council England's recognition of Digital Arts as a new supported artform, encompassing immersive installations, generative systems and data-driven practice, reflects the growing professional significance of this field within the UK creative industries (Arts Council England, 2026).

My interest in this topic arises from personal and artistic experience of synaesthesia, a perceptual condition in which stimulation of one sensory modality can trigger experiences in another (Ward, 2013), which has influenced my compositional practice and led to experimentation with generative systems that translate visual information into musical output.
Existing research has predominantly focused on either the perceptual interpretation of audiovisual relationships or the visualisation of musical structures. While these approaches establish connections between sound and image, they do not address how visual information might function as a direct generative input within computational music systems. A further complication is that the consistency of cross-modal associations remains contested, with some researchers suggesting these relationships are too subjective or culturally variable to function as a reliable basis for generative systems. There is therefore limited exploration of how visual stimuli can actively drive the creation of musical material, and a gap between theories of cross-modal perception and their practical implementation within generative systems.

This study therefore investigates the following question:

How can cross-modal relationships between visual and auditory perception be used to drive generative music systems?

While the subjective and culturally variable nature of cross-modal associations might suggest that visual stimuli cannot function as reliable generative inputs, this study contends that variability itself need not be a limitation. When systems are designed to accommodate rather than resist perceptual difference, visual information can function as a meaningful and dynamic generative source, developed through a mixed methods approach combining a literature review, cross-modal perception survey, practitioner interview, focus group, and practice-based experimentation using Max/MSP.

Literature review

This literature review considers artistic explorations of audiovisual relationships, theories of cross-modal perception, and technological approaches to generative music. Through this analysis, it identifies a gap concerning the use of visual stimulus as a direct generative input for musical systems, which informs the research question explored in this study.

Visual-auditory relationships in art

Relationships between visual and auditory media have been explored in artistic practice for over a century. Early twentieth century painter Wassily Kandinsky proposed that visual elements such as colour and form could function analogously to musical harmony and rhythm. In *Concerning the Spiritual in Art*, Kandinsky (1911) argued that colour possesses expressive qualities capable of evoking emotional responses similar to those produced by music. Based on philosophical interpretation rather than empirical evidence, this position has limited reliability as a framework for understanding consistent audiovisual relationships. However, as a conceptual foundation it remains significant, establishing a basis for thinking about connections between sensory modalities that subsequent cognitive science research has sought to test and substantiate empirically. The value of Kandinsky's contribution therefore lies not in its reliability as evidence, but in its role in framing questions that later research has attempted to answer through more rigorous means.

Systems for representing music visually have a long history, though the relationship between visual notation and musical precision has evolved considerably over time. As Goodall (2013) notes, early notation was highly interpretive and imprecise, with greater consistency and exactness developing gradually through centuries of refinement. This trajectory towards precision reflects an assumption that visual information can reliably encode musical intent, an assumption that is directly challenged by graphic scores such as *Treatise* (Cardew, 1967), which deliberately resist fixed interpretation by using abstract visual forms as a starting point for performance rather than a set of instructions. Performances of *Treatise* vary significantly despite sharing the same visual material, demonstrating that visual-auditory relationships are not fixed but constructed through interpretation. For the purposes of this research, this raises an important question: if even formalised visual systems produce variable outcomes, what are the implications for using visual information as a direct input for generative musical systems?

The tension between structure and interpretation is addressed in later work through generative systems. Scoates (2013) notes that Brian Eno's audiovisual practice extends earlier visual music traditions by creating environments in which sound and image evolve through rule-based processes rather than fixed compositions. Crucially, this represents more than a stylistic shift. It suggests that the variability identified in interpretive approaches can be managed through intentional system design, replacing assumed mappings with structured rules that shape how audiovisual relationships emerge.

Contemporary practitioners continue to develop these approaches using modern digital tools. Artists such as Pauric Freeman generate audiovisual works in which visual output is derived from sound through shared computational processes (Freeman, 2021), demonstrating how generative technologies can produce increasingly complex relationships between visual and musical media. However, Freeman's practice typically operates by translating audio into visual form rather than using visual information to generate sound, highlighting that while the technical and conceptual groundwork exists, the reverse direction remains largely unexplored in contemporary practice.

Cross-modal perception and synaesthesia

Research into relationships between sound and visual perception has emerged within cognitive science and neuroscience, particularly through studies of synaesthesia and cross-modal perception. Synaesthesia refers to a neurological phenomenon in which stimulation of one sensory modality automatically triggers experiences in another. Cytowic and Eagleman (2009) describe this as a condition in which individuals may perceive colours when hearing sounds or associate pitches with visual forms, a phenomenon documented in mainstream media including a BBC Horizon documentary on lexical-gustatory synaesthesia (BBC, 2004). However, synaesthesia represents an extreme and atypical form of cross-modal experience, and broader cross-modal associations are of greater relevance to the present research.

Marks (1975) proposed that humans naturally form correspondences between sensory modalities, linking auditory characteristics such as pitch or loudness with visual qualities including brightness or spatial height. However, a critical question remains as to whether these associations are innate or culturally learned. If primarily the product of cultural conditioning, their reliability as a universal basis for generative system design is significantly weakened, as systems built around such associations may only function meaningfully for audiences sharing similar cultural frameworks.

Levitin (2006) addresses this directly, arguing that associations between sensory domains may arise through both biological factors and learned cultural patterns, meaning no single mapping between visual and auditory elements can be assumed to hold universally. The effectiveness of visual-driven generative systems must therefore depend not on perception alone, but on the quality of the design frameworks through which visual information is interpreted and translated into sound, a conclusion that directly informs the research question explored in this study.

Generative music systems and audiovisual mapping

Generative music refers to music produced by systems that generate musical material through rules or processes rather than direct human composition. Eno (1996) describes generative music as continuously evolving, produced by systems capable of generating potentially infinite variations from a defined set of parameters. Nierhaus (2009) defines the broader field of algorithmic composition as the use of formalised computational systems to generate musical structures, within which non-musical data such as images, movement or environmental variables can be mapped to parameters including pitch, rhythm, timbre or dynamics. While these definitions establish the technical scope of generative practice, they do not address whether systems that generate variation from defined parameters necessarily produce meaningful or musically coherent outcomes.

The mapping of visual information to sound has been explored through digital music technologies. Puckette (2007), creator of Max and Pure Data, describes these environments as flexible platforms for building interactive systems that process and translate data streams in real time, enabling visual characteristics such as pixel brightness or motion to be translated into MIDI or audio through algorithmic mapping. However, such mappings do not in themselves guarantee meaningful or musically coherent outcomes, as relationships between data and sound must be defined through system design. Henke (2009), composer and co-creator of Ableton Live, reflects this in describing his approach as writing structures that then create the music, suggesting that generative systems function less as autonomous creative agents and more as frameworks requiring intentional design and ongoing human input. This positions the designer rather than the system as the primary creative agent, a perspective that has significant implications for how visual-to-audio mapping should be approached.

Research gap and rationale

Existing literature has examined visual music traditions, theories of cross-modal perception, and computational systems capable of linking audiovisual processes, yet much of this work focuses either on how audiovisual relationships are perceived or on the visualisation of musical structures. Less attention has been given to the use of visual stimulus as a direct generative source for musical material in ways that account for perceptual variability and the role of system design.

This gap is particularly relevant within contemporary music technology environments such as Max, which allow real-time translation of visual data into musical parameters. While the technical capability exists, the extent to which visual information can meaningfully and consistently drive generative music systems remains underexplored, and no established framework exists for navigating the relationship between perceptual variability and system design in this context.

This research is also personally motivated. As a practitioner working at the intersection of generative systems, audiovisual performance and installation, the question of how visual information can drive musical output is not purely academic but directly relevant to the development of a creative practice. Understanding how cross-modal relationships can be implemented within generative systems provides a basis for making more informed and intentional decisions about system design, and for situating that practice within a broader field of research and professional activity.

This study therefore represents both an academic investigation and a form of practice-based inquiry, in which the research question is inseparable from the creative work it informs.

Methodology

This study adopts a mixed methods approach, combining qualitative and quantitative methods to investigate how visual stimuli can be used to drive generative music systems (Bryman, 2016). This approach was selected because the research question requires both measurable data on perceptual tendencies and deeper qualitative insight into how these tendencies translate into system design and audience experience.

Cross-modal perception survey

A short online survey explored whether consistent relationships exist in how individuals perceive connections between sound and visual stimuli, using multiple-choice and Likert scale questions (Likert, 1932) alongside limited open responses. This method was selected for its ability to generate comparable data across multiple participants, which qualitative methods alone could not provide. A repeated measure was included to assess internal consistency.

Practitioner interview

A semi-structured interview was conducted with a practitioner working with generative audiovisual systems, selected for the direct relevance of their practice to the research question. Open-ended questions captured reflective responses relating to system design, perception, and audience experience, analysed through thematic coding.

Focus group

A focus group explored how participants engage with and interpret generative audiovisual systems in an interactive setting, using a series of Max-based devices demonstrating different approaches to visual-to-audio mapping. Thematic coding identified patterns across responses.

Practice based experimentation

A series of iterative experiments explored different methods of translating visual data into musical parameters using Max, examining both technical feasibility and perceptual outcomes. This approach aligns with practice-as-research methodologies (Nelson, 2013), allowing knowledge to be generated through making and reflection, and was essential to the research as the question requires the creation and evaluation of working systems.

Ethics, trustworthiness and limitations

Ethical considerations were addressed throughout the research process in line with established guidance (Bell and Waters, 2018), with ethics committee approval obtained prior to data collection (Appendix K). Participation was voluntary, informed consent was obtained for the interview and focus group, and all responses were stored anonymously. The survey included questions about personal sensory experiences including synaesthesia, which carries greater sensitivity than standard survey research, and care was taken to ensure questions were unlikely to cause distress.

A further ethical consideration is the researcher's own experience of synaesthesia and neurodivergence, which informed the design of the research but introduces potential bias in the interpretation of findings. This positionality is acknowledged as both an asset and a limitation. Lived experience of cross-modal perception brought an informed perspective to the research but may have unconsciously shaped which outcomes were found compelling. The use of multiple methods was partly intended to mitigate this, grounding interpretations in a range of data sources rather than the researcher's perspective alone.

The use of multiple methods supports triangulation across survey data, participant discussion, practitioner insight, and practice-based experimentation, strengthening validity by reducing reliance on a single data source. Repeated questions in the survey support internal consistency, and thematic coding across qualitative methods provides a structured approach to analysis, though it is acknowledged that qualitative analysis is inherently interpretive.

Several limitations remain. The self-selecting survey sample of 70 participants likely skewed towards those with existing interest in music or sensory experience, and the cultural backgrounds of participants were not recorded, a significant gap given that cross-modal associations are partly culturally determined. The practitioner interview represents a single perspective from someone working primarily in the opposite direction to the research question. The focus group consisted of five participants in a controlled setting unlikely to reflect real installation or performance contexts. Finally, all experiments were conducted within a single technical environment, and none were tested at scale or with culturally diverse audiences, limiting the generalisability of the findings.

Findings and analysis

This section analyses findings from survey data, interview responses, focus group discussion, and practice-based experimentation.

Consistency and acceptance of cross-modal associations

Survey results (Appendix G) suggest a greater degree of consistency in cross-modal associations than initially expected. When asked to select the image best representing a given sound, 45.71% of participants selected the same option. With five available options, chance alone would predict 20% agreement, meaning the observed consistency is more than double what random response would produce. Similarly, 58.57% described the perceived size of a sound as "big", supported by Marks (1975), who proposes that auditory characteristics such as pitch and intensity are naturally associated with visual qualities including brightness and size. Levitin (2006) further argues such associations arise from both biological processes and learned cultural patterns, resulting in tendencies that are widely shared but not universally fixed.

An important caveat is that sample composition may partly explain this consistency. Survey results indicate that 25.71% of respondents identified as experiencing synaesthesia, with a further 30% responding "maybe", suggesting up to 55% of the sample may have had some form of cross-modal perceptual experience. Given that synaesthesia affects only around 4% of the general population (Cytowic and Eagleman, 2009), this represents a significantly skewed sample, and the observed consistency may reflect group composition rather than universal tendencies.

Despite this, participants appear broadly comfortable engaging with cross-modal associations regardless of synaesthetic experience. A majority reported their choices felt natural and intuitive (34.29%) or that they felt reasonably confident in their decisions (32.86%), suggesting the concept of linking sound with visual elements is widely accepted even where specific associations differ. Focus group findings (Appendix H) corroborate this, with participants describing movement-based interaction as immediately intuitive and quickly influencing system output, demonstrating in a live context the same willingness to engage with cross-modal relationships identified in the survey. Together these findings suggest that while cross-modal mappings may not be universally consistent, users engage with them meaningfully, providing a viable basis for generative system design.

Translation of visual data into musical systems

Early experiments demonstrated that direct mappings between visual and sonic elements are limited by both subjectivity and technical implementation. In Experiment I (Chroma) (Appendix A), visual input was reduced to dominant RGB values and mapped to harmonic groupings. While functional, this reduction of complex visual information to simplified numerical data limited nuance, and latency affected responsiveness, suggesting that direct one-to-one mappings risk losing perceptual richness through abstraction, producing systems that are functional but not necessarily meaningful.

Experiment III (Motion) (Appendix B) produced a more immediately intuitive relationship, with physical movement detected through frame differencing and translated into amplitude and timbral changes. Focus group responses (Appendix H) support this, with participants describing movement-based interaction as easy to understand and quickly influencing the output, suggesting mappings that preserve a sense of physical causality are more readily interpreted as meaningful. However, complex physical gestures were still reduced to a single control value, demonstrating that even seemingly direct mappings rely on abstraction and that perceived intuitiveness does not necessarily reflect the complexity of the underlying system.

Later experiments shifted towards structured generative frameworks. In Experiment VI (Pulse) (Appendix E) and Experiment VII (Evolve) (Appendix F), visual input influenced system behaviour through processes such as probability, constraint, and internal modulation, allowing for greater coherence and musicality. However, focus group participants expressed uncertainty about whether visual input was actually driving the sound in Evolve (Appendix H), questioning whether video was influencing the outcome or vice versa. This suggests that while structured generative frameworks produce more musically coherent results, they risk obscuring the visual-to-audio relationship, raising questions about how transparency of mapping should be balanced against musical quality in system design.

Experiment IV (System) (Appendix C) treated the visual environment as a performable instrument, while Experiment V (Treatise) (Appendix D) demonstrated how visual structures can function as interpretive frameworks rather than data sources. Together these experiments suggest that the relationship between visual input and sonic output exists on a continuum from transparent and direct to complex and mediated, and that position on this continuum has significant implications for both perceived coherence and the audience's ability to engage with the visual-to-audio relationship.

These findings suggest that the translation of visual data into musical systems is not a neutral technical process but a series of design decisions that shape what the relationship between image and sound can mean, with meaningful audiovisual relationships emerging from design choices rather than perceptual tendencies alone.

Human interpretation and system design

The findings suggest that meaningful and resonant results depend on ongoing human intervention in the design, refinement, and interpretation of the system. This challenges the assumption that autonomy is an inherent property of generative systems. The evidence suggests instead that apparent autonomy is a product of careful human design, emerging from the rules and constraints the designer establishes rather than from the system itself.

Practice-based experiments demonstrated that systems require careful calibration to produce coherent outcomes. In Experiment I (Chroma) (Appendix A), early design choices directly constrained the expressive range of the system, while later experiments such as Experiment VI (Pulse) (Appendix E) and Experiment VII (Evolve) (Appendix F) required iterative adjustment of parameters including scaling, smoothing, and gain staging to achieve stable and musically engaging results. This mirrors the principle identified in the literature review through Eno's practice, understood not as the delegation of creativity to a system but as the establishment of conditions within which musical behaviour can emerge, a principle the experiments here suggest applies equally to visual-driven generative systems.

This is supported by established practice within the field. Henke (2009) describes his approach as writing structures that then create the music, aligning closely with the iterative, design-led processes observed throughout this research. The practitioner interview (Appendix J) similarly describes establishing initial rules before refining parameters and characterises cross-modal perception as occurring when certain conditions are met, conditions the designer is responsible for creating. Both perspectives reflect a consistent understanding that meaningful outcomes emerge through a balance of structured processes and intuitive decision making rather than from system behaviour alone.

Focus group findings (Appendix H) further support this view. Participants described the systems as intuitive despite not fully understanding their underlying mechanisms but identified limitations in depth and complexity across all three demonstrations. This suggests that the perceived quality ceiling of these systems is set not by the technology itself but by the level of design investment applied, and that the expressive potential of visual-driven generative systems is directly proportional to the skill and iterative effort of the designer rather than being a fixed property of the approach.

These findings indicate that meaningful audiovisual relationships depend on human involvement at multiple stages. Visual stimuli can function as generative inputs, but the quality of the outcome is determined by the human decisions that govern how that input is captured, interpreted and transformed into sound. The answer to the research question is therefore not simply that cross-modal relationships can drive generative systems, but that they can do so effectively only when the designer takes an active role in shaping how perceptual tendencies are translated into system behaviour.

Professional context and application

The application of cross-modal generative systems extends beyond technical experimentation into professional contexts, introducing legal responsibilities that directly shape how systems can be designed and deployed. Where systems incorporate live video capture, the UK General Data Protection Regulation (UK GDPR, 2021) requires individuals to be informed of how their image is being processed, a genuine challenge in installation contexts where audiences may not realise their image is being used as a generative input. Where systems draw on pre-existing audio material, appropriate licensing must be in place under the Copyright, Designs and Patents Act (1988), as algorithmic use of sampled material does not exempt practitioners from copyright obligations.

Focus group findings (Appendix H) identified potential applications for SEND learners, where the intuitive and responsive nature of these systems may support sensory exploration and accessible engagement in ways conventional instruments cannot, reflecting a broader characteristic of visual-driven generative systems identified throughout this research:

that their value lies partly in their accessibility and low threshold for meaningful interaction.

The professional deployment of cross-modal generative systems therefore requires designers to consider not only how visual information can drive musical output, but how the resulting systems will function within contexts shaped by legal, ethical and institutional frameworks. The design decisions that determine how a system captures, processes and translates visual input are inseparable from the professional responsibilities that govern how those systems can be used.

Conclusion

This study investigated how cross-modal relationships between visual and auditory perception can be used to drive generative music systems, drawing on literature from visual art, cognitive science and music technology alongside primary research comprising a survey, focus group, practitioner interview and practice-based experimentation using Max/MSP.

The findings reveal several interconnected conclusions. Survey data indicated a greater degree of consistency in cross-modal associations than initially anticipated, with observed agreement more than double what chance alone would predict. However, this consistency requires qualification — the sample contained a disproportionately high number of participants with synaesthetic experience, and findings may reflect group composition rather than universal tendencies. Despite this, participants across both the survey and focus group demonstrated a broad willingness to engage with cross-modal relationships meaningfully, suggesting they provide a viable conceptual foundation for generative system design. Practice-based experiments revealed that the relationship between visual input and sonic output exists on a continuum from transparent and direct to complex and mediated, and that apparent autonomy in generative systems is itself a product of human design. Meaningful outcomes emerged not from system behaviour alone but through iterative calibration, parameter refinement and ongoing human interpretation, consistent with both established practitioner approaches and focus group responses.

How can cross-modal relationships between visual and auditory perception be used to drive generative music systems?

The evidence suggests that visual stimuli can function effectively as generative inputs, but that the quality of the outcome is directly proportional to the designer's skill and iterative effort. The most effective systems accommodate perceptual variability through intentional design, allowing relationships to emerge through the interaction between visual input, system behaviour and human engagement. Generative autonomy is not an inherent property of the technology, but a condition created by the designer.

Several limitations qualify these conclusions. The relatively small and culturally homogeneous sample limits generalisability, and all experiments were conducted within a single technical environment without testing at scale or with diverse audiences. Future research could explore how these systems perform with larger and more culturally diverse participants, how different types of visual input shape generative output, and how transparency of mapping affects audience experience across different contexts.

As these systems move into broader professional contexts, practitioners must engage with the legal and ethical frameworks that govern their use. The tension between artistic intent and legal obligation around live video capture and data protection illustrates that design decisions are inseparable from professional responsibilities. Considerations around copyright, accessibility and data protection are not peripheral to system design but integral to responsible professional practice.

References

Arts Council England (2026) 'Digital Arts becomes Arts Council England's tenth supported artform'. Available at: https://www.artscouncil.org.uk/blog/digital-arts-becomes-arts-council-englands-tenth-supported-artform (Accessed: 19 April 2026).

BBC (2004) *'Derek Tastes of Earwax'*, Horizon, BBC Two, 30 September.

Bell, J. and Waters, S. (2018) *Doing Your Research Project: A Guide for First Time Researchers*. 7th edn. Maidenhead: Open University Press.

Bryman, A. (2016) *Social Research Methods*. 5th edn. Oxford: Oxford University Press.

Cardew, C. (1967) *Treatise*. London: Edition Peters.

Cytowic, R.E. and Eagleman, D.M. (2009) *Wednesday is Indigo Blue: Discovering the Brain of Synaesthesia*. Cambridge, MA: MIT Press.

Copyright, Designs and Patents Act 1988, c.48. Available at: https://www.legislation.gov.uk/ukpga/1988/48/contents (Accessed: 19 April 2026).

Eno, B. (1996) *'Generative Music', In Motion Magazine*, 8 June. Available at: https://www.inmotionmagazine.com/eno1.html (Accessed: 8 April 2026).

Equality Act 2010, c.15. Available at: https://www.legislation.gov.uk/ukpga/2010/15/contents (Accessed: 19 April 2026).

Freeman, P. (2021) *'Interview with Pauric Freeman', Generative Hut*, 22 August. Available at: https://www.generativehut.com/post/interview-with-pauric-freeman (Accessed: 8 April 2026).

Goodall, H. (2013) *The Story of Music*. London: Vintage.

Henke, R. (2009) *'An interview with Robert Henke'*, Cycling '74, 22 November. Available at: https://cycling74.com/articles/an-interview-with-robert-henke (Accessed: 19 April 2026).

Kandinsky, W. (1911) *Concerning the Spiritual in Art*. Munich: Piper Verlag. (English translation: Kandinsky, W. (1977) *Concerning the Spiritual in Art*. New York: Dover.)

Levitin, D.J. (2006) *This Is Your Brain on Music: The Science of a Human Obsession*. New York: Dutton.

Likert, R. (1932) *A Technique for the Measurement of Attitudes. Archives of Psychology*.

Marks, L.E. (1975) 'On Coloured-Hearing Synaesthesia: Cross-Modal Translations of Sensory Dimensions', *Psychological Bulletin*, 82(3), pp. 303–331.

Nelson, R. (2013) *Practice as Research in the Arts: Principles, Protocols, Pedagogies, Resistances*. Basingstoke: Palgrave Macmillan.

Nierhaus, G. (2009) *Algorithmic Composition: Paradigms of Automated Music Generation*. Vienna: Springer.

Puckette, M. (2007) *The Theory and Technique of Electronic Music*. Singapore: World Scientific.

Scoates, C. (2013) *Brian Eno: Visual Music*. San Francisco: Chronicle Books.

UK General Data Protection Regulation 2021. Available at: https://www.legislation.gov.uk/eur/2016/679 (Accessed: 19 April 2026).

Ward, J. (2013) *The Student's Guide to Cognitive Neuroscience*. 3rd edn. Hove: Psychology Press.

Appendix

Appendix A: Experiment I (Chroma)

This experiment explored the use of colour as a control input for harmonic change within a generative system. A webcam was used to analyse RGB values from a controlled light source, which were mapped to predefined harmonic groupings and triggered as MIDI output. The system demonstrated a functional relationship between visual input and musical output, while highlighting challenges around subjectivity and responsiveness.

Full documentation available at:
https://isomorphicmusic.com/diary/experiments/

Appendix B: Experiment III (Motion)

This experiment explored how movement within a live video feed could function as a control source for sound. Motion was detected through frame differencing and mapped to amplitude and filter parameters, allowing physical movement to shape the intensity and timbre of the output. The system demonstrated responsive interaction, while highlighting limitations of small-scale testing environments for simulating real world installation contexts.

Full documentation available at:
https://isomorphicmusic.com/diary/experiments/

Appendix C: Experiment IV (System)

This experiment explored the development of a playable audiovisual instrument, where visual parameters could be shaped in real time through MIDI control. Changes in form, rhythm and opacity were mapped to corresponding audio parameters, creating a unified system for performance. The experiment demonstrated the potential of treating visual systems as instruments, while identifying limitations in visual quality and system depth.

Full documentation available at:
https://isomorphicmusic.com/diary/experiments/

Appendix D: Experiment V (Treatise)

This experiment explored the interpretive performance of a graphic score, using visual elements from Cornelius Cardew's *Treatise* as a framework for generating sound. Musical structure was derived from visual features such as spacing and density, supported by a custom Max device for controlled pitch transitions. The experiment highlighted the role of human interpretation alongside system-based approaches, revealing a balance between structure and subjectivity.

Full documentation available at:
https://isomorphicmusic.com/diary/experiments/

Appendix E: Experiment VI (Pulse)

This experiment explored how constrained generative systems can produce evolving rhythmic structures through controlled randomness. A central pulse with probabilistic subdivisions triggered discrete visual and sonic events, while a sustained harmonic layer provided stability. The system demonstrated how constraint and probability can be combined to create coherent and performable generative instruments.

Full documentation available at:
https://isomorphicmusic.com/diary/experiments/

Appendix F: Experiment VII (Evolve)

This experiment explored the use of multiple video sources as control inputs for ambient sound generation. Brightness data from six video streams was mapped to the amplitude of corresponding audio voices, creating a continuously evolving soundscape. The system demonstrated how simple visual mappings, combined with slow modulation, can produce coherent and immersive generative environments.

Full documentation available at:
https://isomorphicmusic.com/diary/experiments/

Appendix G: Cross-modal perception survey results

Question 1: Listen to the audio sample below and select the image that you feel best represents the sound.

Answer	Total	%
Two	32	45.71
Five	13	18.57
Four	10	14.29
Three	9	12.89
One	6	8.57

Question 2: Listen to the audio sample below and select the colour that you feel best represents the sound.

Answer	Total	%
Blue	20	28.57
Violet	17	24.29
Green	13	18.57
Yellow	12	17.14
Red	8	11.43

Question 3: Listen to the audio sample below and select the motion that you feel best represents the sound.

Answer	Total	%
Right	23	32.86
Down	18	25.71
Up	13	18.57
Left	10	14.29
Still	6	8.57

Question 4: Listen to the audio sample below and select the size that you feel best represents the sound.

Answer	Total	%
Big	41	58.57
Average	14	20.00
Huge	10	14.29
Small	4	5.71
Tiny	1	1.43

Question 5: Which sound feels like it fits this image.

Answer	Total	%
Sound A	31	44.29
Sound C	21	30.00
Sound B	18	25.71

Question 6: When you listen to music or sounds, do you ever experience colours in your mind?

Answer	Total	%
Sometimes	23	32.86
Never	20	28.57
Rarely	15	21.43
Often	11	15.71
Always	1	1.43

Question 7: Which of the following best describes your relationship with music?

Answer	Total	%
I am an amateur musician or regularly play music	30	42.86
I enjoy listening to music but do not play an instrument	30	42.86
I am a professional or working musician	8	11.43
I have little involvement with music	2	2.86

Question 8: You heard this sound earlier in the survey. Please listen again and select the colour that best represents the sound based on your current impression.

Answer	Total	%
Blue	22	31.43
Yellow	16	22.86
Violet	16	22.86
Green	9	12.86
Red	7	10.00

Question 9: Based on this description, do you feel you may have or experience Synaesthesia?

Answer	Total	%
No	31	44.29
Maybe	21	30.00
Yes	18	25.71

Question 10: When answering the questions in this survey, how did your choices generally feel?

Answer	Total	%
My choices felt natural and intuitive	24	34.29
I felt reasonably confident in most of my choices	23	32.86
I sometimes felt unsure	13	18.57
I mostly felt like I was guessing	9	12.86
I wasn’t sure how to decide between options	1	1.43

Appendix H: Focus group summary

A focus group was conducted with five participants in person over a 45-minute session. Participants were selected based on their relevance to the research topic and included a mix of musical experience. The session involved demonstrations of three Max devices, each followed by guided discussion using open ended questions.

Demo 1: Motion

Participants interacted with a system in which their movement, captured via webcam, controlled the amplitude and filter cutoff of an oscillator.

Guiding questions:

- How intuitive did it feel to interact with the device?
- To what extent did you feel you could 'play' the instrument?

Demo 2: System

Participants used a MIDI controller to generate visual forms, which were mapped to corresponding audio output.

Guiding questions:

- To what extent did you feel you could 'play' the instrument?
- How would you describe the limitations of the device?

Demo 3: Evolve

Participants observed a system in which brightness data from six video sources controlled the behaviour of an evolving ambient soundscape.

Guiding questions:

- To what extent did you feel that the demonstration was musical?
- To what extent did you feel that the video was influencing the outcome?

Theme	Description	Observations
Engagement and novelty	Participants found the devices interesting and unique, expressing curiosity and engagement during interaction	Described as "interesting" and "different"; strong initial engagement across all demos
Intuitive interaction	Participants generally found the systems intuitive to use, particularly where interaction was direct (e.g. motion)	Movement based interaction described as easy to understand; participants able to quickly influence output
Perceived playability	Participants felt able to "play" the systems, particularly where cause and effect relationships were clear	Motion device described as more performable; System device seen as controllable but more constrained
Understanding vs technical complexity	Participants did not always understand the underlying processes, but were able to interpret outcomes and interact effectively	Questions raised about how systems worked, but interaction remained confident and purposeful

Role of visual input	Participants questioned the extent to which visual input was driving the sound, particularly in more abstract systems	Some uncertainty around whether video was influencing sound or vice versa, especially in Evolve
Perceived limitations	Systems were seen as somewhat limited in scope, often producing relatively simple or repetitive outputs	Comments around lack of depth, variation, or control; desire for more complexity
Purpose and application	Participants questioned the necessity of the systems in a traditional musical context	Some uncertainty around why such systems would be used compared to conventional instruments
Educational potential	Participants identified potential for use in educational or therapeutic contexts, particularly for SEND users	Suggested value for sensory exploration, accessibility, and engagement through interaction

Appendix J: Practitioner interview

A practitioner interview was conducted with an artist working with Max-based audiovisual systems, whose practice focuses on generating visuals from musical input rather than using visual data to drive sound. The participant was selected due to the relevance of their work to the research topic, providing a contrasting perspective on audiovisual mapping.

Informed consent was obtained prior to the interview. The participant agreed for their responses to be included and analysed within this report on the condition of anonymity.

The interview was conducted using open-ended questions, with responses recorded and subsequently analysed thematically. Responses are presented below and have been cleaned for clarity.

Can you describe your creative practice and how sound and visuals feature within it?

I work as a multidisciplinary artist across audiovisual performance and installation. My practice centres on developing custom tools to generate real time compositions from sonic data, where sound and visuals operate as interconnected systems rather than separate elements. These systems produce evolving structures characterised by dense harmonic textures and complex rhythmic behaviour, with visual and sonic outputs influencing and reshaping each other over time. This process allows me to reframe how sound is experienced, shifting audience perception through the interaction between image and audio.

In your work, do sound and visuals operate as separate elements, or as part of a unified system? How would you describe their relationship?

I'm primarily interested in human perception, particularly how we process and combine sensory information. My work explores cross-modal perception, where the brain integrates visual and auditory input into a single experience when certain conditions are met. This means that sound and visuals are not treated as separate elements, but as part of a unified system shaped by how they are perceived together.

I am particularly interested in how flexible this process is, and how it can be shaped through carefully selected data and sensory input. This is something that can be seen in areas such as cinema, where sound is used not just to represent what is seen, but to influence emotion, meaning, and perception more broadly.

My own practice approaches this through the use of synthesis and generative graphics, exploring how abstract audiovisual systems can shape perception and create unified experiences.

To what extent is your work shaped by systems, rules, or generative processes, compared to intuitive or human decision making?

My work is shaped by a combination of structured systems and intuitive decision making. I use control voltage and audio data from a modular synthesiser, which is routed into Touch Designer to drive visual processes. Within this environment, I establish a set of initial rules, often using simple visual forms that align with the characteristics of the sound.

The incoming data is then used to influence and develop these visuals over time, creating more complex behaviour. While the system provides the underlying structure, a significant amount of refinement is required, including adjusting parameters and processing data to achieve a result that fits the intended aesthetic. This process reflects a balance between predefined rules and ongoing human input.

How do audiences typically engage with your work? Is it important that they understand the relationship between sound and visuals, or is the experience itself more significant?

I try not to focus too heavily on the audience, and instead concentrate on my own practice and process. For those who experience the work, it is primarily about the overall experience, rather than understanding the technical relationship between sound and visuals. Ideally, the work is perceived as a unified whole, where it is difficult to separate one element from another.

Alongside this, I also work in educational contexts, where the technical aspects are explored more explicitly, and the relationship between sound and visuals is more clearly defined and understood.

How has your practice developed over time, and what key influences have shaped your approach to sound and visuals?

I initially came to media technology through music, beginning with DJing and later moving into synthesis and sound production. I had limited exposure to visual art early on, but this changed when I began studying, where I was introduced to interaction design, programming, and working with input devices. During this time, I also explored areas such as abstract animation, visual music, psychoacoustics, and musique concrète, which led to a growing interest in visualising sound and understanding audiovisual perception.

Following my studies, I spent time working internationally, where I continued to experiment with tools such as Processing, MaxMSP, and MIDI data to generate real time visuals from sound. This period involved both installation work and live audiovisual performance, and marked a shift towards developing systems that could support these practices. Over time, this led to a focus on creative development, with an emphasis on building custom tools and systems to support audiovisual work.

Appendix K: Ethics committee approval

WB Ethics Form Submission Approval

RE

Research Ethics Committee <Ethics@...

Sunday, 15 March 2026 at 10:40

To: Dominic Goodeve

Thank you for submitting your ethics application form. Based upon the answers you gave, the WaterBear Research Ethics Committee would like to confirm that your project has been assessed as low risk and duly approved. You may now begin your research.

Please note that this approval only applies to your study as described in the application and that any further changes to the study design must be disclosed to the Ethics Committee prior to any research being conducted.

If you need to inform us of any changes or have any further questions, please contact us on Ethics@waterbear.org.uk.