



OPEN ACCESS

EDITED BY

Steven Brown,
McMaster University, Canada

REVIEWED BY

Brigitte Röder,
University of Hamburg, Germany
David Howes,
Concordia University, Canada

*CORRESPONDENCE

Charles Spence
✉ charles.spence@psy.ox.ac.uk

RECEIVED 02 April 2025

ACCEPTED 05 June 2025

PUBLISHED 14 July 2025

CITATION

Spence C and Di Stefano N (2025)
Augmenting art crossmodally: possibilities and
pitfalls. *Front. Psychol.* 16:1605110.
doi: 10.3389/fpsyg.2025.1605110

COPYRIGHT

© 2025 Spence and Di Stefano. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Augmenting art crossmodally: possibilities and pitfalls

Charles Spence ^{1*} and Nicola Di Stefano ²

¹Crossmodal Research Laboratory, Department of Experimental Psychology, Oxford University, Oxford, United Kingdom, ²Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

In this narrative historical review, we take a closer look at the question of whether it is possible to augment works of art through crossmodal (specifically audiovisual) means. We start by highlighting an important distinction between three classes of audiovisual crossmodal correspondence: Namely those operating on individual sensory stimuli (so-called basic correspondences), those operating on dynamically-changing stimuli, or else on combinations of unisensory stimuli (so-called mid-level correspondences), and those operating on complex and often aesthetically-meaningful stimuli, such as music and paintings. We also highlight another important distinction between the literature on crossmodal matching and that dedicated to demonstrating crossmodal effects. The latter distinction aligns, in some sense, onto the distinction between crossmodal mapping and crossmodal effects. Although it may not be possible, in any meaningful sense, to translate works of art from one modality into another, that does not deny the possibility of augmenting a work of art by the deliberate addition of stimulation presented to another sensory modality. The aims and objectives of those who have attempted to augment works of art by introducing additional sensory stimulation are discussed. We also draw attention to a number of challenges and/or pitfalls (such as the distraction offered by recourse to the phenomenon of synaesthesia) for those interested in augmenting auditory/visual art crossmodally.

KEYWORDS

art, audiovisual, crossmodal, Gestalt, intersensoriality, multisensory, sensory translation, sensory augmentation

1 Introduction

In recent decades, there has been a growing awareness within the cognitive neurosciences that the senses do not operate in isolation (Barlow and Mollon, 1982), and that crossmodal interactions are the norm, not the exception (see Calvert et al., 2004).¹ Crossmodality refers to the way in which the stimuli presented in one sense may influence people's perception of the stimuli that happen to be presented in another sense. Crossmodality is different from the term multisensory, with the latter referring to those situations when the senses come together to deliver a multisensory representation of an object or event (Stein et al., 2010). Classic examples of multisensory perception include everything from audiovisual speech perception to multisensory flavor perception. Crossmodal effects, by contrast, are more often observed when seemingly unrelated

¹ As Stein et al. (1996) note in one of their articles: "there is no animal in which there is known to be a complete segregation of sensory processing".

unisensory stimuli are presented at around the same time, such as when the presentation of a bright light makes a simultaneously-presented sound appear louder than it otherwise would (see [Spence, 2018](#), for a review).

One of the most exciting recent areas in crossmodal research concerns the crossmodal correspondences. This is the name given to the almost synaesthesia-like surprising connections between features, attributes, or dimensions of experience presented in different senses. Often confused with synaesthesia (in that both phenomena are surprising when first you hear about them), they are fundamentally different (see [Deroy and Spence, 2013](#)). For instance, there is typically a crossmodal association but no inducer-concurrent mapping in the case of crossmodal correspondences. Furthermore, crossmodal correspondences tend to be consensually shared across groups of people, whereas the inducer-concurrent mapping in the context of synaesthesia is, by definition, idiosyncratic between one synaesthete and the next. It is the crossmodal correspondences that we are interested.

In this narrative historical review, we first introduce an important distinction between three classes of audiovisual crossmodal correspondence: Namely those operating on individual sensory stimuli (so-called basic correspondences), those operating on dynamically-changing stimuli, or else on combinations of unisensory stimuli (so-called mid-level correspondences), and those operating on complex and often aesthetically-meaningful stimuli, such as music and paintings (Section 1.1). We also highlight an important distinction between the literature on crossmodal matching and that dedicated to demonstrating crossmodal effects. The latter distinction aligns, in some sense, with the distinction between crossmodal mapping and crossmodal effects. In Section 1.2, we briefly consider the kinds of explanations that have been put forward to account for crossmodal correspondences to date. Thereafter, in Sections 2–4, we review the evidence concerning audiovisual crossmodal matching and crossmodal effects operating at the level of basic, mid-level, and complex crossmodal correspondences, respectively.

In Section 5, we build on this groundwork by reflecting on the putative existence of crossmodal art. We take a closer look at the question of whether it is possible to augment works of art through crossmodal (specifically audiovisual) means. We show how the phenomenon of synaesthesia, at least as conceptualized by cognitive neuroscientists fails to enlighten theorizing when it comes to thinking about the social/structural construction of meaning in the context of the crossmodal augmentation of art ([Dimova, 2024](#); [Lévi-Strauss, 1997](#); [Spence and Deroy, 2013](#)). We discuss how although it may not be possible, in any meaningful sense, to translate works of art from one modality into another (see [Spence and Di Stefano, 2024a](#)), that does not deny the possibility of augmenting a work of art by the deliberate addition of stimulation presented to another sensory modality. Finally, we discuss the aims and objectives of those who have attempted to augment works of art by introducing additional sensory stimulation. The intriguing notion of intersensoriality is also outlined. This term has been introduced by [Howes \(2025\)](#) to refer to art that crosses sensory borders ([Classen, 1998](#)). We draw attention to a number of the challenges and/or pitfalls for those interested in wanting to augment auditory/visual art crossmodally (Section 6). We

conclude, in Section 7, by challenging the notion, implicit in much creative practice, that elemental correspondences (that is, crossmodal correspondences based on simple stimulus properties), provide the most relevant insights when it comes to augmenting art crossmodally, given that works of art are mostly complex, multi-elemental, and rich in semantic and emotional meaning.

1.1 Basic, mid-level, and complex crossmodal correspondences: a novel typology

Audiovisual crossmodal correspondences can be grouped into different classes based on the (perceived) complexity of the stimuli involved. In particular, while some researchers have chosen to study crossmodal correspondences between basic (i.e., individual) sensory stimuli (i.e., specific features or dimensions), others have chosen to focus on crossmodal correspondences (and Gestalt grouping) in the case of what might be called mid-level correspondences instead (i.e., those correspondences that are experienced between combinations of unisensory stimuli, such as an auditory melody and a sequence of visual stimuli).² A third group of studies, meanwhile, has focused on crossmodal correspondences involving more complex and often semantically-meaningful stimuli, such as, for example, works of art (primarily paintings and classical music) and other popular entertainment media (such as film clips, TV adverts, and popular music selections). While the majority of the research that has been published to date has tended to focus on those correspondences that can be demonstrated between stimuli at the same level of complexity, a number of studies have also investigated crossmodal correspondences operating across levels (as when color properties are associated with pieces of music).

Basic audiovisual crossmodal correspondences operate between discrete unimodal sensory stimuli, such as, for example, pure tones, color patches, and sounds and lights varying in terms of their intensity (i.e., loudness and brightness). It has sometimes been suggested that certain of these basic crossmodal correspondences may reflect the existence of amodal sensory dimensions, such as intensity, shape, texture, brightness, etc. (e.g., [Bond and Stevens, 1969](#); [Ellermeier et al., 2021](#); [Lewkowicz and Turkewitz, 1980](#); [Marks et al., 1987](#); [Root and Ross, 1965](#)). However, that possibility has been rejected recently, at least in the case of sensory correspondences (see [Spence and Di Stefano, 2024b](#), for a review). Rather, at least according to Spence and Di Stefano, putatively amodal dimensions of experience, such as for sensory intensity, should be reconceptualized in terms of a crossmodal correspondence instead (e.g., between auditory intensity and visually-perceived brightness, say). Similarly, [Spence's \(2015\)](#) suggestion from a decade ago that the crossmodal correspondences

² Note that while the descriptor “structural” might appear appropriate for many mid-level correspondences, the term “structural correspondences” has already been used in several other contexts previously (see [Bernard, 1986](#); [Sebba, 1991](#); [Spence, 2011](#)). Hence, to avoid confusion, we will use the descriptor mid-level instead.

between individual sensory attributes might operate on the basis of perceived similarity has also recently been discounted, at least in the case of audition and vision (Di Stefano and Spence, 2024a).³

Mid-level correspondences typically operate between structured combinations of unisensory stimuli, such as short sequences of sounds (that may make up a melody or a spatiotemporally arranged pattern of visual stimuli). Numerous studies, often in the Gestalt tradition, have demonstrated how the grouping of sequences of stimuli presented in one sensory modality may influence the presentation of stimuli presented in the other modality, as for example, in the context of the so-called “crossmodal dynamic capture effect” (Soto-Faraco et al., 2003; see Spence et al., 2007, for a review). Other research, meanwhile, has demonstrated the existence of crossmodal effects in a range of other experimental situations (O’Leary and Rhodes, 1984; see also Huang et al., 2012; Rahne et al., 2008; Schwartz et al., 2012). So, for example, crossmodal effects of auditory perceptual grouping on visual perceptual grouping, and *vice versa*, have been reported. At the same time, however, a number of studies have also demonstrated mid-level crossmodal correspondences that are seemingly more metaphorical or analogical in nature (e.g., Ravnani and Sonnweber, 2017; Wagner et al., 1981; though see also Jewanski, 2010; Lindholm, 2022). A separate literature has demonstrated how temporal synchrony, and the presentation of correlated inputs to different senses, can also facilitate crossmodal binding (e.g., Lin et al., 2022; Parise et al., 2013, 2012).

In parallel, various researchers have explored audiovisual correspondences using more complex and semantically-rich stimuli. In such cases, the complexity of the stimuli involved means that it is harder to explain any crossmodal correspondences that are observed based on specific individual auditory/visual physical stimulus attributes/dimensions (e.g., such as frequency and hue; see Duthie, 2013; Duthie and Duthie, 2015), thus leading researchers to suggest the existence of features that may be capable of mediating those associations at more of a cognitive level. So, for instance, Albertazzi et al. (2015) demonstrated the existence of consistent audiovisual associations between highly-complex stimuli consisting of a series of 15 paintings by an Italian artist and 15 music excerpts from the classic guitar repertoire (e.g., Villas Lobos, Albeniz). These associations were explained using the semantic differential technique based on perceptual and emotional associations (e.g., bright and calm, respectively) (see also Cowles, 1935, for an early study, Spence, 2020a, for a review, and Isifyan et al., 2022, for a more recent study documenting crossmodal associations between paintings and sounds). According to the “emotional mediation hypothesis”, those stimuli that are presented in different sensory domains are more likely to be matched if they share similar affective meanings (see Spence, 2020a, for a review). This has emerged as one of the most powerful accounts of this class of complex crossmodal

correspondences, and is especially powerful in helping to explain audiovisual associations (e.g., Di Stefano et al., 2024, for cross-cultural evidence), though the approach has also been applied to help explain crossmodal associations involving olfactory stimuli (e.g., Di Stefano et al., 2022; Schifferstein and Tanudjaja, 2004; Spence, 2020b) and even taste stimuli as well (Spence and Levitan, 2021).

At the same time, however, a separate distinction can be made between those studies that have merely sought to establish the existence of crossmodal correspondences between the stimuli presented at a particular level (i.e., basic, mid-level, or complex)—crossmodal matching studies—and those studies that have looked to determine whether there are any crossmodal effects, evidenced by the presence of stimuli in one modality having an impact on those presented in the other modality (see Table 1). The general belief amongst researchers would appear to be that there should be an increased probability of crossmodal binding or multisensory integration (occasionally resulting in the perceived unification of the component sensory stimuli; Eisenstein, 1947; Wells, 1980) when stimuli are crossmodally congruent as compared to when presented with crossmodally incongruent pairings instead (Iwamiya, 2013).

While crossmodal correspondences operate between sensory stimuli presented at different levels of complexity (e.g., between individual auditory stimuli and complex visual stimuli), the majority of research that has been published to date has tended to present stimuli at the same level of complexity in both the auditory and visual modalities. Nevertheless, a few studies have been published documenting crossmodal correspondences between stimuli at different levels, such as, for example, demonstrating crossmodal correspondences between short snippets of music and isolated color patches (e.g., Palmer et al., 2013; Whiteford et al., 2018).

There are two possibly controversial claims (or observations) that we wish to make in this review: The first is that the mere existence of a crossmodal correspondence, as revealed by research on crossmodal matching, has no necessary implications for the existence of a crossmodal effect; The second is that the existence of crossmodal correspondences operating between individual sensory stimuli has no necessary implications as far as the existence of crossmodal effects operating between complex stimuli are concerned. That is, many crossmodal effects have been demonstrated in the literature without any consideration of the crossmodal congruency, or otherwise, of the component unimodal stimuli (see London, 1954; Welch and Warren, 1980, 1986, for reviews). So, for example, the presentation of a tone has been reported to enhance the rated brightness of a simultaneously-presented light (Stein et al., 1996; though see Odgaard et al., 2003), or vice versa (Odgaard et al., 2004). In this case, there was absolutely no consideration of whether the frequency of the sound was in any sense congruent (or incongruent) with the color of the light. Meanwhile, many studies have demonstrated how crossmodal effects (albeit different ones) occur in the case of both congruent and incongruent combinations of sensory stimuli (Chen et al., 2025). So, for example, Chen and colleagues reported that auditory stimuli influenced people’s aesthetic perception of pictures regardless of the semantic or spatial congruency of the component stimuli.

³ Here supporting the early position adopted by Helmholtz (1878/1971, p. 77), and *contra* Marks (2011; cf. Hartshorne, 1934). In brief, Helmholtz claimed that sensory modalities are fundamentally distinct, preventing direct transitions or comparisons between them. For instance, one cannot determine whether sweetness is more similar to red or blue. Comparisons are meaningful only within the same modality, such as identifying yellow as closer to orange than to blue.

TABLE 1 Broad categorization of different types of crossmodal correspondence.

Type of correspondence	Stimuli involved	Crossmodal matching (CM)	Crossmodal effect (CE) or multisensory interaction
Basic (or low-level) correspondence (more sensory than emotional)	Pure tone, loudness, pitch; colour patch, hue, brightness.	Multiple crossmodal correspondences documented, though disagreement about the pitch-hue mapping.	Coupling priors, unity effect (Parise and Spence, 2009)
Mid-level correspondence	Rhythm, musical mode; colour harmony, spatiotemporal patterns	Metaphoric matching, structural similarity; numerosity correspondence, analogical mapping.	Gestalt grouping (O’Leary and Rhodes, 1984); Modulation of crossmodal dynamic capture effect (Spence, 2015; Spence et al., 2007)
Complex (or high- level) more emotional than sensory; Stylistic correspondences (Hasenfus et al., 1983; Siefkes and Arielli, 2015)	Painting, piece of music, film clips, TV adverts; ballet/dance (Opera excluded as it is intrinsically audiovisual)	Music matched to painting; Painting chosen to match music. Visuals matched to music; movements/body shapes or gesture and music? Light show during concert?	Does music the change viewer’s interpretation of painting or film? What is the impact of presenting incongruent music? Occasional reports of unification of inputs (e.g., in Scriabin’s Prometheus, Wells, 1980; Video jockeys matching visuals to music; e.g., Jean-Michel Jarré)

It sometimes appears to be implicitly assumed by researchers working in the area (i.e., in crossmodal correspondences research) that the existence of a crossmodal effect can be taken as evidence of the existence of a crossmodal correspondence between one or several of the sensory features or dimensions involved. As such, the view that we wish to advocate here is that the existence of a crossmodal correspondence can only be convincingly demonstrated by the results of crossmodal matching studies (that is, in studies where the participants are explicitly asked to pick the best match of the available options that have been provided to them⁴ or, in more refined/informative protocols, to rate the extent to which all of the presented stimuli from different sensory domains match one another). While it would seem likely that crossmodal correspondences, as demonstrated in so many crossmodal matching studies, will likely give rise to measurable differences when a participant’s responses to congruent vs. incongruent combinations of crossmodal stimuli are compared (Parise and Spence, 2009, 2012), it is important to remember that many (perhaps even the majority of) crossmodal effects may actually have little to do with crossmodal correspondences.

1.2 Explaining crossmodal correspondences

Over the years, a number of different mechanisms have been put forward to try to explain crossmodal correspondences (Motoki et al., 2023; Spence, 2011; Spence and Di Stefano, 2024a). These include statistical, structural (perhaps better referred to as physiological correspondences; see Spence and Di Stefano, 2024a), semantic (though again possibly better referred to as “lexical”, see Walker, 2012) as well as the emotional account (Spence, 2020a) (see Table 2). It is important to recognize that the most appropriate explanation for a given crossmodal correspondence is to some extent independent of the complexity

4 It is worth noting that, in similar protocols, the existence of a statistically significant crossmodal correspondence between stimuli in two arbitrary perceptual dimensions (such as pitch and hue; Walker-Andrews, 1994) only demonstrates that the chosen pairing was the best of the options that happened to be available to participants at the time they were asked (see Spence and Levitan, 2021).

TABLE 2 Summary of the various different types of crossmodal correspondence that have been proposed that have to connect auditory and visual stimuli and selected literature sources suggested and/or supported them.

Type of correspondence	Description/explanation
Structural	Correspondence based on structural alignment of stimulus dimensions (e.g., both pitch and hue being circular dimensions)
Physical	Correspondence based on fact that stimuli themselves are of a similar type (e.g., both sound and light being considered as waves)
Physiological*	Correspondence based on similar neural hypothetical similar neural encoding principle (e.g., increases in intensity being represented by increased neural firing)
Psychological -perceptual (-Amodal)	Correspondence based on perceptual similarity between the component stimuli (sometimes referred to as intersensory equivalence)
Mood-based/affective/emotional-mediation	Correspondence based on mood/emotion associated with each of the component stimuli being similar
Associative/statistical	Correspondence based on associative learning/internalization of crossmodal statistical regularities in environment

*Note that Spence (2011) originally labeled this category ‘Structural’. However, given that this label had already been introduced 20 years earlier by Sebba (1991) to describe dimensional alignment of perceptual space, it would seem more appropriate to call this category “Physiological”, given that that is the putative cause/source of the correspondence.

of the component unisensory stimuli that are involved. That said, claims regarding the existence of amodal sensory and similarity-based crossmodal correspondences have typically been made in relation to individual sensory stimuli. Meanwhile, Gestalt grouping, as well as spatiotemporal synchrony, is more often mentioned as the most probable explanation for any crossmodal effects operating at the level of mid-level correspondences (Spence, 2015). Emotional mediation tends to be the most frequently mentioned explanation of audiovisual correspondences between complex and aesthetically-meaningful stimuli, such as, for example, music and painting/images (e.g., Di Stefano et al., 2024). That said, emotional mediation is occasionally raised as a possible explanation in the case of crossmodal correspondences between individual perceptual stimuli, while basic sensory correspondences (e.g., between pitch and size) could potentially also contribute to what might, at first,

be taken as stylistic correspondences (see [Duthie, 2013](#); [Duthie and Duthie, 2015](#); [Siefkes and Arielli, 2015](#)).

2 Crossmodal correspondences between simple sensory stimuli

One of the most extensively studied areas in research on audiovisual correspondences concerns the associations that have been demonstrated between individual sensory stimuli (or attributes), such as pitch, brightness, or loudness. Over the last 50 years or so, a very wide array of crossmodal matching studies has been published. More recently, a number of studies demonstrating crossmodal effects on perceptual binding as well as in a variety of speeded response tasks that are modulated by the crossmodal correspondences that exist between individual sensory stimuli (attributes or sensory dimensions) have also appeared in the literature; it is to this research that we turn next.

2.1 Crossmodal matching

Many published studies have documented crossmodal correspondences between simple stimuli, or stimulus dimensions (see [Spence, 2011](#); [Spence and Sathian, 2020](#), for reviews of the literature on basic audiovisual correspondences). In fact, the literature reveals a multitude of crossmodal correspondences between, for instance, auditory pitch and a variety of visually-presented dimensions: These include size (e.g., [Anikin and Johansson, 2019](#); [Bonetti and Costa, 2018](#); [Evans and Treisman, 2010](#); [Gallace and Spence, 2006](#); [Mondloch and Maurer, 2004](#)), shape/angularity ([Marks, 1987](#); [Murari et al., 2020](#); [Parise and Spence, 2012](#)), lightness, brightness (e.g., [Brunel et al., 2015](#); [Hubbard, 1996](#); [Klapetek et al., 2012](#); [Marks, 1974, 1987, 1989](#)), hue (e.g., [Melara, 1989](#); [Spence and Di Stefano, 2022a](#), for a review) and elevation ([Bonetti and Costa, 2019](#); [Wrembel, 2009](#); [Chiou and Rich, 2012](#)), etc. (see also [Sun et al., 2018](#)). The existence of such correspondences has often been explained in terms of the statistical or structural/physiological account ([Spence, 2011](#)).

There has, however, been widespread disagreement in the case of a putative crossmodal correspondence between auditory pitch and visual hue. Some commentators have been convinced that such a correspondence must exist based on the structural similarity of the underlying sensory dimensions ([Gombrich, 1960](#); [Sebba, 1991](#)) while others deny the existence of such a correspondence, or else disagree concerning what the most appropriate mapping would be ([Bernstein et al., 1971](#); [Caivano, 1994](#); [Davis, 1979](#); [Garner, 1978](#); [Newton, 1704](#)). At the same time, however, other researchers have argued that methodological issues (i.e., confounding hue and brightness) may have compromised the interpretation of several earlier studies ([O'Malley, 1957](#); [Simpson et al., 1956](#); [Wicker, 1968](#)). Despite the structural features that might be shared by the visual dimension of hue and the auditory dimension of pitch (e.g., categorical perception in the continuous nature of wavelengths/frequency range scales), as highlighted in the reviews from [Spence and Di Stefano \(2022a, 2024a\)](#), there would appear to be little agreement as far as which hue should, in fact, be matched with which musical note. What is more,

structural issues soon emerge when it comes to explaining color-sound correspondences based on any perceptual, psychophysical, or physical similarity between the two component stimuli. For instance, when considering the color spectrum and audible sound range, the former fundamentally differ in terms of the non-linear distribution of tones along the audible frequency spectrum as compared to colors in the visible range. Moreover, a fundamental property of auditory perception, namely octave periodicity, seems to have no parallel in vision, apart from some metaphorical and indirect usages of the same terms for referring to colors by some commentators (e.g., [Pridmore, 1992](#)).

More recent research has drawn attention to the existence of a number of robust crossmodal correspondences between visual features (i.e., surface textures) and auditory timbre, due to the allegedly intrinsic multisensoriality of timbre characterization/semantics (e.g., [Reuter et al., 2018](#); [Reymore and Lindsey, 2025](#); [Saitis et al., 2020](#); [Wallmark, 2019](#)).⁵ For instance, [Reymore et al. \(2023\)](#) demonstrated that timbre semantic associations mostly refer to the visual appearance of objects, such as the material or haptic features of surface textures (e.g., raspy, smooth, brilliant). Several other recent studies have, though, deepened our understanding of the relationship between auditory roughness and shapes, by demonstrating that listeners tend to match more dissonant sounds to spikier and rougher objects/shapes (see also [Murari et al., 2015](#)). [Giannos et al. \(2021\)](#) assessed whether non-tonal and highly dissonant harmonic stimuli are associated with rough images, while more consonant stimuli are associated with the images of low visual roughness. To test this particular hypothesis, the researchers harmonized a fixed melody in seven different styles, including highly tonal, non-tonal, and random variations. The participants in this study were asked to match the melodies to images of variable roughness (i.e., black and white 2D and 3D images that represented surfaces with different degrees smoothness/roughness). In this case, the results confirmed that auditory dissonance was indeed highly correlated with visual roughness (see also [Di Stefano and Spence, 2022](#), for a review).

Other researchers have demonstrated the existence of audiovisual crossmodal correspondences between musical timbre and visual shapes ([Adeli et al., 2014](#); [Gurman et al., 2021](#)). For instance, the participants in a study by [Adeli et al. \(2014\)](#), were asked to select the visual equivalent of a given sound, i.e., its shape, color (or grayscale) and vertical position. A strong association between timbre and visual shapes was observed, with soft timbres being more frequently associated with blue, green or light gray rounded shapes, while harsh timbres were associated with red, yellow or dark gray sharp angular shapes. Importantly, these findings were subsequently replicated by [Gurman et al. \(2021\)](#).

2.2 Crossmodal effects

A wide range of crossmodal correspondence effects have been demonstrated across a wide range of experimental paradigms (e.g., [Evans and Treisman, 2010](#); [Gallace and Spence, 2006](#);

⁵ And see [Menouti et al. \(2015\)](#) for the synesthetic angle to this phenomenon.

Parise and Spence, 2012; and see Spence, 2011; Spence and Sathian, 2020, for reviews). So, for example, Parise and Spence (2009) published a series of psychophysical studies demonstrating increased spatiotemporal integration for crossmodally congruent combinations of auditory and visual stimuli (e.g., relative high-pitched sound with a relatively large visual stimulus) as compared to incongruent combinations (e.g., relatively high pitch and small visual size in an unspeeded psychophysical task).⁶ Sound frequency has also been reported to influence people's responses to color lightness (Hagtvedt and Brasel, 2016).

Other researchers have started to address the question of whether timbre modulates visual perception. So, for example, in a series of two experiments, Wallmark et al. (2021) investigated whether the timbre of a musical note (an acoustic prime) would affect the subsequent perception of, in the first experiment, brightness (dark–bright dimension) and, in a second study, both brightness and spatial texture (smooth–rough dimension). The participants had to identify a shift in roughness/brightness between two consecutively-presented target squares of subtly contrasting levels (rougher/brighter, smoother/darker, or the same) in a speeded-response paradigm. In a second experiment, prior to the presentation of the target square, the participants could be exposed to sounds that varied in terms of their roughness (smooth/rough). For visual stimuli, photos of sandpaper patches of contrasting grit sizes were used: baseline, medium roughness (100-grit); low roughness (150 grit); and high roughness (50 grit). A sine wave and a sawtooth wave they used as auditory stimuli. Modest evidence was found that timbres increase response bias in a semantically congruent manner when participants identify visual stimuli (e.g., when a “rough” saw-tooth wave accompanies the second of two identical spatial textures, the “rough” sound increased the probability of participants judging the second texture as rougher), thus suggesting that rough sounds may increase of the perceived roughness of the visual stimuli (see also Wallmark and Allen, 2020). Other researchers, meanwhile, have chosen to investigate timbral brightness by means of multisensory interference phenomena (Saitis and Wallmark, 2024).

Hence, in summary, there is currently robust experimental evidence for the existence of a wide range of audiovisual crossmodal correspondences involving pitch, such as pitch-size, pitch-brightness, pitch-height, with the exception of pitch-hue correspondences. In the latter case, despite the fact that many commentators are convinced of what the most appropriate crossmodal mapping is (see Spence and Di Stefano, 2024a, for a review), there would appear to be little agreement except in odd cases (such as the color scarlet supposedly matching the sound of the trumpet; e.g., Kandinsky, 1977 (originally published in 1911); Leibniz, 1896 (originally published 1704); Locke, 1690;⁷

Ortmann, 1933; see also Donnell-Kotrozo, 1978). However, the fact that pitch-based crossmodal correspondences are relative in nature for normal listeners (Spence, 2019),⁸ means the consequences for those wanting to art crossmodally are unclear, though the question remains open for those with absolute pitch (see Di Stefano and Spence, 2024b).⁹ What is more, the sheer range of crossmodal correspondences that might be primed by a given auditory pitch means that it is difficult to see how a unique mapping could be established in (e.g.,) a sensory-substitution setting (Cavazos Quero et al., 2021), or else when considering the question of sensory translation between one sense and another (Spence and Di Stefano, 2024b). This challenge presumably also applies for those wanting to translate works of art from one sensory modality into another (e.g., see Müller-Eberstein and van Noord, 2019).

3 Structural correspondences

Mid-level audiovisual crossmodal correspondences operate on dynamically-changing stimuli and/or on combinations of unisensory stimuli (attributes or dimensions; Guo et al., 2023). They sit between the correspondences that have been observed with individual stimuli and those that operate at the level of more aesthetically-meaningful complex sets of stimuli.¹⁰ Mid-level crossmodal matching is sometimes based on the common numerosity, or spatiotemporal organization, of the individual elements that happen to be presented in each sensory modality (see Spence, 2015, for a review), while at other times it may be based on correlated unisensory input (Parise et al., 2012, 2013). One might also think of crossmodal correspondence that are based on the harmony of the stimuli presented in each modality (Alves, 2005; Spence and Di Stefano, 2022b; Wells, 1980; cf. Pimentel et al., 2025). At the same time, however, metaphorical (e.g., Ravignani and Sonnweber, 2017; Wagner et al., 1981) and emotional mediation has also been reported in at least some cases.

3.1 Crossmodal matching

To date, far fewer crossmodal matching studies have been conducted at this level. What is more, the majority of research that has been published to date would appear to involve cross-level correspondences, such as music-to-color associations for single-line piano melodies (Lindborg and Friberg, 2015; Palmer et al., 2016). Researchers have assessed crossmodal correspondences between colors and music clips (Chen, 2013), musical genres (Holm et al., 2009; see also Bresin, 2005), music tempo (Guo and Jiang, 2023; see also Timmers, 2022), melody (Cuddy, 1985), and musical

⁶ That said, Sourav et al. (2025) were unable to replicate at least one of these specific findings.

⁷ Note though that Locke does not present this example in order to endorse its solution, but to comment ironically: “For the hope to produce an idea of light or color by a sound, however formed, is to expect that sounds should be visible or colors audible, and to make the ears do the office of all the other senses. Which is all one as to say that we might taste, smell, and see by the ears: a sort of philosophy worthy only of *Sancho Pancha*, who had

the faculty to see *Dulcinea* by hearsay.” (Locke, 1690, pp. 29–30—emphasis in the original).

⁸ Note, though, that the relative nature of crossmodal correspondences might not be restricted to pitch (see Cohen, 1934).

⁹ Note though that the Pantone color of the year a few years ago was released along with a matching soundscape.

¹⁰ One challenging case here is consonant and dissonant stimuli (Di Stefano et al., 2022). Some consider dissonant stimuli unitary, others believe that people are aware of the component stimuli.

intervals Di Stefano et al. (submitted)¹¹. Gaiger (2018) has even suggested that paintings can have rhythm. It is also easy to see how the emotional associations listeners have with major vs. minor chords could also be linked to colors (Parncutt, 2014, 2024; see also Carraturo et al., 2025; Wanke et al., 2025).

3.2 Crossmodal effects

When, for instance, the rate of repetitive auditory stimuli is increased or decreased, while a simultaneously-presented visual flicker remains constant, the latter appears to change accordingly with the auditory stimulus, an effect known as “auditory driving” (Gebhard and Mowbray, 1959; Recanzone, 2003, 2009; Shipley, 1964; Welch et al., 1986; see also Boltz, 2017, 2018; Occelli et al., 2011; Shams et al., 2000). By contrast, changes in the rate of visual flicker do not appear to change the perceived rate of auditory flutter to anything like the same extent. This, then, is an example of auditory dominance (see Di Stefano and Spence, 2025, for a number of other crossmodal influences in the perception of temporal structure).

O’Leary and Rhodes (1984) conducted a study demonstrating a bidirectional influence between vision and audition in a task involving perceptual organization. Multiple studies have demonstrated how intramodal perceptual grouping precedes crossmodal influences in the context of the crossmodal dynamic capture effect (Soto-Faraco et al., 2003; Spence et al., 2007; see also Cook and Van Valkenburg, 2009; Gilbert, 1938; Maeda et al., 2004; Takeshima and Gyoba, 2013; Watanabe and Shimojo, 2001).¹² Crossmodal congruency effects have also been documented based on stimulus identity, in this case reflecting a short temporal pattern (Frings and Spence, 2010; cf. Huddleston et al., 2008; Julesz and Hirsh, 1972).

Consider here only how the affective meaning of musical scales depends on their direction—namely, whether they are ascending or descending (Collier and Hubbard, 2001). On the one hand, this association hints at the importance of the structural organization of the individual stimuli presented in a given modality, over-and-above whatever sensory qualities any of the individual elements may have when presented in isolation. This observation may also be linked to the relative nature of crossmodal correspondences, especially those involving auditory pitch (Spence, 2019; see also Eysenck, 1940). More broadly, the way in which individual elements are structured within a given sensory modality—and how they relate across the senses—brings us to the concept of harmony,

a principle that has long been central to both music and visual aesthetics. In the Greek tradition, the term harmony originally related to the physical unification of different elements (Lippman, 1963). As such, the concept of harmony found a natural application in the domain of music, conceived of as an artistic practice based on the juxtaposition of different auditory elements (i.e., sounds) fitted together to generate pleasant effects. In the field of vision science, a number of researchers have investigated the concept of harmony between color pairs (e.g., Allen and Guilford, 1936; Burchett, 2002; Field, 1835; Palmer and Griscom, 2013; Schloss and Palmer, 2011). McNeill Whistler (1978) talked of paintings as harmonies (though see Kargon, 2011). Some suggest that music and painting may be combined harmoniously (e.g., consider only the literature on “color music”; Argüelles, 1972; Plummer, 1915; Rimington, 1915; Sullivan, 1914; Watkins, 2018; Zilczer, 1987). Given such considerations, it becomes natural to consider whether the concept of crossmodal harmony makes any sense (Kargon, 2011; see Spence and Di Stefano, 2022b, for a review).

4 Crossmodal correspondences between complex auditory and visual stimuli

Experimental psychologists have long been interested in the crossmodal matching of complex auditory and visual stimuli, as well as on the crossmodal effects of one on the other (see Spence and Di Stefano, 2025, for a review). For whatever reason, the majority of studies that have been published to date in this area would appear to have been conducted with images of paintings and short classical music selections (see Spence, 2020a, for a review).

4.1 Crossmodal matching

A number of studies have presented a small selection of stimuli in one sensory modality, with participants being required to match from a pre-selected set of stimuli presented in the other modality (see Table 3). It should, though, be noted that many of the early studies would appear to have been statistically underpowered by contemporary standards. Moreover, there have been relatively few attempts to replicate previous studies, as has become an increasingly common practice given the replication crisis in psychological science. That said, the majority of the research that has been published to date appears to demonstrate a consensual (i.e., a significantly non-random) tendency (though, of course, one should always be mindful of the so-called “file drawer problem”; see Rosenthal, 1979).

The majority of research that has been published to date would appear to be consistent with the claim that complex crossmodal associations between music and paintings are primarily mediated by emotion (see also Hung, 2000).¹³ More refined protocols typically present the auditory stimulus—specifically, a musical

11 Di Stefano, N., Ansani, A., Schiavio, A., Saarikallio, S., Toiviainen, P., Brattico, E., et al. (submitted). “Minor seconds are bitter and blueish while major sixths are sweet and yellowish: a cross-cultural study of multisensory mappings of musical intervals,” in *Cognition*.

12 Of course, it is entirely possible that any Gestalt perceptual grouping might only occur unimodally in the auditory and/or visual modalities, and crossmodal art might emerge from the parallel unisensory Gestalt perceptual grouping observed in each of the component modalities (cf. Kubovy and Yu, 2012)—one might consider this a kind of crossmodal correspondence between unisensory Gestalt attributes (see also Bender et al., 1954; Spence et al., 2007).

13 That said, spatial conceptual associations between music and pictures have been revealed by researchers looking at the N400 effect (Zhou et al., 2014).

TABLE 3 Broad categorization of different types of audiovisual crossmodal matching and crossmodal effects involving complex multi-element aesthetically-meaningful sensory stimuli (see [Spence and Di Stefano, 2025](#)).

Type of correspondence	Stimuli involved	Crossmodal matching (CM)	Crossmodal effect (CE)
Complex (or high-level)	Paintings and pieces of music	Walker, 1927 ; Cowles, 1935 ; Wehner, 1966 ; Peretti, 1972 ; Minnigerode et al., 1976 ; Parrott, 1982 ; Albertazzi et al., 2015 ; Rančić and Marković, 2019 ; Iosifyan et al., 2022	Lindner and Hynan, 1987 ; Limbert and Polzella, 1998 ; Actis-Grosso et al., 2017 ; Rančić and Marković, 2019 ; Braun Janzen et al., 2023 ; Fekete et al., 2023 ; Fink et al., 2024

excerpt—while having the participants rate the extent to which it matched each of the pre-selected visual stimuli. This kind of experimental design allows researchers to better understand association trends (e.g., [Di Stefano et al., 2024, 2025](#)) while, at the same time, avoiding the risk of forcing a “best” choice response if none of the presented stimuli happen to clearly match the auditory excerpt. There have also been a number of studies looking at cross-level correspondences involving complex stimuli as but one component. For instance, [Isbilen and Krumhansl \(2016\)](#) conducted research highlighting the existence of emotion-mediated color associations to Bach’s Well-Tempered Clavier. Furthermore, and similar to what has just been reported, the suggestion is that crossmodal correspondences between music and ambient color are mediated by emotion ([Hauck et al., 2022](#); though see [Lipscomb and Kim, 2004](#)).

4.2 Crossmodal effects

Researchers have demonstrated that music (or musical emotion) can crossmodally influence the perceived (or at least rated) brightness of paintings ([Bhattacharya and Lindsen, 2016](#); see also [Logeswaran and Bhattacharya, 2009](#)). For example, the participants in one study reported by [Hong et al. \(2024\)](#) were presented with a wide range of emotional music pieces alongside various visual stimuli. The results revealed that lower-pitched music tended to result in darker judgments of visual objects than when higher-pitched music was presented instead (again see [Table 3](#)).¹⁴ Meanwhile, a study by [Albertazzi et al. \(2020\)](#) investigated the existence of crossmodal correspondences between contemporary art (paintings by Kandinsky) and music (excerpts from Schönberg). The experiment was conducted in two phases. In the first phase, the participants evaluated the perceptual characteristics first of visual stimuli (some pictures of Kandinsky’s paintings, with varying perceptual characteristics and contents) and then of auditory stimuli (musical excerpts taken from the repertoire of Schönberg’s piano works) relative to 11 pairs of adjectives tested on a continuous bipolar scale (i.e., based on Osgood’s semantic differential technique; [Osgood et al., 1957](#)). In the second phase, participants were required to associate pictures and musical excerpts. The results of the semantic differential demonstrated that certain paintings and musical excerpts were evaluated as semantically more similar, while others were evaluated as semantically more different. At the same time, the results of the

direct association between musical excerpts and paintings showed both attractions and repulsions among the stimuli. The overall results provide significant insights into the relationship between concrete and abstract concepts and into the process of perceptual grouping in cross-modal phenomena (see also [Ansani et al., 2020](#); [Boswell, 2012](#); [Ellis and Simons, 2005](#); [Iwamiya, 1994](#)).

4.3 Interim summary

Given the research that has been reviewed so far, it would appear clear that crossmodal matching occurs between all possible combinations of individual, organized, and more complex, semantically-meaningful stimuli. Having demonstrated such crossmodal correspondences the next question is how such findings inform attempts at sensory translation or sensory augmentation. At the outset here, one might consider whether crossmodal matching is perhaps more relevant when considering sensory translation, whereas the literature on crossmodal effects may be more pertinent for those wanting to augment art crossmodally. Returning to the two important claims that we put forward earlier, we might say that the mere existence of a crossmodal correspondence (as revealed by research on crossmodal matching), has no necessary implications for the existence of a crossmodal effect. A second claim is that the existence of crossmodal correspondences operating between individual sensory stimuli has no necessary implications as far as the existence of crossmodal effects operating between complex stimuli are concerned. Furthermore, it can also be argued that the absence of a demonstrable crossmodal effect cannot be taken as evidence of the absence of a crossmodal correspondence either. That is, it is at least theoretically plausible, that people might consensually rate certain combinations of stimuli as matching vs. mismatching, without there being any crossmodal effects on performance (e.g., in a speeded classification task). All three of these claims would appear consistent with the various sources of evidence that have been reviewed here.

5 Attempts to augment auditory/visual art and entertainment crossmodally

There is an extensive literature on “color music”, sometimes called “visual music” (e.g., [Adams, 1995](#); [Baker, 2002](#); [Dimova, 2024](#); [Galeyev, 2003](#); [Klein, 1937](#); [Liu, 2022](#); [Lupton, 2018](#); [Peacock, 1988](#); [Poast, 2000](#); [Sabaneev and Pring, 1929](#); [Sabanev, 1911](#); [Scholes, 1970](#); [Shaw Miller, 2006](#); [South, 2001](#); [Vergo, 2012](#); [Zika,](#)

¹⁴ It might be worth considering whether the concept of halo dumping ([Clark and Lawless, 1994](#)) could be relevant in those studies where the participants are only given a single dimension on which to respond.

2013, 2018; Zilcher, 1987, 2016). These phrases refer to artistic interest in both translating music into color but also in augmenting one art form by additional stimulation in the other modality. When translation is attempted in the opposite direction it is sometimes referred to as “graphical sound” instead (Benčić, 2021). While it has been argued at length elsewhere that literal sensory translation may not be possible (Spence and Di Stefano, 2024a; see also Zika, 2013, 2018), there is an interesting body of artistic work that probes the possibility of augmenting art crossmodally (see Spence and Di Stefano, 2025).

Many of those interested in this area invoke the notion of synaesthesia when thinking about engaging multiple senses simultaneously (e.g., see Dimova, 2024; Karwowski et al., 1942; Spence and Di Stefano, 2025; Waterworth, 1997). However, as Deroy and Spence (2013) have previously argued at length, such an appeal (to synaesthesia) is fundamentally misguided, and that what may be much more important is to build on the crossmodal correspondences that are consensually shared by the majority of people if one wants to augment art crossmodally in a way that communicates meaningfully with the majority (cf. Spence, 2012). Polina Dimova uses the term “synaesthetic metaphor” as a heuristic device (Dimova, 2024), grouping congenital synaesthesia, literary synaesthesia (aka pseudo-synaesthesia), intermediality in the arts, and cultural synaesthesia together. As Howes (2025, p. 147) puts it, she: “engages with them contrapuntally, in the manner of a fugue. The right-thinking neuroscientist would be aghast at such heterodoxy and abominate it”. The reason being the Criterion of Consistency test that is now used as a gold standard to determine who is a “genuine synaesthete” according to neuroscience (Baron-Cohen et al., 1993; Johnson et al., 2013; Root et al., 2025; Svardal and Iversen, 1989; though see also Harrison and Baron-Cohen, 1996). Such a view, although aligning with the “neuromania” (Legrenzi and Umiltà, 2011) or “cerebrocentrism” (Howes’ terminology) of so much contemporary cognitive neuroscience research, can seem overly constraining when considering the crossmodal augmentation of art (e.g., see Howes and Classen, 2014; Young, 2005). What is needed is rather more of a social/structuralist account of synaesthesia, consistent with the views put forward by Lévi-Strauss (1997). Echoing much the same point, half a century earlier, Vanechikina (1973) noted how almost all of those he spoke to specially emphasized the associative, metaphoric nature of “color hearing” in music and excluded from the artistic sphere any clinical cases of color hearing.

Separate from any attempt to use synaesthesia in crossmodal compositions, it should be recognized that certain art forms, such as live opera, inherently stimulate multiple senses; Live music performance too, though this varies all the way from simple inevitability of seeing musicians play in a classic music concert through to the visuals that augment more experimental music (see also McDonald et al., 2022). In the latter cases, though, it is by no means always clear which is the dominant/primary modality. However, separate from these issues, there are genuinely unisensory experiences of art, painting, sculpture, at least as it is displayed in contemporary galleries.¹⁵ At the other extreme, there may be accidental/incidental background music presented in a gallery

environment (see Table 4). In such cases, there is unlikely to be any intentional connection, or expectation of relevance or resonance (Loureiro et al., 2019).

Iwamiya (2013) investigated the interaction between auditory and visual processing when listening to music in an audiovisual context. C. D. Friedrich, for example, exhibited pictures in low light accompanied by music (Siegel, 1974). P. O. Runge exhibited the *Times of the Day* painting series in J. W. von Goethe music room and by doing so, the suggestion is that the aesthetic experience was escalated into a full multisensory experience (The Getty Research Institute, 2013).

5.1 Sensory translation: turning music into light

There has long been artistic interest in the possibility of sensory translation (Kargon, 2011; see Spence and Di Stefano, 2022a, for a review). Relevant here, Stechow (1953) highlighted an important distinction between: “translations from the visual arts into music and parallelisms between the visual arts and music” (p. 324, italics in original). Later, he observed that “it would seem to me that comparability of structure reveals a more ‘real’ relationship between such works of art than a mere affinity of ‘mood’ or ‘texture’ could suggest” (Stechow, 1953, p. 325); Notice how the latter comment presumably emphasizes the structural rather than the affective nature of underlying correspondences (though see Kargon, 2011, on the contrary position). Just consider Alexander B. Rimington (1854–1918) of the Royal College of Arts in London, who first presented his color organ to the public in 1895, and which he performed with up until 1911 in London (Rimington, 1895, 1911, 1915). Father Castel, the French priest’s earlier suggestion that every home would one day have an ocular harpsicord never came to pass (see Conrad, 1999; Moritz, 1997). The project was likely doomed to failure (Hankins, 1994).¹⁶ That is, the best that you can hope to do is to convey the emotional tone of a specific work (again, see Spence and Di Stefano, 2022a, 2024a, for reviews). Furthermore, sensory translation has typically been considered in terms of the element-by-element translation of individual sensory stimuli (Gibson, 2023; Spence and Di Stefano, 2024a). However, as we have just seen, mid-level and complex correspondences emerge as a result of intramodal perceptual grouping, and hence what might be more appropriate to translate is the unimodal Gestalt, in the augmentation of art and entertainment crossmodally (Bragdon, 1916, 1918).

5.2 Sensory augmentation: Scriabin’s Poem of Fire

A somewhat distinct literature has emerged on the question of the crossmodal sensory augmentation of art. Here, one might consider Scriabin’s (1910) “Prometheus: Poem of Fire” as a

¹⁵ The shiny patina on many public brass sculptures clearly speaks to people’s desire to haptically interact with many works of art when allowed.

¹⁶ As was the *Pyrophone* from Kastner (1870) that incorporated 13 foil-covered gas nozzles, lighted by crystal tubes (Kastner, 1875).

¹⁵ Though it is perhaps worth noting here that some early sculptural pieces were explicitly designed to be felt not seen (see Gallace and Spence, 2014).

TABLE 4 Summary of various kinds of audiovisual artistic/entertainment experience, where music complements/augments the visuals, or vice versa.

Artistic/entertainment experience w. music	Relation between auditory and visual elements	Example (relevant literature)
Attending a classical music concert in person	The seen gestures of the musicians and vocalists are the cause of the sounds that are heard	Vuoskowsky et al., 2014, 2016
Dance/ballet performance, or a rhythmic gymnastics routine	Seen dancer's movements typically a response to the auditory score (e.g., The Nutcracker; or Swan Lake)	Chiat et al., 2013
Going to the opera	The background music is designed to set a mood for the on-stage action	Though see Kierkegaard, 1988, on Mozart's opera, Don Giovanni
Attending a performance of Scriabin's Poem of Fire with <i>luce</i> accompaniment	Music composed first and <i>luce</i> designed to enhance the experience	Scriabin, 1910; Gawboy and Townsend, 2012
Watching an audiovisual Film or TV advert	Music normally added post-production in order to convey a particular mood, emotion, or atmosphere	
Video-jockey (VJ); Jean-Michel Jarré concert (Boltz, 2013)	VJ mixes a visual accompaniment in response to pre-existing music	Synaesthetic visual augmentation of Mahler's 2nd symphony (Deutsch, 2012; cf. Haverkamp, 2020)
Multimedia artworks composed to engage both eye and ear	Audiovisual stimuli may be synchronized, and/or may share a common, or intentionally conflicting emotional message	Basquin, 2020; Benčić, 2021; Whitney, 1980
Experiencing architecture w. musical accompaniment	Given stylistic similarities, music sometimes used to augment videos of architecture	Muecke and Zach, 2007; Siefkes and Arielli, 2015
Visual artwork augmented by auditory (or other sensory elements)	The auditory augmentation is designed to modify the viewer's experience of the artwork; Some examples of "color music" or "visual music" fit here	Tate Sensorium (Pursey and Lomas, 2018)

Notice how the synchrony between the auditory and visual elements may sometimes exogenously encourage the integration of the audiovisual inputs (e.g., when attending a classical music concert; see also Muller, 2010), while in other cases, the context in which the two channels of communication are experienced may encourage more of an endogenous, or top-down, attempt by the viewer (or audience) to combine the multisensory inputs.

paradigmatic example (Hull, 1927). This musical score is unusual in that it came with an accompanying score for the *luce* (Brent-Smith, 1926a,b; Gawboy and Townsend, 2012).¹⁷ Although not performed in Scriabin's lifetime, the code for the light score (which had been thought lost) was discovered some decades later (Galeyev and Vanechkina, 2001). Unfortunately, however, Scriabin died before the first performance of his work with *luce* accompaniment took place (in New York City in 1915; Baker, 2002). Nevertheless contemporary performances of this multisensory artwork have sometimes incorporated both auditory and visual elements (Galeyev, 1988; see Spence and Di Stefano, 2022a, for a review). One can think of this as an attempt to augment (i.e., rather than to translate) music with a color organ. That said, Scriabin's intention for the *luce* remains shrouded in mystery. Confusing matters somewhat, Scriabin himself was a synaesthete (including color music; Triarhou, 2016; Witztum and Lerner, 2016).¹⁸ Some commentators have suggested that the introduction of the *luce* may have been intended to help disambiguate elements in the musical score (Gawboy and Townsend, 2012). However, this remains a question of ongoing theoretical debate (see also Howes, 2025).

Intriguingly, although formal evaluation has not yet been conducted, those commentators who have experienced such intentional multisensory, or crossmodal, artistic presentations have, on occasion, hinted at an emergent experience, one that supports the role of Gestalt perceptual grouping in a crossmodal

artistic context (though note that the rich notion of multisensory emergence likely reaches beyond the insights offered by the Gestalt approach). Just take the following quote from Wells (1980, p. 101–105): “One has only to trace the remarkable unfolding of harmonic color in Scriabin's later works to realize that it was harmony above all that prompted him to seek a visual counterpart to his music. An equivalence of aural harmony and visual is so strongly implied in Scriabin's later works that for me they seem to emanate from a single source... The author tells of events leading up to finding a possible link between musical harmony and visual color based on complementarity... The parallelism of musical harmony with visual color departs radically from current hypotheses of tonal organization.” (cf. Spence and Di Stefano (submitted)¹⁹, on the notion of crossmodal counterpoint). Note here, though, that Wells is not a disinterested observer inasmuch as in his article he goes on to propose a crossmodal pitch-hue mapping of his own.

In principle, multisensory emergence might be just one example of an extraordinary multisensory experience (e.g., Critchley, 1977; Ehrenzweig, 1965). Sirius and Clarke (1994, p. 119) similarly write of how: “There is a long history of the association of music with visual spectacles of one sort or another (opera, theater, and in this century film, television and video), and a substantial amount of anecdotal evidence that the two media can interact in powerful and effective ways.” (see also Anderson, 1993). The unification of the unimodal elements (as hinted at by Wells, 1980) is another possible outcome that is sometimes mentioned in the literature albeit infrequently, as is the notion of resonance (perhaps a kind of crossmodal modulation) is also mentioned in more artistic contexts (e.g., Muecke and Zach, 2007; cf. McQuarrie and

17 The piece was initially conceptualized as a total work of art combining music, a light score or “luce”, and even fragrance (Dimova, 2024; Peacock, 1985; Plummer, 1915; Scriabin, 1910). Although what fragrance Scriabin intended to use was never specified by the composer.

18 In fact, he was tested by C. S. Myers at the University of Cambridge, and the results written-up in a couple of academic papers (Myers, 1911, 1914).

19 Spence, C., and Di Stefano, N. (submitted). Crossmodal counterpoint: From music to multimedia – Incongruity, cognitive dissonance, irony, and surrealism. *i-Perception*.

Mick, 1992). Researchers have demonstrated that visual cues can modulate the integration and segregation of objects in auditory scene analysis (Rahne et al., 2007). Audiovisual Gestalten likely rely as much on temporal synchrony as much as any crossmodal correspondence at the level of individual sensory elements, or even the Gestalt impression (see also Alves, 2005; Battey and Fischman, 2016; HaCohen, 2016; Haverkamp, 2020; Kaduri, 2016; Muller, 2010; Russet and Starr, 1988; Whitney, 1980). At the same time, however, it has also become increasingly clear that there is no straightforward structural or psychological mapping between audition and vision.

While the term multimedia is broad, and covers multiple forms, there is a sense that some multimedia art works fit the description of crossmodal art in that the stimuli presented to eye and ear are in some sense independent, yet the case is different from the addition of music to film, where the music is seemingly always added as an afterthought (Lipscomb, 2013; Tan et al., 2013; Truckenbrod, 1992). At the same time, however, there are empirical phenomena that are undoubtedly worthy of some form of explanation. Consider here only the phenomenal success of the music of Jean-Michel Jarré, with the largest live audience for an entertainment event ever recorded at the time (in excess of 750,000 people; cf. Boltz, 2013; Tan et al., 2013) when a light show would accompany his musical performances. It would seem hard to explain why such multimedia performances were (at least for a time) so successful, without having recourse to some kind of emergent experience that elevated the multisensory experience beyond what would have been experienced at a traditional music concert (or color organ show; see Spence and Di Stefano, 2022a).²⁰ At the time, these were the largest live entertainment audiences ever recorded, and thus speak to the power of simultaneous audiovisual stimulation.²¹ In this case, the artist was augmenting his music, which came first, with visuals (see also Collopy, 2000, and Chmiel and Schubert, 2019, for a history).

5.3 Intersensoriality

Howes (2025) has recently introduced the notion of intersensoriality to describe art forms that deliberately cross sensory borders (Classen, 1998). The notion of intersensory, somewhat distinct from both multisensory integration and synaesthetic translation, emphasizes relationality, complementarity, and emergent meaning. Here, the senses are not unified in an essentialist manner, nor merely co-activated, but brought into dialogue in a way that preserves and indeed accentuates their difference. Howes (2025, p. 143) writes of how: “the new intersensory music and art history” departs from the conventional distribution and hierarchization of the sensible by focusing on the exchanges between—and intercalibration of—the senses. This redirection of attention is manifest in such works

as Classen’s (1998) chapter on “Crossing sensory borders in the arts” in *The Color of Angels*, Shaw-Miller’s (2010) *Eye hEar: The Visual in Music*, and Veldhorst’s (2018) *Van Gogh and Music: A Symphony in Blue and Yellow*, among other works (see especially Lauwrens, 2012). Howes (2025, p. 147) concludes his piece as follows: “here was an exhibition entitled ‘The Romantic Eye’ going on at the Swedish Nationalmuseum at the time of the Stockholm conference. It was a little too ocularcentric for my taste, but one writing on a wall caught my eye: ‘An artist is each and every one for whom the aim of life is to develop one’s senses’—Friedrich Schlegel”. Fair enough, but for the avant-garde it’s different: for them, making music or making art involves straining at the bounds of sense, or crossing over (Classen, 1998). Contemporary historians of music and art need to follow suit if they are to make any headway in the understanding of the art-music nexus at the conjuncture of the nineteenth and twentieth centuries as at the conjuncture of the twentieth and twenty-first centuries (see Jones et al., 2006; Lauwrens, 2012). As we have seen, it is the renvoi, the interplay, the interrelationships of the senses that define the most lasting aesthetic expressions of the two periods”.

The notion of intersensoriality can thus help to promote a relational view of the senses, one that resonates with Lévi-Strauss’ (1997) idea, expressed in *Look, Listen, Read*, that systems of perception in different senses can be homologous without necessarily being identical. In particular, his commentary on Father Castel’s color organ (see Lévi-Strauss, 1997, p. 129 and ff) exemplifies this: the device did not conflate sound and color into a single unified percept, but rather highlighted their relational differences. This relational interpretation also aligns with Baumgarten (1750) interpretation of aesthetics as, in Leibnizian terms, the capacity to grasp “unity-in-multiplicity” among sensible qualities. Here, one might also consider Septimus Piesse’s “Gamut of Odors” (Piesse, 1867) in which he matched 24 musical notes to a range of scents. While a casual reading might suggest that Piesse is proposing a direct crossmodal mapping between specific musical notes and odors, a closer reading of this English chemist’s work, suggests instead that he had more of a relational mapping in mind; meaning that successful combinations of odors can be predicted on the basis of pairing associated notes that go well together. In his treatise, Piesse explicitly noted how sounds and odors blend together similarly, producing different degrees of “a nearly similar impression” in the sensory nerves (Piesse, 1867, p. 39). Piesse also writes about how the mixture needed to prepare the odors for the handkerchief evokes effects on the smelling nerve “similar to that which music or the mixture of harmonious sounds produces upon the nerve of hearing, that of pleasure” (Piesse, 1867, p. 219). He also claims that creating a mixture of scents is like creating a mixture of sounds, i.e., chords. At one point, he even writes that: “We have citron, lemon, orange peel, and verbena, forming a higher octave of smells, which blend in a similar manner” (Piesse, 1867, p. 39). Piesse was seemingly convinced that the pleasantness of musical harmony resembles that of perfumes (consisting of various base notes), and he presented a scale of correspondences between sounds and odors, as he believed that “there is, as it were, an octave of odors like an octave in music” (Piesse, 1867, p. 38).

In this light, mid-level crossmodal correspondences deserve renewed attention—not as mere by-products of low-level sensory

20 I.e., just as has been suggested in the case of Scriabin due to the crossmodal correspondence between light and sound in *Poem of Fire* (Wells, 1980; see also Boltz, 2013; see also Spence and Di Stefano, 2025).

21 Although such audiovisual entertainments seemingly lost some of their magic subsequently.

matching or high-level conceptual metaphor, but as hybrid, emergent forms of intersensory pattern recognition. One key research question here thus becomes: What sensory feature, structural property, or dynamic pattern drives these mid-level correspondences? They often resist easy categorization because they operate at the cusp of the perceptual and the conceptual, exhibiting characteristics of both. One might refer here to the “collideroscopic sensorium” (Howes, 2023) as a useful metaphor for understanding how such intersensory correspondences open novel aesthetic dimensions without collapsing sensory boundaries.

5.4 Augmenting cinematic entertainment

In their 1928 “Statement on Sound”, Eisenstein, Pudovkin, and Alexandrov advocated the role of sound in film and for “the creation of a new orchestral counterpoint of sight-images and sound-images” (see Eisenstein et al., 1999). Still more radically, in an essay called “Synchronization of Senses”, Eisenstein describes how a “single, unifying sound-picture image” might be developed as a “polyphonic structure” that “achieves its total effect through the composite sensation of all the pieces as a whole” [Eisenstein, 1947; see Spence and Di Stefano (submitted)¹⁹]. Pointing to examples of synesthetic art including Rimbaud’s “color sonnet” and James McNeill Whistler’s “color symphonies”, Eisenstein proposed that such effects could also be achieved in the cinema. At one point Eisenstein writes that: “To remove the barriers between sight and sound, between the seen world and the heard world ... To bring about a unity and a harmonious relationship between these two opposite spheres. What an absorbing task! The Greeks and Diderot, Wagner and Scriabin—who has not dreamt of this ideal?” It is the ideal of the Gesamtkunstwerk, the total work of art that synthesizes multiple media: precisely the reverse of the Laocoön argument (see also Cardullo, 2017; Kandinsky, 1977; Smith, 2007). Here, note we see common themes emerging across different artistic/entertainment media.

Note that film and advertising are different inasmuch as people assume that the music and visuals have been deliberately paired in order to achieve some emotional effect or convey some form of semantic meaning (see Bolivar et al., 1994; Boltz, 2001, 2004; Boltz et al., 1991; Bullerjahn and Gildenring, 1994; Chion, 1990, 2000; Clemente et al., 2023; Cohen, 1990, 2010; Herget and Albrecht, 2022; Hung, 2000; Klein et al., 2021; Lipscomb and Kendall, 1994; Lipscomb and Tolchinsky, 2005; Millet et al., 2021; Nosal et al., 2016; Parke et al., 2007; Silas et al., 2024; Steffens, 2020; Thayer and Levenson, 1983; Tan et al., 2017, 2007; see also Carr and Rickard, 2016; Marin et al., 2012). For instance, Hung highlights how music in TV advertising can be analyzed in terms of its sensual properties, but also in terms of its emotional effect, and there again at the level of any semantic effects.²² Meanwhile, Gorbman (1987) suggested that any music applied to a film segment will affect the spectator because the latter automatically imposes meaning on such combinations. As such, there is likely to be an important difference between (artistic) media. Nevertheless,

it is still interesting to note how many of the same themes and experimental manipulations (e.g., relating to the impact of congruent vs. incongruent combinations of film content and music stimuli) crop up in the literature on the effect of music on visual perception. Unfortunately, however, there is simply no opportunity (nor space) to review this parallel literature here. All we can do here is to highlight the existence of this large and highly-relevant experimental literature.

6 Challenges to the crossmodal augmentation of art

Before concluding this narrative historical review, it is worth noting how there are a number of challenges, theoretical, practical, attentional, and potentially also ethical when it comes to the crossmodal augmentation of art.

6.1 Theoretical challenges

A clear distinction between crossmodal augmentation in art (as in Scriabin, 1910, *Poem of Fire*) and apparently related attempts to translate auditory stimuli into visual ones, as in the case of Rimington’s Color Organ (see also Kargon, 2011), should be highlighted here. In the latter case, at least according to the creator’s intention, we are dealing with the straightforward attempt as sensory translation between audition and vision (Spence and Di Stefano, 2024b). In the case of Scriabin, there is a possible translation based on the composer’s own color-music synaesthesia followed by a combination of the two stimuli, possibly in order to help disambiguate certain elements of the musical composition visually. It is, however, likely that some attempts at augmentation may result in nothing more than redundancy of information across sensory domains—similar to both seeing a battle and reading a description of it. This is what Banes (2001) describes, in the case of olfactory stimuli being used to provide the smell of what can be seen on stage, as pleonastic. It can be argued, by contrast, that crossmodal augmentation should by definition involve, more than mere translation or redundancy; it requires the addition of non-redundant sensory inputs from a different domain, enriching (or possibly disambiguating) the original experience.

Relatedly, one might draw attention to issues related to the sensory nature of artworks by observing that to have either sensory translation or crossmodally augmented art, the mere simultaneous perception of two sensory inputs is insufficient. For example, hearing Beethoven while viewing a Munch exhibit does not constitute sensory translation or augmentation. And what if a contemporary artist *deliberately* broadcasts Beethoven during the exhibition, explicitly presenting it as multisensory art, a ticket paid for it, and people made aware of this? Would that turn the mere coexistence of music and paintings into an example of crossmodal art? Let us return, then, to Danto (1981) with his key questions on the ontology of art, such as, for example: What makes something an artwork? What distinguishes art from mere objects? How does context shape art’s identity? Danto’s reflections challenge the idea

²² Note that semantic correspondences have also been shown to influence perception crossmodally (see Chen and Spence, 2010, 2011).

that art is defined by its material properties alone, emphasizing instead its conceptual and contextual dimensions.

The crossmodal augmentation of art can therefore be seen as requiring more than mere simultaneous presentation of the relevant auditory and visual stimuli. It involves a deliberate, non-redundant integration of both, framed within an artistic context that invites interpretation. Crucially, the observer's awareness of this intentional combination transforms the experience from mere coexistence into a meaningful, augmented artistic phenomenon.

6.2 Practical challenges

It is worth pausing here to consider Behne's (1999) "habituation hypothesis", according to which the effect of background music has steadily decreased due to the growing omnipresence and availability of music in day-to-day life. In a meta-analysis, Behne observed that the effect of music on emotions, perceptions, and behavior had decreased by roughly 8% per decade. Such habituation, were it to be a replicable phenomenon, would obviously raise a practical challenge for those wanting to augment visual art with music, say. That said, Kämpfe et al. (2011) were unable to replicate this general historical trend in a more recent and more methodically controlled meta-analysis.

Separately, should any putative crossmodal augmentation rely on precise crossmodal temporal synchronization then there may be challenges around ensuring that different sensory inputs/elements align perfectly in real time (e.g., matching sound with visual effects). This can be problematic: Interactive or AI-driven experiences sometimes suffer from technology-induced lag in one or more senses, thus reducing the immersiveness of the experience. One might be reminded here of Bruce Naumann's (1969) artwork "Lip Sync" (see <https://www.moma.org/collection/works/107669>). The work consists of a close-up of a male face presented upside down repeatedly say "lip-sync". The key point is that the asynchrony between the sound of the voice and the associated lip-movements continuously change in terms of the asynchrony thus drawing the viewer is as they can never quite pin down the relationship between the auditory and visual channels. One might consider this a genuinely audiovisual work of art, and as such, the notion of crossmodal augmentation doesn't really have any relevance.

6.3 Attentional constraints and the danger of sensory overload

The Danish philosopher Kierkegaard (1988, originally published in 1843) once wrote that: "It is a common experience that to strain two senses at the same time is not pleasant, and thus it is often disruptive to have to use the eyes a great deal at the same time as the ears are being used. Therefore, one is inclined to shut the eyes when listening to music. This is more or less true of all music, and *in sensu eminentiori* [in an eminent sense] of [Mozart's opera] Don Giovanni. As soon as the eyes are involved, the impression is disrupted, for the dramatic unity that presents itself to the eye is altogether subordinate and deficient in comparison with the musical unity that is heard simultaneously.

My own experience has convinced me of this". One might think of a view of artistic experience as grounded in sensory fragmentation compared to the multisensory inputs we ordinarily have to manage in daily life.

One argument suggested here is that part of what makes art powerful is its ability to isolate, emphasize, or distill specific aspects of perception. Unlike daily life, where we are bombarded with multisensory input that we must constantly filter, artistic experiences often provide a kind of structured reduction—a selective framing of reality that allows for deeper contemplation. For instance, a still life with fruits and bottles strips away movement, sound, odors, and so forces us to focus on composition, color, and form instead. This could explain why we value these aesthetic experiences, because they allow us to maximize sensory engagement. Reintegrating sensory modalities in aesthetic experience to augment them implies adding more sensory information (haptics, olfaction, spatial audio, interactive visuals, etc.). The danger is that this may merely increase cognitive load, making it harder to focus on any one aspect, thus undermining the essence of art and artistic contemplation. Is multisensoriality in art a form of redundancy—where different sensory modalities repeat the same message—or does it truly open new aesthetic dimensions?²³

The argument seemingly has some cognitive implications. Freeing the senses from the burden of managing excessive stimuli implies freeing intellectual resources for deeper reflection and cognitive engagement. A key argument in favor of art as sensory fragmentation is that it does that to free the mind, allowing for more profound conceptual engagement. By narrowing the range of sensory input, art can liberate cognitive capacity for interpretation, imagination, and emotional resonance. In minimalist art, the deliberate reduction of visual complexity forces the viewer into an active intellectual engagement with form and space. By contrast, technologically-augmented multisensory experiences might do the opposite. Consider only the use of virtual/augmented reality (VR/AR) to augment aesthetic experience (e.g., Cho et al., 2020; He et al., 2018). If the mind is busy processing an excess of sensory inputs, does it have the capacity left to engage deeply with the ideas behind the work? One conclusion that could be drawn here is that while more sensory input might feel like an enrichment, it can also lead to passivity if the experience becomes too immersive or overwhelming. Consider here only Aitamurto et al.'s (2018) study in which the impact of a mobile video see-through AR tour guide on users' engagement with art was assessed. Besides showing many positives related to the AR technology itself, the results highlighted the downsides of the same technological augmentation related to the users feeling distracted by the AR guide and worried about excessive screen time.

Certainly, not all theorists/critics have necessarily appreciated the desire to engage more senses in art: As Arnheim stated in an essay entitled, with clear reference to the distinction among the arts in Lessing's 18th-century aesthetics, *A New Laocoön*: "in their attempts to attract the audience, two media are fighting

²³ One might also be reminded here of Nordau's (1898) suggestion that synaesthetic art (as popularized in the contemporary context by the likes of Waterworth, 1997; Whitelaw, 2008) was somehow degenerate (Nordau, 1898).

each other instead of capturing it by a united effort” (Arnheim, 1957, p. 199; Babbitt, 1910)—this sounds like a problem of divided attention, and indeed problems of working memory have been suggested as a limiting factor when auditory and visual inputs are somehow independent. So, while in the best-case scenario, some sort of multisensory emergence may be experienced in the context of crossmodal art, there is always the danger of sensory overload should the various unisensory inputs not be sufficiently congruent (Malhotra, 1984). One other practical challenge that it is worth noting relates to the observation that deliberate attempts to enhance experience by deliberately introducing additional sensory inputs (e.g., in the field of sensory marketing) have, counterintuitively, often proved unsuccessful (Spence, 2021).

6.4 Is it ethical to deliberately augment a work of art crossmodally?

Finally, it is worth briefly considering the ethical question that was raised by David Lomas (from the Art History department at the University of Manchester), in relation to Flying Object’s crossmodal augmentation of four works of visual art from the permanent collection at Tate Britain through the use of scent, sound, virtual touch, and even chocolate in 2015 (see Pursey and Lomas, 2018). Given that the artists of the original works were all dead by the time of the exhibition, one can ask the question of whether it is ethical to deliberately try to modify the meaning/viewer’s experience of these important works of art by deliberately augmenting them through some form of crossmodal stimulation? Nevertheless, the situation feels importantly different from those cases where the original artist chooses to augment their own work (think Scriabin, and Jean-Michel Jarré). At the same time, the augmentation is clearly differentiated from the original experience. Here, one might consider whether one arrives at a different conclusion when thinking about the crossmodally augmented art as a (partially) different artwork. Consider here Eugène Bataille’s, *La Joconde fumant la pipe*, (1897) or Duchamp’s ready-made depicting Leonardo’s Mona Lisa with mustache (*L.H.O.O.Q.*, 1919) or Dalí’s later alteration (Self-portrait, 1954).

7 Conclusions

One of the key points to emerge from the present narrative historical review is that it may not be possible to use elemental correspondences (that is, crossmodal correspondences based on simple stimulus properties), given that works of art are mostly complex, multi-elemental, and semantically/emotionally meaningful. This conclusion seemingly contradicting Eisenstein’s suggestion that there is no “pervading law of absolute meanings and correspondences between colors and sound.” (as quoted in Harrison, 2001, p. 133–134). One might also want to consider whether crossmodal effects are more likely to be experienced if the auditory and visual stimuli are clearly linked (i.e., rather than, say, when music is treated as no more than incidental background music; see Liu, 2022). Returning to the main claims that were made in the Introduction: The division into basic, mid-level, and complex crossmodal correspondences certainly appears helpful

when considering crossmodal correspondences/effects, especially in the context of translation/augmenting art experiences. At the same time, our sense is that more impressive outcomes are likely to be achieved should researchers and practitioners move away from the (idiosyncratic) synaesthesia mind-set, and rather focus instead on the consensually share crossmodal correspondences (see also Kargon, 2011). It will undoubtedly be interesting to see how this “intersensoriality” in art develops now that theorizing has moved away from the distraction of synaesthesia that hindered the first wave of interest a little over a century ago (Dimova, 2024; Howes, 2023, 2025).

One question for the future that we haven’t had the opportunity to consider here is whether musical knowledge/expertise may affect audiovisual crossmodal correspondences and hence the way in which music and visuals interact in experience (Di Stefano et al., 2024, 2025; Di Stefano et al. (submitted)¹¹; Guo et al., 2023; Mok, 2022; Vanechkina, 1968; Walker, 1987; see also Cutietta and Haggerty, 1987). Ultimately, though, as we have just seen, there are a number of theoretical, practical, attentional, and ethical questions that may sometimes arise for those interested in augmenting works of auditory or visual art crossmodally.

Author contributions

CS: Conceptualization, Writing – original draft, Writing – review & editing. ND: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. AHRC grant helped fund this work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Actis-Grosso, R., Zavagno, D., Lega, C., Zani, A., Daneyko, O., and Cattaneo, Z. (2017). Can music be figurative? Exploring the possibility of crossmodal similarities between music and visual arts. *Psihologija* 50, 285–306. doi: 10.2298/PSI1703285A
- Adams, C. S. (1995). Artistic parallels between Arnold Schoenberg's music and painting (1908–1912). *College Music Sympos.* 35, 5–21.
- Adeli, M., Rouat, J., and Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. *Front. Hum. Neurosci.* 8:352. doi: 10.3389/fnhum.2014.00352
- Aitamurto, T., Boin, J. B., Chen, K., Cherif, A., and Shridhar, S. (2018). "The impact of augmented reality on art engagement: Liking, impression of learning, and distraction," in *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry: 10th International Conference, VAMR 2018, Held as Part of HCI International 2018* (Las Vegas, NV: Springer International Publishing), 153–171.
- Albertazzi, L., Canal, L., and Micciolo, R. (2015). Cross-modal association between materic painting and classical Spanish music. *Front. Psychol.* 6:424. doi: 10.3389/fpsyg.2015.00424
- Albertazzi, L., Canal, L., Micciolo, R., and Hachen, I. (2020). Cross-modal perceptual organization in works of art. *i-Perception*. 11, 1–22. doi: 10.1177/2041669520950750
- Allen, E. C., and Guilford, J. P. (1936). Factors determining the affective values of color combinations. *Am. J. Psychol.* 48, 643–648. doi: 10.2307/1416516
- Alves, B. (2005). Digital harmony of sound and light. *Comp. Music J.* 29, 45–54. doi: 10.1162/0148926050757179982
- Anderson, J. D. (1993). Sound and image together: Cross-modal confirmation. *Wide Angle* 15, 30–43.
- Anikin, A., and Johansson, N. (2019). Implicit associations between individual properties of color and sound. *Attent. Percp. Psychophys.* 81, 764–777. doi: 10.3758/s13414-018-01639-7
- Ansani, A., Marini, M., D'Errico, F., and Poggi, I. (2020). How soundtracks shape what we see: Analyzing the influence of music on visual scenes through self-assessment, eye tracking, and pupillometry. *Front. Psychol.* 11:556697. doi: 10.3389/fpsyg.2020.02242
- Argüelles, J. A. (1972). *Charles Henry and the Formation of a Psychophysical Aesthetic*. Chicago, IL: University of Chicago Press.
- Arnheim, R. (1957). "New Laocoon: artistic composites and the talking film," in *Film as Art* (Berkeley, CA: University of California Press), 199–230.
- Babbitt, I. (1910). *The new Laocoon: An Essay on the Confusion of the Arts*. Boston, MA: Houghton.
- Baker, J. (2002). "Prometheus and the quest for color-music: the world premiere of Scriabin's Poem of Fire with lights, New York, March 20, 1915," in *Music and Modern Art*, ed. J. Leggio (New York, NY: Routledge), 61–95.
- Banes, S. (2001). Olfactory performances. *TDR/The Drama Review* 45, 68–76. doi: 10.1162/105420401300079040
- Barlow, H., and Mollon, J. (1982). *The Senses*. Cambridge, UK: Cambridge University Press.
- Baron-Cohen, S., Harrison, J., Goldstein, L. H., and Wyke, M. (1993). Coloured speech perception: Is synaesthesia what happens when modularity breaks down? *Perception* 22, 419–426. doi: 10.1068/p220419
- Basquin, K. S. (2020). *Mary Ellen Bute: Pioneer Animator*. New Barnet, UK: John Libbey Publishing.
- Batthey, B., and Fischman, R. (2016). "Convergence of time and space," in *The Oxford Handbook of Sound and Image in Western Art*, ed. T. Kaduri (Oxford, UK: Oxford University Press), 61–82.
- Baumgarten, A. G. (1750). *Aesthetica*. Kleyb.
- Behne, K.-E. (1999). Zu einer Theorie der Wirkungslosigkeit von (Hintergrund-)Musik [To a theory of the ineffectiveness of (background) music]. *Musikpsychologie* 14, 7–23.
- Benčić, L. (2021). Graphical sound – From inception up to the masterpieces. *Academia Lett.* 2021:1108. doi: 10.20935/AL1108
- Bender, M. B., Green, M. A., and Fink, M. (1954). Patterns of perceptual organization with simultaneous stimuli. *Arch. Neurol. Psychiatry* 72, 233–255. doi: 10.1001/archneurpsyc.1954.02330020101009
- Bernard, J. W. (1986). Messiaen's synaesthesia: the correspondence between colour and sound structure in his music. *Music Percept.* 4, 41–68. doi: 10.2307/40285351
- Bernstein, I. H., Eason, T. R., and Schurman, D. L. (1971). Hue-tone interaction: A negative result. *Percept. Motor Skills* 33, 1327–1330. doi: 10.2466/pms.1971.33.3f.1327
- Bhattacharya, J., and Lindsen, J. P. (2016). Music for a brighter world: brightness judgment bias by musical emotion. *PLoS ONE* 11:e0148959. doi: 10.1371/journal.pone.0148959
- Bolivar, V. J., Cohen, A. J., and Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: effects on the interpretation of visual action. *Psychomusicology* 13, 28–59. doi: 10.1037/h0094102
- Boltz, M. (2013). "Music videos and visual influences on music perception and appreciation: Should you want your MTV?," in *The Psychology of Music in Multimedia*, eds. S. L. Tan, A. Cohen, S. Lipscomb, and R. Kendall (Oxford, UK: Oxford University Press), 217–235.
- Boltz, M., Schulkind, M., and Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Mem. Cognit.* 19, 593–606. doi: 10.3758/BF03197154
- Boltz, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Percept.* 18, 427–454. doi: 10.1525/mp.2001.18.4.427
- Boltz, M. G. (2004). The cognitive processing of film and musical soundtracks. *Mem. Cognit.* 32, 1194–1205. doi: 10.3758/BF03196892
- Boltz, M. G. (2017). Auditory driving in cinematic art. *Music Percept.* 35, 77–93. doi: 10.1525/mp.2017.35.1.77
- Boltz, M. G. (2018). Auditory driving and affective influences. *Appl. Cogn. Psychol.* 32, 512–517. doi: 10.1002/acp.3420
- Bond, B., and Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Percept. Psychophys.* 6, 337–339. doi: 10.3758/BF03212787
- Bonetti, L., and Costa, M. (2018). Pitch-verticality and pitch-size cross-modal interactions. *Psychol. Music* 46, 340–356. doi: 10.1177/0305735617710734
- Bonetti, L., and Costa, M. (2019). Musical mode and visual-spatial cross-modal associations in infants and adults. *Musicae Scientiae* 23, 50–68. doi: 10.1177/1029864917705001
- Boswell, S. (2012). "Music in cinema: how soundtrack composers act on the way people feel, keynote speech on Music and emotions: compositions perspectives," in *Presented at the 9th International Symposium CMMR 2012 on Music and Emotions, London, UK, June 19–22* (Berlin: Springer).
- Bragdon, C. (1916). Song and light. *Architect. Record* 4, 169–172.
- Bragdon, C. (1918). *Architecture and Democracy*. New York, NY: Lector House.
- Braun Janzen, T., de Oliveira, B., Venter Ferreira, G., Sato, J. R., Feitosa-Santana, C., and Vanzella, P. (2023). The effect of background music on the aesthetic experience of a visual artwork in a naturalistic environment. *Psychology of Music* 51, 16–32. doi: 10.1177/03057356221079866
- Brent-Smith, A. (1926a). Some reflections on the work of Scriabin. *The Musical Times* 67, 593–595. doi: 10.2307/911826
- Brent-Smith, A. (1926b). Some reflections on the work of Scriabin. *The Musical Times* 67, 692–694. doi: 10.2307/911978
- Bresin, R. (2005). "What is the color of that music performance?," in *Proceedings of the International Computer Music Conference* (Barcelona, Spain: International Computer Music Association), 367–370.
- Brunel, L., Carvalho, P. F., and Goldstone, R. L. (2015). It does belong together: Cross-modal correspondences influence cross-modal integration during perceptual learning. *Front. Psychol.* 6:358. doi: 10.3389/fpsyg.2015.00358
- Bullerjahn, C., and Guldénring, M. (1994). An empirical investigation of effects of film music using qualitative content analysis. *Psychomusicology* 13, 99–118. doi: 10.1037/h0094100
- Burchett, K. E. (2002). Color harmony. *Color Research and Application* 29, 28–31. doi: 10.1002/col.10004
- Caivano, J. L. (1994). Color and sound: physical and psychophysical relations. *Color Res. Appl.* 19, 126–133. doi: 10.1111/j.1520-6378.1994.tb00072.x
- Calvert, G., Spence, C., and Stein, B. E. (eds.). (2004). *The handbook of multisensory processing*. Cambridge, MA: MIT Press.
- Cardullo, R. J. (2017). Gesamtkunstwerk, synesthesia, and the Avant-garde: Wassily Kandinsky's The Yellow Sound as a work of art. *Hermeneia* 18, 5–21.
- Carr, S. M., and Rickard, N. S. (2016). The use of emotionally arousing music to enhance memory for subsequently presented images. *Psychol. Music* 44, 1145–1157. doi: 10.1177/0305735615613846
- Carraturo, G., Pando-Naude, V., Costa, M., Vuust, P., Bonetti, L., and Brattico, E. (2025). The major-minor mode dichotomy in music perception. *Phys. Life Rev.* 52, 80–106. doi: 10.1016/j.plrev.2024.11.017
- Cavazos Quero, L., Lee, C.-H., and Cho, J.-D. (2021). Multi-sensory color code based on sound and scent for visual art appreciation. *Electronics* 10:1696. doi: 10.3390/electronics10141696
- Chen, L. (2013). "Synaesthetic correspondence between auditory clips and colors: an empirical study," in *Intelligent Science and Intelligent Data Engineering. ISIDE 2012. Lecture Notes in Computer Science*, eds. J. Yang, F. Fang, and C. Sun (Berlin, Heidelberg: Springer), 7751.

- Chen, Y.-C., and Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition* 114, 389–404. doi: 10.1016/j.cognition.2009.10.012
- Chen, Y.-C., and Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *J. Exp. Psychol.* 37, 1554–1568. doi: 10.1037/a0024329
- Chen, Z., Liu, Y., Wang, X., Wu, L., and Huang, J. (2025). Can sounds change the perception of pictures? Exploring the influence of semantic and spatial congruency on aesthetic perception. *Collabra: Psychol.* 11:133241. doi: 10.1525/collabra.133241
- Chiat, L. F., Ying, L. F., and Piaw, C. Y. (2013). Perception of congruence between music and movement in a rhythmic gymnastics routine. *J. Basic Appl. Sci. Res.* 3, 259–268.
- Chion, M. (1990). *Audio-Vision: Sound on Screen*. New York, NY: Columbia University Press.
- Chion, M. (2000). “Audio-vision and sound,” in *Sound*, eds. P. Kruth, and H. Stobart (Cambridge, UK: Cambridge University Press), 201–221.
- Chiou, R., and Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception* 41, 339–353. doi: 10.1068/p7161
- Chmiel, A., and Schubert, E. (2019). Psycho-historical contextualization for music and visual works: a literature review and comparison between artistic mediums. *Front. Psychol.* 10:182. doi: 10.3389/fpsyg.2019.00182
- Cho, J. D., Jeong, J., Kim, J. H., and Lee, H. (2020). Sound coding color to improve artwork appreciation by people with visual impairments. *Electronics* 9:1981. doi: 10.3390/electronics9111981
- Clark, C. C., and Lawless, H. T. (1994). Limiting response alternatives in time-intensity scaling: an examination of the halo dumping effect. *Chem. Senses* 19, 583–594. doi: 10.1093/chemse/19.6.583
- Classen, C. (1998). *The Color of Angels: Cosmology, Gender and the Aesthetic Imagination*. London, UK: Routledge.
- Clemente, A., Friberg, A., and Holzapfel, A. (2023). Relations between perceived affect and liking for melodies and visual designs. *Emotion* 23, 1584–1605. doi: 10.1037/emo0001141
- Cohen, A. J. (1990). Associationism and musical soundtrack phenomena. *Contemp. Music Rev.* 9, 163–178. doi: 10.1080/07494469300640421
- Cohen, A. J. (2010). “Music as a source of emotion in film,” in *Handbook of Music and Emotion: Theory, Research, Applications*, eds. P. N. Juslin and J. A. Sloboda (Oxford, UK: Oxford University Press), 879–908.
- Cohen, N. E. (1934). Equivalence of brightness across modalities. *Am. J. Psychol.* 46, 117–119. doi: 10.2307/1416240
- Collier, W. G., and Hubbard, T. L. (2001). Musical scales and evaluations of happiness and awkwardness: Effects of pitch, direction, and scale mode. *Am. J. Psychol.* 114, 355–375. doi: 10.2307/1423686
- Collopy, F. (2000). Colour, form, and motion – dimensions of a musical art of light. *Leonardo* 33, 355–360. doi: 10.1162/002409400552829
- Conrad, D. (1999). The Dichromaccord: reinventing the elusive color organ. *Leonardo* 32, 393–398. doi: 10.1162/002409499553631
- Cook, L. A., and Van Valkenburg, D. L. (2009). Audio-visual organization and the temporal ventriloquism effect between grouped sequences: evidence that unimodal grouping precedes cross-modal integration. *Perception* 38, 1220–1233. doi: 10.1068/p6344
- Cowles, J. T. (1935). An experimental study of the pairing of certain auditory and visual stimuli. *J. Exp. Psychol.* 18, 461–469. doi: 10.1037/h0062202
- Critchley, M. (1977). “Ecstatic and synaesthetic experiences during music perception,” in *Music and the Brain*, eds. M. Critchley, and R. A. Henson (Springfield, IL: C. C. Thomas), 217–232.
- Cuddy, L. L. (1985). The color of melody. *Music Percept.* 2, 345–360. doi: 10.2307/40285303
- Cutietta, R. A., and Haggerty, K. J. (1987). A comparative study of color association with music at various age levels. *J. Res. Music Educ.* 35, 78–91. doi: 10.2307/3344984
- Danto, A. C. (1981). *The Transfiguration of the Commonplace: A Philosophy of Art*. Cambridge, MA: Harvard University Press.
- Davis, J. W. (1979). A response to Garner’s observations on the relationship between colour and music. *Leonardo* 12, 218–219. doi: 10.2307/1574213
- Deroy, O., and Spence, C. (2013). Why we are not all synesthetes (not even weakly so). *Psychon. Bull. Rev.* 20, 643–664. doi: 10.3758/s13423-013-0387-2
- Deutsch, J. (2012). “Synaesthesia and synergy in art. Gustav Mahler’s “Symphony No. 2 in C minor” as an example of interactive music visualization,” in *Sensory Perception—Mind and Matter*, eds. F. G. Barth, P. Giampieri-Deutsch, and H.-D. Klein (Vienna: Springer), 215–235.
- Di Stefano, N., Ansani, A., Schiavio, A., Saarikallio, S., and Spence, C. (2025). Audiovisual associations in Saint-Saëns’ carnival of the animals: a cross-cultural investigation on the role of timbre. *Empi. Stud. Arts.* 43:2. doi: 10.1177/02762374241308810
- Di Stefano, N., Ansani, A., Schiavio, A., and Spence, C. (2024). Prokofiev was (almost) right: A cross-cultural exploration of auditory-conceptual associations in Peter and the Wolf. *Psychon. Bull. Rev.* 31, 1735–1744. doi: 10.3758/s13423-023-02435-7
- Di Stefano, N., and Spence, C. (2022). Roughness: A multisensory/crossmodal perspective. *Atten. Percept. Psychophys.* 84, 2087–2114. doi: 10.3758/s13414-022-02550-y
- Di Stefano, N., and Spence, C. (2024a). Perceptual similarity: Insights from the crossmodal correspondences. *Rev. Philos. Psychol.* 15, 997–1026. doi: 10.1007/s13164-023-00692-y
- Di Stefano, N., and Spence, C. (2024b). Should absolute pitch be considered as a unique example of context-free sensory judgments in humans? *Cognition* 249:105805. doi: 10.1016/j.cognition.2024.105805
- Di Stefano, N., and Spence, C. (2025). “Perceiving temporal structure within and between the senses: A crossmodal/multisensory perspective,” in *Attention, Perception, and Psychophysics* (Cham: Springer). doi: 10.3758/s13414-025-03045-2
- Di Stefano, N., Vuust, P., and Brattico, E. (2022). Consonance and dissonance perception. A critical review of the historical sources, multidisciplinary findings, and main hypotheses. *Phys. Life Rev.* 43, 273–304. doi: 10.1016/j.plrev.2022.10.004
- Dimova, P. (2024). *At the Crossroads of the Senses: The Synaesthetic Metaphor Across the arts in European Modernism*. University Park, PA: Pennsylvania State University Press.
- Donnell-Kotrozo, C. (1978). Intersensory perception of music: Color me trombone. *Music Educators J.* 65, 32–37. doi: 10.2307/3395546
- Duthie, A. C. (2013). *Do Music and Art Influence One Another? Measuring Cross-Modal Similarities in Music and Art* (PhD thesis). Ames: Iowa State University.
- Duthie, C., and Duthie, B. (2015). Do music and art influence one another? Measuring cross-modal similarities in music and art. *Polymath.* 5, 1–22.
- Ehrenzweig, A. (1965). *The Psychoanalysis of Artistic Vision and Hearing*. New York, NY: Braziller.
- Eisenstein, S., Pudovkin, W. I., and Alexandrov, G. V. (1999). “Statement on sound,” in *Close up: 1927-1933*, eds. J. Donald, A. Friedberg, and L. Marcus (Princeton: Princeton University Press), 83–86.
- Eisenstein, S. M. (1947). *The Film Sense* (New York, NY: Harcourt).
- Ellermeier, W., Kattner, F., and Raum, A. (2021). Cross-modal commutativity of magnitude productions of loudness and brightness. *Atten. Percept. Psychophys.* 83, 2955–2967. doi: 10.3758/s13414-021-02324-y
- Ellis, R. J., and Simons, R. F. (2005). The impact of music on subjective and physiological indices of emotion while viewing films. *Psychomusicology* 19, 15–40. doi: 10.1037/h0094042
- Evans, K. K., and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *J. Vision* 10, 1–12. doi: 10.1167/10.1.6
- Eysenck, H. J. (1940). The general factor in aesthetic judgements. *Br. J. Psychol.* 31, 94–102. doi: 10.1111/j.2044-8295.1940.tb00977.x
- Fekete, A., Specker, E., Mikuni, J., Trupp, M. D., and Leder, H. (2023). “When the painting meets its musical inspiration: The impact of multimodal art experience on aesthetic enjoyment and subjective well-being in the museum,” in *Psychology of Aesthetics, Creativity, and the Arts*. Available online at: <https://psycnet.apa.org/doi/10.1037/aca0000641>
- Field, G. (1835). *Chromatics: Or the Analogy, Harmony, and Philosophy of Colours*. London, UK: David Bogue.
- Fink, L., Fiehn, H., and Wald-Fuhrmann, M. (2024). The role of audiovisual congruence in aesthetic appreciation of contemporary music and visual art. *Sci. Rep.* 14:20923. doi: 10.1038/s41598-024-71399-y
- Frings, C., and Spence, C. (2010). Crossmodal congruency effects based on stimulus identity. *Brain Res.* 1354, 113–122. doi: 10.1016/j.brainres.2010.07.058
- Gaiger, J. (2018). Can a painting have a rhythm? *Br. J. Aesthet.* 58, 363–383. doi: 10.1093/aesth/ajay026
- Galeyev, B. M. (1988). The fire of Prometheus: Music-kinetic art experimentation in the USSR. *Leonardo* 21, 383–391. doi: 10.2307/1578701
- Galeyev, B. M. (2003). Evolution of gravitational synesthesia in music: to color and light. *Leonardo* 36, 129–134. doi: 10.1162/002409403321554198
- Galeyev, B. M., and Vanechikina, I. L. (2001). Was Scriabin a synesthete? *Leonardo* 34, 357–361. doi: 10.1162/00240940152549357
- Gallace, A., and Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Percept. Psychophys.* 68, 1191–1203. doi: 10.3758/BF03193720

- Gallace, A., and Spence, C. (2014). "The neglected power of touch: What cognitive neuroscience tells us about the importance of touch in artistic communication," in *Sculpture and Touch*, ed. P. Dent (Farnham: Ashgate Publishers), 107–124.
- Garner, W. (1978). The relationship between colour and music. *Leonardo* 11, 225–226. doi: 10.2307/1574153
- Gawboy, A. M., and Townsend, J. (2012). Scriabin and the possible. *MTO* 18, 1–21. doi: 10.30535/mto.18.2.2
- Gebhard, J. W., and Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *Am. J. Psychol.* 72, 521–528. doi: 10.2307/1419493
- Giannos, K., Athanasopoulos, G., and Cambouropoulos, E. (2021). Cross-modal associations between harmonic dissonance and visual roughness. *Music Sci.* 4:20592043211055484. doi: 10.1177/20592043211055484
- Gibson, S. (2023). "Moving towards the performed image (colour organs, synesthesia and visual music): early Modernism (1900–1955)," in *Live Visuals: History, Theory, Practice*, eds. S. Gibson, S. Arisona, D. Leishman, and A. Tanaka (London, UK: Routledge), 41–61.
- Gilbert, G. M. (1938). A study in inter-sensory Gestalten. *Psychol. Bull.* 35:698.
- Gombrich, E. H. (1960). *Art and Illusion*. London, UK: Phaidon Press.
- Gorbman, C. (1987). *Unheard Melodies: Narrative Film Music*. Bloomington, IA: Indiana University Press.
- Guo, Q., and Jiang, T. (2023). "Using crossmodal correspondence between colors and music to enhance online art exhibition visitors' experience," in *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity* (Cham: Springer Nature Switzerland).
- Guo, X., Qu, J., Liu, M., Liu, C., and Huang, J. (2023). Dynamic audio-visual correspondence in musicians and non-musicians. *Psychol. Music* 52, 175–186. doi: 10.1177/03057356231185467
- Gurman, D., McCormick, C., and Klein, R. M. (2021). Crossmodal correspondence between auditory timbre and visual shape. *Multisens. Res.* 35, 221–241. doi: 10.1163/22134808-bja10067
- HaCohen, R. (2016). "Between generation and suspension," in *The Oxford Handbook of Sound and Image in Western Art*, ed. T. Kaduri (Oxford, UK: Oxford University Press), 36–60.
- Hagtvedt, H., and Brasel, S. A. (2016). Crossmodal communication: Sound frequency influences consumer responses to color lightness. *J. Market. Res.* 53, 551–562. doi: 10.1509/jmr.14.0414
- Hankins, T. L. (1994). The ocular harpsichord of Louis-Bertrand Castel; Or, the instrument that wasn't. *Osiris* 9, 141–156. doi: 10.1086/368734
- Harrison, J. (2001). *Synaesthesia: The Strangest Thing*. Oxford, UK: Oxford University Press.
- Harrison, J., and Baron-Cohen, S. (1996). Acquired and inherited forms of cross-modal correspondence. *Neurocase* 2, 245–249. doi: 10.1080/13554799608402401
- Hartshorne, C. (1934). *The Philosophy and Psychology of Sensation*. Chicago, IL: University of Chicago Press.
- Hasenbus, N., Martindale, C., and Birnbaum, D. (1983). Psychological reality of cross-media artistic styles. *J. Exp. Psychol.* 9, 841–863. doi: 10.1037/0096-1523.9.6.841
- Hauck, P., von Castell, C., and Hecht, H. (2022). Crossmodal correspondence between music and ambient color is mediated by emotion. *Multisens. Res.* 35, 407–446. doi: 10.1163/22134808-bja10077
- Haverkamp, M. (2020). "Light, color and motion as crossmodal elements of baroque music," in *e-Forum Acusticum 2020, Lyon, December 7–10th*. doi: 10.48465/fa.2020.0021
- He, Z., Wu, L., and Li, X. R. (2018). When art meets tech: The role of augmented reality in enhancing museum experiences and purchase intentions. *Tour. Managem.* 68, 127–139. doi: 10.1016/j.tourman.2018.03.003
- Helmholtz, H. (1878/1971). *Treatise on Physiological Optics (Vol. II)*. New York, NY: Dover Publications.
- Herget, A. K., and Albrecht, J. (2022). Soundtrack for reality? How to use music effectively in non-fictional media formats. *Psychol. Music* 50, 508–529. doi: 10.1177/0305735621999091
- Holm, J., Aaltonen, A., and Siirtola, H. (2009). Associating colours with musical genres. *J. New Music Res.* 38, 87–100. doi: 10.1080/09298210902940094
- Hong, Y. J., Choi, A., Lee, C.-E., Cho, W., Yoon, S., and Lee, K. (2024). Concurrent musical pitch height biases judgment of visual brightness. *Psychol. Music* 53:3. doi: 10.1177/03057356231216950
- Howes, D. (2025). The new intersensory music and art history: Review of "Sensuous Consonance: The Visual Arts in Conjunction with Music Around 1900", organized by Linda Hinners and Tobias Lund, Royal Swedish Academy of Music and the Swedish Nationalmuseum, Stockholm, 4 October 2024. *Senses Soc.* 20, 143–150. doi: 10.1080/17458927.2025.2463766
- Howes, D., and Classen, C. (2014). *Ways of Sensing: Understanding the Senses in Society*. London, UK: Routledge.
- Howes, H. (2023). The colliderscopic sensorium. *HAU: J. Ethnographic Theory* 13, 730–734. doi: 10.1086/728409
- Huang, J., Gamble, D., Sarniertsophon, K., Wang, X., and Hsiao, S. (2012). Feeling music: Integration of auditory and tactile inputs in musical meter perception. *PLoS ONE* 7:e48496. doi: 10.1371/journal.pone.0048496
- Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *Am. J. Psychol.* 109, 219–238. doi: 10.2307/1423274
- Huddleston, W. E., Lewis, J. W., Phinney, R. E., and DeYoe, E. A. (2008). Auditory and visual attention-based apparent motion share functional parallels. *Percept. Psychophys.* 70, 1207–1216. doi: 10.3758/PP.70.7.1207
- Hull, A. E. (1927). *A great Russian tone poet: Scriabin* (2nd Ed.). London, UK: Trübner and Co.
- Hung, K. (2000). Narrative music in congruent and incongruent TV advertising. *J. Advert.* 29, 25–34. doi: 10.1080/00913367.2000.10673601
- Iosifyan, M., Sidoroff-Dorso, A., and Wolfe, J. (2022). Cross-modal associations between paintings and sounds: effects of embodiment. *Perception* 51, 871–888. doi: 10.1177/03010066221126452
- Isbilen, E. S., and Krumhansl, C. L. (2016). The color of music: emotion-mediated associations to Bach's Well-Tempered Clavier. *Psychomusicol.: Music, Mind Brain* 26, 149–161. doi: 10.1037/pmu0000147
- Iwamiya, S. (1994). Interaction between auditory and visual processing when listening to music in an audiovisual context: 1. Matching 2. Audio quality. *Psychomusicology* 13, 133–154. doi: 10.1037/h0094098
- Iwamiya, S. (2013). "Perceived congruence between auditory and visual elements in multimedia," in *The Psychology of Music in Multimedia*, eds. S.-L. Tan, A. J. Cohen, S. D. Lipscomb, and R. A. Kendall (Oxford, UK: Oxford University Press), 141–164.
- Jewanski, J. (2010). "Color-tone analogies: A systematic presentation of the principles of correspondence," in *Audiovisuality: A Multidisciplinary Survey of Audiovisual Culture*, eds. D. Daniels, S. Naumann, and J. Thoben (Köln: König), 77–87.
- Johnson, D., Allison, C., and Baron-Cohen, S. (2013). "The prevalence of synesthesia: The consistency revolution," in *Oxford Handbook of Synaesthesia*, eds. J. Simner and E. Hubbard (Oxford: Oxford Academic).
- Jones, C. A., Hasegawa, Y., Jacobson, M., and Arning, B. (2006). *Sensorium: Embodied Experience, Technology, and Contemporary Art*. Cambridge, MA: MIT Press.
- Julesz, B., and Hirsh, I. J. (1972). "Visual and auditory perception - An essay of comparison," in *Human Communication: A Unified View*, eds. E. E. David, Jr., and P. B. Denes (New York, NY: McGraw-Hill), 283–340.
- Kaduri, T. (2016). *The Oxford Handbook of Sound and Image in Western Art*. Oxford, UK: Oxford University Press.
- Kämpfe, J., Sedlmeier, P., and Renkewitz, F. (2011). The impact of background music on adult listeners. A meta-analysis. *Psychol. Music* 39, 424–448. doi: 10.1177/0305735610376261
- Kandinsky, W. (1977). *Concerning the Spiritual in art, Especially in Painting*. New York, NY: Dover Publications.
- Kargon, J. (2011). Harmonizing these two arts: Edmund Lind's The Music of Color. *J. Des. Hist.* 24, 1–14. doi: 10.1093/jdh/epq042
- Karwowski, T. F., Odbert, H. S., and Osgood, C. E. (1942). Studies in synesthetic thinking. II. The rôle of form in visual responses to music. *J. General Psychol.* 26, 199–222. doi: 10.1080/00221309.1942.10545166
- Kastner, F. (1875). *Invention du pyrophone: Expériences nouvelles sur les flammes chantantes* (2nd Ed.). Paris: Typographie de A. Parent.
- Kierkegaard, S. (1988). *Either/Or: Part 1*, eds. H. V. Hong and E. H. Hong (Princeton, NJ: Princeton University Press), 119–120.
- Klapetek, A., Ngo, M. K., and Spence, C. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? *Attent. Percept. Psychophys.* 74, 1154–1167. doi: 10.3758/s13414-012-0317-9
- Klein, A. B. (1937). *Coloured Light: An Art Medium (Third Enlarged Edition of The Art of Light: Colour Music)*. London, UK: Crosby Lockwood and Son.
- Klein, K., Melnyk, V., and Völckner, F. (2021). Effects of background music on evaluations of visual images. *Psychol. Market.* 38, 2240–2246. doi: 10.1002/mar.21588
- Kubovy, M., and Yu, M. (2012). Multistability, cross-modal binding and the additivity of conjoint grouping principles. *Philosoph. Trans. Royal Soc. B.* 367, 954–964. doi: 10.1098/rstb.2011.0365
- Lauwrens, J. (2012). Welcome to the revolution: the sensory turn and art history. *J. Art Historiography* 7, 1–17.
- Legrenzi, P., and Umiltà, C. (2011). *Neuromania: On the limits of brain science (translated by F. Anderson)*. Oxford: Oxford University Press.
- Leibniz, G. W. (1896). *On Solidity. New Essays Concerning Human Understanding (Originally published, 1704)*. New York, NY: Macmillan.
- Lévi-Strauss, C. (1997). *Look, Listen, and Read*. New York, NY: Basic Books.

- Lewkowicz, D. J., and Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Dev. Psychol.* 16, 597–607. doi: 10.1037/0012-1649.16.6.597
- Limbert, W. M., and Polzella, D. J. (1998). Effects of music on the perception of paintings. *Empir. Stud. Arts* 16, 33–39. doi: 10.2190/V8BL-GBJK-TLFP-R321
- Lin, C., Yeh, M., and Shams, L. (2022). Subliminal audio visual temporal congruency in music videos enhances perceptual pleasure. *Neurosci. Lett.* 779:136623. doi: 10.1016/j.neulet.2022.136623
- Lindborg, P. M., and Friberg, A. K. (2015). Colour association with music is mediated by emotion: Evidence from an experiment using a CIE Lab interface and interviews. *PLoS ONE* 10:e0144013. doi: 10.1371/journal.pone.0144013
- Lindholm, P. (2022). “A light in sound, a sound-like power in light”: Poetry, synaesthesia and multimedia. *Études de lettres* 319, 77–103. doi: 10.4000/edl.4009
- Lindner, D., and Hynan, M. T. (1987). Perceived structure of abstract paintings as a function of structure of music listened to on initial viewing. *Bull. Psychon. Soc.* 25, 44–46. doi: 10.3758/BF03330072
- Lippman, E. A. (1963). Hellenic conceptions of harmony. *J. Am. Musicol. Soc.* 16, 3–35. doi: 10.2307/829917
- Lipscomb, S. D. (2013). “Cross-modal alignment of accent structures in multimedia. The psychology of music in multimedia,” in *The psychology of Music in Multimedia*, eds. S. Tan, A. J. Cohen, S. D. Lipscomb, and R. A. Kendall (Oxford, UK: Oxford University Press), 192–215.
- Lipscomb, S. D., and Kendall, R. A. (1994). Perceptual judgment of the relationship between musical and visual components in film. *Psychomusicology* 13, 60–98. doi: 10.1037/h0094101
- Lipscomb, S. D., and Kim, E. M. (2004). “Perceived match between visual parameters and auditory correlates: an experimental multimedia investigation,” in *Proceedings of the 8th International Conference on Music Perception and Cognition* (Evanston, IL), 72–75.
- Lipscomb, S. D., and Tolchinsky, D. E. (2005). “The role of music communication in cinema,” in *Musical Communication*, eds. D. Miell, R. Macdonald, and D. J. Hargreaves (Oxford: Oxford University Press), 383–404.
- Liu, M. (2022). Paintings of music. *J. Aesthet. Art Critic.* 80, 151–163. doi: 10.1093/jaac/kpac003
- Locke, J. (1690). *An Essay Concerning Human Understanding, Book III*. Paris, France: Basset.
- Logeswaran, N., and Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neurosci. Lett.* 455, 129–133. doi: 10.1016/j.neulet.2009.03.044
- London, I. D. (1954). Research on sensory interaction in the Soviet Union. *Psychol. Bull.* 51, 531–568. doi: 10.1037/h0056730
- Loureiro, S. M. C., Roschk, H., and Lima, F. (2019). The role of background music in visitors' experience of art exhibitions: Music, memory and art appraisal. *Int. J. Arts Managem.* 22, 4–24.
- Lupton, E. (2018). “Visualizing sound,” in *The Senses: Design Beyond Vision*, eds. E. Lupton and A. Lipps (Hudson, NY: Princeton Architectural Press), 204–217.
- Maeda, F., Kanai, R., and Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Curr. Biol.* 14, R990–R991. doi: 10.1016/j.cub.2004.11.018
- Malhotra, N. K. (1984). Information and sensory overload. Information and sensory overload in psychology and marketing. *Psychol. Market.* 1, 9–21. doi: 10.1002/mar.4220010304
- Marin, M. M., Gingras, B., and Bhattacharya, J. (2012). Crossmodal transfer of arousal, but not pleasantness, from the musical to the visual domain. *Emotion* 12, 618–631. doi: 10.1037/a0025020
- Marks, L. E. (1974). On associations of light and sound: the mediation of brightness, pitch, and loudness. *Am. J. Psychol.* 87, 173–188. doi: 10.2307/1422011
- Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *J. Exp. Psychol.* 13, 384–394. doi: 10.1037//0096-1523.13.3.384
- Marks, L. E. (1989). On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness. *J. Exp. Psychol.* 15, 586–602. doi: 10.1037//0096-1523.15.3.586
- Marks, L. E. (2011). Synesthesia, then and now. *Intellectica* 55, 47–80. doi: 10.3406/intel.2011.1161
- Marks, L. E., Hammeal, R. J., Bornstein, M. H., and Smith, L. B. (1987). Perceiving similarity and comprehending metaphor. *Monogr. Soc. Res. Child Dev.* 52, 1–100. doi: 10.2307/1166084
- McDonald, J., Canazza, S., Chmiel, A., De Poli, G., Houbert, E., Murari, M., et al. (2022). Illuminating music: Impact of color hue for background lighting on emotional arousal in piano performance videos. *Front. Psychol.* 13:828699. doi: 10.3389/fpsyg.2022.828699
- McNeill Whistler, J. A. (1978). *Letter to the world. May 22nd. [Reprinted in the Gentle Art of Making Enemies]*. New York, NY: Dover, 127–128.
- McQuarrie, E., and Mick, D. G. (1992). On resonance: A critical pluralistic inquiry into advertising rhetoric. *J. Consumer Res.* 19, 180–197. doi: 10.1086/209295
- Melara, R. D. (1989). Dimensional interaction between color and pitch. *J. Exp. Psychol.* 15, 69–79. doi: 10.1037//0096-1523.15.1.69
- Menouti, K., Akiva-Kabiri, L., Banissy, M. J., and Stewart, L. (2015). Timbre-colour synaesthesia: exploring the consistency of associations based on timbre. *Cortex* 63, 1–3. doi: 10.1016/j.cortex.2014.08.009
- Millet, B., Chattah, J., and Ahn, S. (2021). Soundtrack design: The impact of music on visual attention and affective responses. *Appl. Ergon.* 93:103301. doi: 10.1016/j.apergo.2020.103301
- Minnigerode, F. A., Ciancio, D. W., and Sbarboro, L. A. (1976). Matching music with paintings by Klee. *Percept. Mot. Skills* 42, 269–270. doi: 10.2466/pms.1976.42.1.269
- Mok, J. (2022). *Cross-Modal Correspondences: Different Modes, Common Codes? Investigating Musical Engagement with an Ecological Cognitive Approach* (Master's programme in Music, Communication and Technology). Department of Musicology, University of Oslo, Oslo, Norway.
- Mondloch, C. J., and Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognit. Affect. Behav. Neurosci.* 4, 133–136. doi: 10.3758/CABN.4.2.133
- Moritz, W. (1997). “The dream of color music, and machines that made it possible,” in *Animation World Magazine*, 2.1, April.
- Motoki, K., Marks, L. E., and Velasco, C. (2023). Reflections on cross-modal correspondences: Current understanding and issues for future research. *Multisens. Res.* 37, 1–23. doi: 10.1163/22134808-bja10114
- Muecke, M. W., and Zach, M. S. (2007). *Resonance: Essays on the Intersection of Music and Architecture*. Ames, IA: Culicidae Press.
- Muller, J. P. (2010). “Synchronization as a sound-image relationship,” in *See This Sound: Audiovisuality An Interdisciplinary Survey of Audiovisual Culture*, eds. D. Daniels, S. Naumann, and J. Thoben (New York, NY: Walter König), 401–413.
- Müller-Eberstein, M., and van Noord, N. (2019). “Translating visual art into music,” in *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* (Seoul: IEEE), 3117–3120.
- Murari, M., Chmiel, A., Tiepolo, E., Zhang, J. D., Canazza, S., Rodà, A., et al. (2020). “Key clarity is blue, relaxed, and maluma: Machine learning used to discover cross-modal connections between sensory items and the music they spontaneously evoke,” in *KEER 2020, AISC 1256*, eds. H. Shoji, S. Koyama, T. Kato, K. Muramatsu, T. Yamanaka, P. Lévy, K. Chen, and A. M. Lokman (Singapore: Springer Nature Singapore), 214–223.
- Murari, M., Rodà, A., Canazza, S., De Poli, G., and Da Pos, O. (2015). Is Vivaldi smooth and take? Nonverbal sensory scales for describing music qualities. *J. New Music Res.* 44, 359–372. doi: 10.1080/09298215.2015.1101475
- Myers, C. S. (1911). A case of synaesthesia. *Br. J. Psychol.* 4, 228–238. doi: 10.1111/j.2044-8295.1911.tb00045.x
- Myers, C. S. (1914). Two cases of synaesthesia. *Br. J. Psychol.* 7, 112–117. doi: 10.1111/j.2044-8295.1914.tb00246.x
- Newton, I. (1704). *Opticks*. Available online at <https://archive.org/details/opticksortreatis00newt> (Accessed June 15, 2025).
- Nordau, M. (1898). *Degeneration [English Translation of Nordau, 1892]*. London, UK: Heinemann.
- Nosal, A. P., Keenan, E. A., Hastings, P. A., and Gneezy, A. (2016). The effect of background music in shark documentaries on viewers' perceptions of sharks. *PLoS ONE* 11:e0159279. doi: 10.1371/journal.pone.0159279
- Ocelli, V., Spence, C., and Zampini, M. (2011). Audiotactile interactions in temporal perception. *Psychon. Bull. Rev.* 18, 429–454. doi: 10.3758/s13423-011-0070-4
- Odgaard, E. C., Arie, Y., and Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. *Percept. Psychophys.* 65, 123–132. doi: 10.3758/BF03194789
- Odgaard, E. C., Arie, Y., and Marks, L. E. (2004). Brighter noise: Sensory enhancement of perceived loudness by concurrent visual stimulation. *Cognit. Affect. Behav. Neurosci.* 4, 127–132. doi: 10.3758/CABN.4.2.127
- O'Leary, A., and Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. *Percept. Psychophys.* 35, 565–569. doi: 10.3758/BF03205954
- O'Malley, G. (1957). Literary synesthesia. *J. Aesthet. Art Critic.* 15, 391–411. doi: 10.2307/427153
- Ortmann, O. (1933). Theories of synesthesia in the light of a case of colored hearing. *Hum. Biol.* 5, 155–211.
- Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Urbana: University of Illinois Press.
- Palmer, S. E., and Griscorn, W. (2013). Accounting for taste: Individual differences in preference for harmony. *Psychon. Bull. Rev.* 20, 453–461. doi: 10.3758/s13423-012-0355-2
- Palmer, S. E., Langlois, T. A., and Schloss, K. B. (2016). Music-to-color associations of single-line piano melodies in non-synesthetes. *Multisens. Res.* 29, 157–193. doi: 10.1163/22134808-00002486

- Palmer, S. E., Schloss, K. B., Xu, Z., and Prado-León, L. R. (2013). Music-color associations are mediated by emotion. *Proc. Natl. Acad. Sci. USA*. 110, 8836–8841. doi: 10.1073/pnas.1212562110
- Parise, C., and Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE* 4:e5664. doi: 10.1371/journal.pone.0005664
- Parise, C. V., Harrar, V., Ernst, M. O., and Spence, C. (2013). Cross-correlation between auditory and visual signals promotes multisensory integration. *Multisens. Res.* 26, 307–316. doi: 10.1163/22134808-00002417
- Parise, C. V., and Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Exp. Brain Res.* 220, 319–333. doi: 10.1007/s00221-012-3140-6
- Parise, C. V., Spence, C., and Ernst, M. (2012). When correlation implies causation in multisensory integration. *Curr. Biol.* 22, 46–49. doi: 10.1016/j.cub.2011.11.039
- Parke, R., Chew, E., and Kyriakakis, C. (2007). Quantitative and visual analysis of the impact of music on perceived emotion of film. *Comp. Entertainm.* 5:5. doi: 10.1145/1316511.1316516
- Parncutt, R. (2014). The emotional connotations of major versus minor tonality: one or more origins? *Musicae Scientiae* 18, 324–353. doi: 10.1177/1029864914542842
- Parncutt, R. (2024). *Psychoacoustic Foundations of Major-Minor Tonality*. Cambridge, MA: MIT Press.
- Parrott, A. C. (1982). Effects of paintings and music, both alone and in combination, on emotional judgment. *Percept. Mot. Skills* 54, 635–641. doi: 10.2466/pms.1982.54.2.635
- Peacock, K. (1985). Synesthetic perception: Alexander Scriabin's color hearing. *Music Percept.* 2, 483–506. doi: 10.2307/40285315
- Peacock, K. (1988). Instruments to perform color-music: Two centuries of technological experimentation. *Leonardo* 21, 397–406. doi: 10.2307/1578702
- Peretti, P. (1972). A study of student correlations between music and six paintings by Klee. *J. Res. Music Educ.* 20, 501–504. doi: 10.2307/3343810
- Piesse, G. W. S. (1867). *The Art of Perfumery and the Methods of Obtaining the Odors of Plants: With Instructions for the Manufacture of Perfumes for the Handkerchief, Scented Powders, Olfactory Vinegars, Dentifrices, Pomatums, Cosmetics, Perfumed Soap, Etc., to which is Added an Appendix on Preparing Artificial Fruit-essences, Etc.* Philadelphia: Lindsay and Blakiston.
- Pimentel, O., Chuquichambi, E. G., Spence, C., and Velasco, C. (2025). The diatonic sound of scent imagery. *Perception*. 3:3010066251342011. doi: 10.1177/03010066251342011
- Plummer, H. C. (1915). Color music—A new art created with the aid of science. The color organ used in Scriabin's symphony "Prometheus". *Scient. Am.* 112, 343, 350–351. doi: 10.1038/scientificamerican04101915-343
- Poast, M. (2000). Color music: Visual color notation for musical expression. *Leonardo* 33, 215–221. doi: 10.1162/002409400552531
- Pridmore, R. W. (1992). Music and color: Relations in the psychophysical perspective. *Color Res. Appl.* 17, 57–61. doi: 10.1002/col.5080170110
- Pursey, T., and Lomas, D. (2018). Tate Sensorium: An experiment in multisensory immersive design. *Senses Soc.* 13, 354–366. doi: 10.1080/17458927.2018.1516026
- Rahne, T., Böckmann, M., Von Specht, H., and Sussman, E. S. (2007). Visual cues can modulate integration and segregation of objects in auditory scene analysis. *Brain Res.* 1144, 127–135. doi: 10.1016/j.brainres.2007.01.074
- Rahne, T., Deike, S., Selezneva, E., Brosch, M., König, R., Scheich, H., et al. (2008). A multilevel and cross-modal approach towards neuronal mechanisms of auditory streaming. *Brain Res.* 1220, 118–131. doi: 10.1016/j.brainres.2007.08.011
- Rančić, K., and Marković, S. (2019). The perceptual and aesthetic aspects of the music-paintings congruence. *Vision* 3:65. doi: 10.3390/vision3040065
- Ravignani, A., and Sonnweber, R. (2017). Chimpanzees process structural isomorphisms across sensory modalities. *Cognition* 161, 74–79. doi: 10.1016/j.cognition.2017.01.005
- Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *J. Neurophysiol.* 89, 1078–1093. doi: 10.1152/jn.00706.2002
- Recanzone, G. H. (2009). Interactions of auditory and visual stimuli in space and time. *Hear. Res.* 258, 89–99. doi: 10.1016/j.heares.2009.04.009
- Reuter, C., Jewanski, J., Saitis, C., Czedik-Eysenberg, I., Siddiq, S., and Oehler, M. (2018). "Colors and timbres—consistent color-timbre mappings at non-synesthetic individuals," in *Proceedings of the 34th Jahrestagung der Deutschen Gesellschaft für Musikpsychologie: Musik im audiovisuellen Kontext* (Gießen: Waxmann Verlag).
- Reymore, L., and Lindsey, D. T. (2025). Color and tone color: audiovisual crossmodal correspondences with musical instrument timbre. *Front. Psychol.* 15:1520131. doi: 10.3389/fpsyg.2024.1520131
- Reymore, L., Noble, J., Saitis, C., Traube, C., and Wallmark, Z. (2023). Timbre semantic associations vary both between and within instruments: an empirical study empirical study incorporating register and pitch height. *Music Percept.* 40, 253–274. doi: 10.1525/mp.2023.40.3.253
- Rimington, A. W. (1895). "A new art: Colour-music," in *A paper read at St. James's Hall on June 6, 1895, published in pamphlet form by Messrs Spottiswoode and Co., New St. Square. June 13, 1895. Reprinted in 'Colour Music, the Art of Light', ed. A. B. Klein* (London: Lockwood), 256–261.
- Rimington, A. W. (1911). *Colour Music: The art of Mobile Colour*. New York, NY: Frederick A. Stokes.
- Rimington, A. W. (1915). Color music. *Sci. Am.* 113, 79. doi: 10.1038/scientificamerican07241915-79
- Root, N., Chkhaidze, A., Melero, H., Sidoroff-Dorso, A., Volberg, G., Zhang, Y., et al. (2025). How "diagnostic" criteria interact to shape synesthetic behavior: the role of self-report and test-retest consistency in synesthesia research. *Conscious. Cogn.* 129:103819. doi: 10.1016/j.concog.2025.103819
- Root, R. T., and Ross, S. (1965). Further validation of subjective scales for loudness and brightness by means of cross-modality matching. *Am. J. Psychol.* 78, 285–289. doi: 10.2307/1420502
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychol. Bull.* 86, 638–641. doi: 10.1037/0033-2909.86.3.638
- Russet, R., and Starr, C. (1988). *Experimental Animation: Origins of a New Art*. Boston: Da Capo Press.
- Sabaneev, L., and Pring, S. W. (1929). The relation between sound and colour. *Music Letters* 10, 266–277. doi: 10.1093/ml/10.3.266
- Sabanev, L. (1911). On sound-color accordance. *Muzyka* 9, 196–200.
- Saitis, C., and Wallmark, Z. (2024). Timbral brightness perception investigated through multimodal interference. *Attent. Percept. Psychophys.* 86, 1835–1845. doi: 10.3758/s13414-024-02934-2
- Saitis, C., Weinzierl, S., von Kriegstein, K., Ystad, S., and Cuskey, C. (2020). Timbre semantics through the lens of crossmodal correspondences: a new way of asking old questions. *Acoust. Sci. Technol.* 41, 365–368. doi: 10.1250/ast.41.365
- Schiffstein, H. N. J., and Tanudjaja, I. (2004). Visualizing fragrances through colors: The mediating role of emotions. *Perception* 33, 1249–1266. doi: 10.1068/p5132
- Schloss, K. B., and Palmer, S. E. (2011). Aesthetic response to color combinations: Preference, harmony, and similarity. *Attent. Percept. Psychophys.* 73, 551–571. doi: 10.3758/s13414-010-0027-0
- Scholes, P. (1970). "Colour and music," in *The Oxford Companion to Music* (Oxford, UK: Oxford University Press), 202–210.
- Schwartz, J. L., Grimault, N., Hupé, J. M., Moore, B. C., and Pressnitzer, D. (2012). Multistability in perception: Binding sensory modalities, an overview. *Philosoph. Trans. Royal Soc. B: Biol. Sci.* 367, 896–905. doi: 10.1098/rstb.2011.0254
- Scriabin, A. (1910). "Prometheus, op. 60, tone-poem for orchestra and projected colour lights," in *Original Score in British Museum Library*.
- Sebba, R. (1991). Structural correspondence between music and color. *Color Res. Appl.* 16, 81–88. doi: 10.1002/col.5080160206
- Shams, L., Kamitani, Y., and Shimojo, S. (2000). What you see is what you hear. *Nature* 408, 788–788. doi: 10.1038/35048669
- Shaw Miller, S. (2006). Visual music and the case for rigorous thinking. *The Art Book* 13, 3–5. doi: 10.1111/j.1467-8357.2006.00623.x
- Shaw-Miller, S. (2010). *Eye hEar: The Visual in Music*. London, UK: Routledge.
- Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science* 145, 1328–1330. doi: 10.1126/science.145.3638.1328
- Siefkes, M., and Arielli, E. (2015). "An experimental approach to multimodality: How musical and architectural styles interact in aesthetic perception," in *Building Bridges for Multimodal Research: International Perspectives on Theories and Practices of Multimodal Analysis*, 246–265.
- Siegel, L. (1974). Synaesthesia and the paintings of Caspar David Friedrich. *Art J.* 33, 196–204. doi: 10.1080/00043249.1974.10793214
- Silas, S., Baker, D. J., and Müllensiefen, D. (2024). Musical manipulation of visual scenes in video, film, and TV advertisements: a large-scale investigation into the implicit effects of sonic branding. *J. Advert. Res.* 64, 192–212. doi: 10.2501/JAR-2024-013
- Simpson, R. H., Quinn, M., and Ausubel, D. P. (1956). Synaesthesia in children: association of colors with pure tone frequencies. *J. Genetic Psychol.* 89, 95–103. doi: 10.1080/00221325.1956.10532990
- Sirius, G., and Clarke, E. F. (1994). The perception of audiovisual relationships: a preliminary study. *Psychomusicology* 13, 119–132. doi: 10.1037/h0094099
- Smith, M. W. (2007). *The Total Work of Art: From Bayreuth to Cyberspace*. London, UK: Routledge.
- Soto-Faraco, S., Kingstone, A., and Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia* 41, 1847–1862. doi: 10.1016/S0028-3932(03)00185-4
- Sourav, S., Röder, B., Ambsdorf, F., Melissari, A., Arvaniti, M., and Vatakis, A. (2025). Five experiments, including a pre-registered replication, and mini

meta-analyses find no evidence that sound-shape associations modulate audiovisual temporal order judgements. *J. Exp. Psychol.: General* 154, 522–532. doi: 10.1037/xge0001641

South, W. (2001). *Color, Myth, and Music: Stanton Macdonald-Wright and Synchronism*. Raleigh: North Carolina Museum of Art.

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attent. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7

Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *J. Consumer Psychol.* 22, 37–54. doi: 10.1016/j.jcps.2011.09.004

Spence, C. (2015). “Cross-modal perceptual organization,” in *The Oxford Handbook of Perceptual Organization*, ed. J. Wagemans (Oxford, UK: Oxford University Press), 649–664.

Spence, C. (2018). “Multisensory perception,” in *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, eds. J. Wixted and J. Serences (Hoboken, NJ: John Wiley and Sons), 1–56.

Spence, C. (2019). On the relative nature of (pitch-based) crossmodal correspondences. *Multisens. Res.* 32, 235–265. doi: 10.1163/22134808-20191407

Spence, C. (2020a). Assessing the role of emotional mediation in explaining crossmodal correspondences involving musical stimuli. *Multisens. Res.* 33, 1–29. doi: 10.1163/22134808-20191469

Spence, C. (2020b). Olfactory-colour crossmodal correspondences in art, science, and design. *Cognit. Res.: Princ. Implicat.* 5:52. doi: 10.1186/s41235-020-00246-1

Spence, C. (2021). *Sensehacking: How to use the Power of Your Senses for Happier, Healthier Living*. London, UK: Viking Penguin.

Spence, C., and Deroy, O. (2013). “Crossmodal mental imagery,” in *Multisensory Imagery: Theory and Applications*, eds. S. Lacey and R. Lawson (New York, NY: Springer), 157–183.

Spence, C., and Di Stefano, N. (2022a). Coloured hearing, colour music, colour organs, and the search for perceptually meaningful correspondences between colour and pitch. *i-Perception*. 13, 1–42. doi: 10.1177/20416695221092802

Spence, C., and Di Stefano, N. (2022b). Crossmodal harmony: Looking for the meaning of harmony beyond hearing. *i-Perception*. 13, 1–40. doi: 10.1177/20416695211073817

Spence, C., and Di Stefano, N. (2024a). Sensory translation between audition and vision. *Psychon. Bull. Rev.* 31, 599–626. doi: 10.3758/s13423-023-02343-w

Spence, C., and Di Stefano, N. (2024b). What, if anything, can be considered an amodal sensory dimension? *Psychon. Bull. Rev.* 31, 1915–1933. doi: 10.3758/s13423-023-02447-3

Spence, C., and Di Stefano, N. (2025). “Gestalt perceptual grouping and crossmodal art,” in *Handbook of Gestalt-Theoretic Psychology of Art*, ed. W. Coppola (London: Routledge), 202–230.

Spence, C., and Levitan, C. A. (2021). Explaining crossmodal correspondences between colours and tastes. *i-Perception*. 12, 1–28. doi: 10.1177/20416695211018223

Spence, C., Sanabria, D., and Soto-Faraco, S. (2007). “Intersensory Gestalten and crossmodal scene perception,” in *Psychology of Beauty and Kansei: New Horizons of Gestalt Perception*, ed. K. Noguchi (Tokyo: Fuzanbo International), 519–579.

Spence, C., and Sathian, K. (2020). “Audiovisual crossmodal correspondences: Behavioural consequences and neural underpinnings,” in *Multisensory Perception: From Laboratory to Clinic*, eds. K. Sathian and V. S. Ramachandran (San Diego, CA: Elsevier), 239–258.

Stechow, W. (1953). Problems of structure in some relations between the visual arts and music. *J. Aesthet. Art Critic.* 11, 324–333. doi: 10.1111/1540_6245.jaac11.4.0324

Steffens, J. (2020). The influence of film music on moral judgments of movie scenes and felt emotions. *Psychol. Music* 48, 3–17. doi: 10.1177/0305735618779443

Stein, B. E., Burr, D., Constantinides, C., Laurienti, P. J., Meredith, A. M., Perrault, T. J. Jr, Ramachandran, R., et al. (2010). Semantic confusion regarding the development of multisensory integration: a practical solution. *Eur. J. Neurosci.* 31, 1713–1720. doi: 10.1111/j.1460-9568.2010.07206.x

Stein, B. E., London, N., Wilkinson, L. K., and Price, D. P. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J. Cogn. Neurosci.* 8, 497–506. doi: 10.1162/jocn.1996.8.6.497

Sullivan, J. W. N. (1914). An organ on which color compositions are played. The new art of color music and its mechanism. *Scient. Am.* 110, 163:170. doi: 10.1038/scientificamerican02211914-163

Sun, X., Li, X., Ji, L., Han, F., Wang, H., Liu, Y., et al. (2018). An extended research of crossmodal correspondence between color and sound in psychology and cognitive ergonomics. *PeerJ* 6:e4443. doi: 10.7717/peerj.4443

Svartdal, F., and Iversen, T. (1989). Consistency in synesthetic experience to vowels and consonants: five case studies. *Scand. J. Psychol.* 30, 220–227. doi: 10.1111/j.1467-9450.1989.tb01084.x

Takeshima, Y., and Gyoba, J. (2013). Changing pitch of sounds alters perceived visual motion trajectory. *Multisens. Res.* 26, 317–332. doi: 10.1163/22134808-00002422

Tan, S. L., Cohen, A. J., Lipscomb, S. D., and Kendall, R. A. (2013). *The Psychology of Music in Multimedia*. Oxford, UK: Oxford University Press.

Tan, S. L., Pfordresher, P., and Harré, R. (2017). *Psychology of Music: From Sound to Significance* (London: Routledge).

Tan, S. L., Spackman, M. P., and Bezdek, M. A. (2007). Viewers' interpretations of film characters' emotions: effects of presenting film music before or after a character is shown. *Music Percept.* 25, 135–152. doi: 10.1525/mp.2007.25.2.135

Thayer, J. F., and Levenson, R. W. (1983). Effects of music on psychophysiological responses to a stressful film. *Psychomusicology* 3, 44–52. doi: 10.1037/h0094256

The Getty Research Institute (2013). *Philipp Otto Runge's Times of Day*. Available online at: http://www.getty.edu/research/special_collections/notable/runge.html (Accessed June 15, 2025).

Timmers, R. (2022). “Cross-modality and embodiment of tempo and timing,” in *The Oxford Handbook of Time in Music*, eds. M. Doffman, E. Payne, and T. Young (Oxford, UK: Oxford University Press), 215–234.

Triarhou, L. C. (2016). Neuromusicology or musiconeurology? ‘Omni-art’ in Alexander Scriabin as a fount of ideas. *Front. Psychol.* 7:364. doi: 10.3389/fpsyg.2016.00364

Truckenbrod, J. (1992). Integrated creativity: Transcending the boundaries of visual art, music and literature. *Leonardo Music J.* 2, 89–95. doi: 10.2307/1513214

Vanechkina, I. (1968). “A résumé of inquiry on consistencies in ‘colored hearing’ among members of composers’ association of the USSR,” in *Papers of VI All-Union Acoustic Conference* (Moscow: Academy of Sciences of the U.S.S.R.).

Vanechkina, I. (1973). “Soviet musicians and light-music,” in *The Art of Luminous Sounds* (Kazan, Russia: KAI), 89–110.

Veldhorst, N. (2018). *Van Gogh and Music: A Symphony of Blue and Yellow*. New Haven: Yale University Press.

Vergo, P. (2012). *The Music of Painting: Music, Modernism, and the Visual Arts from the Romantics to John Cage*. London, UK: Phaidon.

Vuoskowi, J., Spence, C., Thompson, M., and Clarke, E. (2014). Cross-modal interactions in the perception of expressivity in musical performance. *Attent. Percept. Psychophys.* 76, 591–604. doi: 10.3758/s13414-013-0582-2

Vuoskowi, J., Spence, C., Thompson, M., and Clarke, E. (2016). Interaction of sight and sound in the perception and experience of musical performance. *Music Percept.* 33, 457–471. doi: 10.1525/mp.2016.33.4.457

Wagner, S., Winner, E., Cicchetti, D., and Gardner, H. (1981). “Metaphorical” mapping in human infants. *Child Dev.* 52, 728–731. doi: 10.2307/1129200

Walker, E. (1927). *Das musikalische Erlebnis und seine Entwicklung*. Gottingen, Germany: Vandenhoeck and Ruprecht.

Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: visual angularity is hard, high-pitched, and bright. *Attent. Percept. Psychophys.* 74, 1792–1809. doi: 10.3758/s13414-012-0341-9

Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Percept. Psychophys.* 42, 491–502. doi: 10.3758/BF03209757

Walker-Andrews, A. (1994). “Taxonomy for intermodal relations,” in *The Development of Intersensory Perception: Comparative Perspectives*, eds. D. J. Lewkowicz and R. Lickliter (Hillsdale, NJ: Lawrence Erlbaum), 39–56.

Wallmark, Z. (2019). Semantic crosstalk in timbre perception. *Music Sci.* 2, 1–18. doi: 10.1177/2059204319846617

Wallmark, Z., and Allen, S. E. (2020). Preschoolers' crossmodal mappings of timbre. *Attent. Percept. Psychophys.* 82, 2230–2236. doi: 10.3758/s13414-020-02015-0

Wallmark, Z., Nghiem, L., and Marks, L. E. (2021). Does timbre modulate visual perception? Exploring crossmodal interactions. *Music Perception* 39, 1–20. doi: 10.1525/mp.2021.39.1.1

Wanke, R., Ansani, A., Di Stefano, N., and Spence, C. (2025). Exploring auditory morphodynamics: audio-visual associations in sound-based music. *i-Perception*. 16, 1–21. doi: 10.1177/20416695251338718

Watanabe, K., and Shimojo, S. (2001). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychol. Sci.* 12, 109–116. doi: 10.1111/1467-9280.00319

Waterworth, J. A. (1997). Creativity and sensation: The case for synaesthetic media. *Leonardo* 30, 327–330. doi: 10.2307/1576481

Watkins, J. (2018). “Creating affective visual music,” in *EVA '18: Proceedings of the Conference on Electronic Visualisation and the Arts*, 374–381. doi: 10.14236/ewic/EVA2018.70

Wehner, W. L. (1966). The relation between six paintings by Paul Klee and selected musical compositions. *J. Res. Music Educ.* 14, 220–224. doi: 10.2307/3344055

Welch, R. B., DuttonHurt, L. D., and Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Percept. Psychophys.* 39, 294–300. doi: 10.3758/BF03204939

- Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* 88, 638–667. doi: 10.1037/0033-2909.88.3.638
- Welch, R. B., and Warren, D. H. (1986). “Intersensory interactions,” in *Handbook of Perception and Performance: Vol. 1. Sensory Processes and Perception*, eds. K. R. Boff, L. Kaufman, and J. P. Thomas (New York, NY: Wiley).
- Wells, A. (1980). Music and visual color: a proposed correlation. *Leonardo* 13, 101–107. doi: 10.2307/1577978
- Whiteford, K. L., Schloss, K. B., Helwig, N. E., and Palmer, S. E. (2018). Color, music, and emotion: Bach to the blues. *i-Perception*. 9, 1–27. doi: 10.1177/2041669518808535
- Whitelaw, M. (2008). Synesthesia and cross-modality in contemporary audiovisuals. *Senses Soc.* 3, 259–276. doi: 10.2752/174589308X331314
- Whitney, J. (1980). *Digital Harmony: On the Complementarity of Music and Visual Art*. Peterborough, NH: Byte Books.
- Wicker, F. W. (1968). Mapping the intersensory regions of perceptual space. *Am. J. Psychol.* 81, 178–188. doi: 10.2307/1421262
- Witztum, E., and Lerner, V. (2016). Alexander Nicolaevich Scriabin (1872–1915): Enlightenment or illness? *J. Med. Biogr.* 24, 331–338. doi: 10.1177/0967772014537151
- Wrembel, M. (2009). On hearing colours—cross-modal associations in vowel perception in a non-synaesthetic population. *Poznań Studi. Contemp. Linguist.* 45, 595–612. doi: 10.2478/v10010-009-0028-0
- Young, D. (2005). The smell of greenness: Cultural synaesthesia in the Western desert. *Etnofoor* 18, 61–77. Available online at: <https://www.jstor.org/stable/25758086>
- Zhou, L., Jiang, C., Delogu, F., and Yang, Y. (2014). Spatial conceptual associations between music and pictures as revealed by N400 effect. *Psychophysiology* 51, 520–528. doi: 10.1111/psyp.12195
- Zika, F. (2013). “Color and sound: Transcending the limits of the senses,” in *Light, Image, Imagination*, eds. M. Blassnigg, G. Deutsch, and H. Schimek (Amsterdam: Amsterdam University Press), 29–46.
- Zika, F. (2018). “Colour and sound: transcending the limits of the senses,” in *Senses and Sensation: Critical and Primary Sources, Vol. 2: History and Sociology*, ed. D. Howes (Abingdon: Routledge), 303–316.
- Zilczer, J. (1987). “Color music”: Synaesthesia and nineteenth-century sources for abstract art. *Artibus et Historiae* 8, 101–126. doi: 10.2307/1483303
- Zilczer, J. (2016). “American Rhapsody: From Modern to Postmodern in visual music,” in *The Oxford Handbook of Sound and Image in Western Art*, ed. T. Kaduri (Oxford, UK: Oxford University Press), 17–35.