Article

PERCEPTION

Perception
2022, Vol. 51(12) 871–888

DOI: 10.1177/03010066221126452
journals.sagepub.com/home/pec
SAGE

Cross-modal associations between paintings and sounds: Effects of embodiment

Marina Iosifyan
University of St Andrews, School of Divinity, UK

Anton Sidoroff-Dorso
Moscow State University of Psychology and Education, Russian Federation

Judith Wolfe
University of St Andrews, School of Divinity, UK

Abstract
The present study investigated cross-modal associations between a series of paintings and sounds. We studied the effects of sound congruency (congruent vs. non-congruent sounds) and embodiment (embodied vs. synthetic sounds) on the evaluation of abstract and figurative paintings. Participants evaluated figurative and abstract paintings paired with congruent and non-congruent embodied and synthetic sounds. They also evaluated the perceived meaningfulness of the paintings, aesthetic value and immersive experience of the paintings. Embodied sounds (sounds associated with bodily sensations, bodily movements and touch) were more strongly associated with figurative paintings, while synthetic sounds (non-embodied sounds) were more strongly associated with abstract paintings. Sound congruency increased the perceived meaningfulness, immersive experience and aesthetic value of paintings. Sound embodiment increased immersive experience of paintings.

Keywords
cross-modal associations, embodied, aesthetics, art, sound

Date Received: 20 December 2021; accepted: 30 August 2022

Corresponding author:
Marina Iosifyan, School of Divinity, University of St Andrews, St Mary's College, St Andrews, KY16 9UB, UK.
Email: marina.iosifyan@gmail.com

Everyday life perception is multisensory: we perceive the world as a unity of sounds, movements, colours, odours, shapes and textures, although we receive information about it via separate senses. Although research gained insights as to how the mind integrates information from different senses into a coherent perception, much remains unknown (Stein et al., 2020). For instance, the role of embodiment in multisensory integration (Tsakiris, 2017). One way of investigating how different modalities interact is understanding cross-modal associations. Cross-modal associations are associations between stimuli from different modalities, for example, colour and sound. Cross-modal associations do not imply the experience of a concurrent stimulus (like synaesthesia) but associating sensations with each other. Contrary to synaesthesia, which is a rare condition present in only 4.4% of the population (Simner et al., 2006), cross-modal associations are much more widespread. Some neuroscientists consider them as essential and natural part of regular cognition as human mind is intrinsically multisensory (Pascual-Leone & Hamilton, 2001). Researchers have found cross-modal associations between odours and music (Levitan et al., 2015), colour and music (Palmer et al., 2013), taste and music (Wang & Spence, 2015), touch and film (Iosifian, 2020; Iosifyan et al., 2017), and music and paintings (Albertazzi et al., 2015). All these associations were found in the general population, among individuals who do not have synaesthesia.

Researchers continue to delineate associations between sensations from different modalities (Giannos et al., 2021; Mok et al., 2019; Motoki et al., 2020; Saluja & Stevenson, 2018). They also attempt to understand the mechanisms which underly these associations. Understanding these mechanisms may contribute to our understanding of multisensory integration, which is one of the key characteristics of perception. There are different mechanisms which underly cross-modal associations. They involve emotional, structural, statistical, and semantic correspondences (Spence, 2011). Emotional correspondences involve two sensations which are associated because they both are associated with a similar emotion. For example, a yellow colour and a major music tone are associated with happiness. Structural cross-modal correspondences are mediated by peculiarities of the neural systems which code sensory information. For example, loudness and brightness are related to increased stimulus intensity, so light and bright stimuli are associated. Statistical cross-modal correspondences involve two sensations which are associated because they are often correlated in the natural environment. For instance, smaller objects often create sounds of higher pitch, while larger objects – of low pitch, thus high-pitched sounds are associated with smaller shapes and low-pitched sounds – with larger shapes. Semantical cross-modal correspondences involve two sensations which are semantically congruent and thus associated with each other. For example, pitch and elevation are associated because our language uses the words 'low' and 'how' both to describe pitch and elevation. In the present study, we are interested in the role of embodied cross-modal correspondences: correspondences which are based on sensorimotor associations (Pezzulo et al., 2011; Spence & Deroy, 2012).

The role of embodied or sensorimotor associations might contribute to our understanding of cross-modal associations, because embodied representations may explain why certain cross-modal associations develop. For instance, mirror-touch synaesthesia is discussed in the literature: people with this synaesthesia experience tactile sensations when they observe someone else being touched (Blakemore et al., 2005; Holle et al., 2013). Some researchers assume that the underlying mechanism of this synaesthesia is based on mirror-touch neurons: mirror neurons which fire both when humans observe others being touched and when they are being touched themselves (Linkovski et al., 2017; Ward & Banissy, 2015). Although some researchers claim that there is no evidence to link cross-modal associations with the neurological phenomena of synaesthesia (Deroy & Spence, 2013; Parise & Spence, 2013), others assume that the two are related (Marks, 2017; Martino & Marks, 2001; Ludwig et al., 2011). Moreover, some researchers assume that mirror-touch synaesthesia is an extreme endpoint of a normal mirror neuron system, which is also present in people without synaesthesia (Linkovski et al., 2017). In non-synaesthetes, observing others being touched activates certain brain regions which are also activated during actual touch (secondary somatosensory cortex, Ebisch et al., 2008; Keysers et al., 2004; Keysers et al., 2010). It is assumed that the purpose of mirror neurons is to translate sensory information into motor actions (Rizzolatti & Fogassi, 2014). That is, a mirror neuron mechanism can partly explain how embodied representations underly cross-modal associations: two sensations from different modalities can be associated because they are related to similar embodied representations.

Some researchers assume that cross-modal associations between words and shapes can be explained by embodied representations. Certain words are associated with angular shapes because of the audio-motor associations between listening and performing articulatory movements: the word 'maluma' is associated with rounded shapes, because the mouth is rounded when we pronounce this word (Galantucci et al., 2006; Sapir, 1929; Spence & Deroy, 2012).

Another study showed that speech sounds affect the way people perceive mouth-like shapes. Participants listened to a speech sound ('wee' or a 'woo' sound) or an environmental sound (e.g., ice cracking) and after that they very briefly saw an oval which varied in aspect ratio (Sweeny et al., 2012). When participants were asked to choose the oval they briefly saw from a group of 10 different ovals, their choice was affected by the sound they heard. The oval was evaluated as taller after the 'wee' sounds and as wider after the 'woo' sounds. It was assumed that while listening to the speech sounds, individuals automatically perform articulatory movements similar to these sounds: performing 'wee' sound movements involves a narrower mouth shape, hence the oval was evaluated as taller; performing 'woo' sounds movements involves a wider mouth shape, hence the oval was evaluated as wider (Sweeny et al., 2012).

The sound symbolism approach assumes that certain cross-modal correspondences (e.g., the mil-mal phenomenon[1]) can be explained by the movements of the mouth when pronouncing certain sounds, rather than by other factors (such as statistical correspondences in the environment: e.g., larger objects make lower-pitched sounds; Sapir, 1929). Thus, audiomotor (sound and movement) and not only audiovisual (sound and shape) correspondences play an important role in these associations (Galantucci et al., 2006). This does not mean that other mechanisms of cross-modal associations do not play a role in them (e.g., emotion, semantic evaluation, statistical correspondences). However, the role of embodied representations in these associations is not yet well reported in the literature.

To fill this gap, in the present study we investigate the role of embodiment in associations between sounds and complex visual stimuli (paintings). More specifically, we will investigate the role of embodiment in perception of abstract and figurative paintings. Earlier research in neuroaesthetics has found that abstract and figurative paintings are approached differently and elicit different mental processes. Figurative paintings elicit local gaze scan paths and activate category-specific brain areas (e.g., tuned for faces and objects). Abstract paintings elicit global gaze scan paths and activate brain areas which involve perception of shape and colour (features of intermediate complexity; Epstein & Kanwisher, 1998; Kawabata & Zeki, 2004; Yago & Ishai, 2006; Zangemeister et al., 1995). Thus, embodied representations can play a different role in cross-modal associations with abstract and figurative paintings.

Previous research investigated the role of embodiment in the perception of visual art, but not cross-modal associations. Researchers observed mu-rhythm suppression (which indicates mirror neuron activation) in response to abstract paintings, as opposed to graphically modified versions of these paintings (Umiltà et al., 2012). They suggested that this activation reflects the imitation of movements made to create the artwork (knowing the object [e.g., a painting] is equivalent to knowing the process by which it has been created). The same effects were found in film perception: researchers compared brain activity while their participants watched the same scene shot by a fixed camera and by a moving, zoomed camera (Heimann et al., 2014). They found that motor regions were more intensely activated when the camera was moving. Artworks can, it would seem, be perceived in an embodied manner. This can be related to embodied representations associated with the process of art-making and representations associated with the content of an artwork.

In the present study, we investigate cross-modal associations between embodied and synthetic sounds and figurative and abstract paintings. Embodied sounds in this study are understood as sounds associated with bodily sensations, bodily movements, and touch. Embodied nature of sound is related to the fact that sounds accompany actions and are processed together (Vainio & Vainio, 2021). Sound characteristics (higher-lower pitch) affect hand movements (upper-lower hand locations; Salgado-Montejo et al., 2016). Earlier research found neurons which are sensitive to both sounds and actions (Keysers et al., 2003; Kohler et al., 2002). Thus, in the present study we use the sounds which accompany footsteps, coughing, touching pebbles, etc. Synthetic sounds are sounds which, to the contrary, are not associated with these sensations. Previous research demonstrated that abstract and figurative art elicit different mental processes (Aviv, 2014). Construal level theory (CLT), earlier applied to understand differences in the processing of figurative and abstract art, suggests that there are differences in abstract and concrete mindsets (Durkin et al., 2020): namely, they vary in the cognitive availability of abstract or concrete representations. Embodied sounds are related to specific actions, objects, or events (e.g., the sound of breathing, the sound of tearing paper), while synthetic sounds cannot be easily attributed to concrete events. We expected that embodied sounds would be more strongly associated with figurative paintings, because they involve object representation, and that synthetic sounds would be more strongly associated with abstract paintings.

Previous research investigated links between visual and auditory stimuli. Interestingly, the idea that certain shapes are associated with particular sounds was developed by artists, for example, Vasiliy Kandinsky (Kandinsky, 1926/1994). Cross-modal associations between complex auditory stimuli (music) and paintings have been empirically investigated. For instance, it was found that hue, lightness, timbre, and musical tempo are among the key characteristics which underly associations between music and paintings (Albertazzi et al., 2015). However, the role of embodied representations in cross-modal associations between sounds and paintings was not investigated.

Sounds can be associated with images, but they can also affect the way we perceive these images. That is, cross-modal associations between embodied/synthetic sounds and paintings might affect the perception of paintings. In the Sweeny et al. (2012) study, described above, speech sounds influenced the perception of ovals. Research on cross-modal correspondences in general revealed several perceptual effects: high/low pitched sounds affect the speed and accuracy of brightness and shape perception (Marks, 1987; Odgaard et al., 2003; Stein et al., 1996), descending and ascending pitch affects the perception of ambiguous visual motion (Maeda et al., 2004), sounds of a male/female voice affect the perception of androgynous faces (Smith et al., 2007), visual cues facilitate identification of auditory tests and vice versa (Schneider et al., 2008; Vallet et al., 2010). Moreover, cross-modal effects are also related to perceived value of stimuli: congruent sounds intensify hedonic evaluations of food (Crisinel et al., 2012), and congruent music intensifies the aesthetic appreciation of paintings (Limbert & Polzella, 1998). For example, jazz music increases liking of abstract paintings, while Classical music – liking of figurative paintings (Actis-Grosso et al., 2016; Koning & van Lier, 2013).

To further investigate hedonic and perceptual effects of sounds on paintings, we will measure how sounds affect aesthetic appreciation, perceived meaningfulness and immersive experience of paintings. Although the broad concept of aesthetic value includes various evaluative responses one may have to art (e.g., emotional reaction, absorption, being moved, in Leder and Nadal's model of aesthetic appreciation and aesthetic judgments; Leder & Nadal, 2014), in the present study we focus on a narrower definition of aesthetic value as aesthetic evaluation of an artwork. In line with the congruency effects described above, it is possible that congruent sounds (both embodied and synthetic) might increase aesthetic appreciation of paintings. These effects of congruency can be related to the fluency theory of aesthetic pleasure (Forster et al., 2013; Reber et al., 2004). That is, congruent sound may increase processing of the painting as more fluent. Fluency, in its turn, may increase positive affect and aesthetic evaluation of the painting. Moreover, the effects of embodiment may partly underly congruency between sounds and paintings. In relation to embodiment effects on the perception of art, it was found that sensory-motor engagement impacts aesthetic appreciation of an artwork (Ardizzi et al., 2020). In one study, in the experimental condition, participants were asked to contract muscles while looking at a figurative painting with a painful expression. In the control condition, participants were asked to refrain from making any facial movements. Participants in the experimental group enjoyed the artworks more than did participants in the control group (Ardizzi et al., 2020). That is, embodiment influences the hedonic effects of figurative paintings. It is thus possible that congruent embodied sounds might increase aesthetic appreciation (perceived aesthetic value) of figurative paintings. Moreover, previous research showed that effects of embodiment are involved in perception of abstract paintings as well: cortical motor system is involved in observation of abstract paintings (Umiltà et al., 2012). Thus, we might expect that congruent embodied sounds also increase aesthetic appreciation of abstract paintings. The effect of synthetic sounds on aesthetic value of figurative versus abstract paintings is more unclear, so this aspect of the study remains exploratory.

Based on the previous research, congruent visual/auditory stimuli facilitate identification of auditory/visual tests (Schneider et al., 2008; Vallet et al., 2010). Exposed to a congruent sound (e.g., a sheep sound), people identify a visually depicted object (e.g., a sheep) faster and make less errors (Schneider et al., 2008). It is thus possible that congruent sounds might increase perceived meaningfulness of paintings. Moreover, sound embodiment may play a role in cross-modal effects on perceived meaningfulness of paintings. In line with Freedberg and Gallese embodied simulation theory, when perceiving figurative paintings depicting acting bodies, viewers simulate their actions, emotions, and sensations (Freedberg & Gallese, 2007). Moreover, even perception of abstract paintings involves this corporeal dimension: observing abstract paintings viewers can simulate the artist's creative gestures (e.g., dynamic brush strokes; Freedberg & Gallese, 2007). Thus, we might expect that embodied sounds increase perceived meaningfulness of both figurative and abstract paintings. For figurative paintings, congruent embodied sounds might facilitate the understanding of the depicted scenes in these paintings. For abstract paintings, congruent embodied sounds might facilitate the understanding of the artist's action that produced these paintings.

Finally, together with aesthetic appreciation and perceived meaningfulness, we will investigate immersive experience of paintings. Immersive experiences in art were mainly investigated in fiction studies and are described as a feeling of 'getting lost in a book' (Nell, 1988). In visual aesthetics, immersive experience allows the viewer to engage in transportation experience, be fully involved in a story or an image (Burke, 2015; Di Dio et al., 2016). Since perception in real environment is multisensory, being able to hear a sound congruent with a painting may increase the feeling of immersion. Earlier studies showed that perception in one modality (haptic feedback) increases the feeling of immersion in virtual reality (Fan et al., 2013). Sound embodiment may play a role in cross-modal effects on immersive experience of paintings. Camera movements and audio effects imitating experiences in everyday life (e.g., zooming imitates visual experience of walking towards the scene in the real environment) increase sensory immersion in cinema (Gallese & Guerra, 2020). It is possible that embodied sounds (e.g., imitating the sound of walking on a wooden floor) increase immersive experience of paintings.

In this study, we presented figurative and abstract paintings paired with embodied and synthetic sounds. We measured the strength of associations between paintings and sounds, as well as perceived meaningfulness, perceived aesthetic value, and immersive experience of paintings. Based on the literature reviewed above, we tested the following hypotheses:

1. Embodied sounds would be more strongly associated with figurative paintings than with abstract paintings; synthetic sounds would be more strongly associated with abstract paintings than with figurative paintings.
2. Congruent sounds would increase perceived meaningfulness, aesthetic value and immersive experience of paintings more compared to non-congruent sounds.
3. Embodied sounds would increase perceived meaningfulness, aesthetic value and immersive experience of figurative and abstract paintings more than synthetic sounds.

Method

Materials

Sounds. We collected 80 sounds, including 40 embodied sounds and 40 synthetic sounds. Embodied sounds included sounds associated with bodily movements (e.g., the sound of steps on a wooden floor), bodily sensations (e.g., the sound of human drinking), and touch (e.g., the sound of human fingertips running over pebbles). Synthetic sounds included electronic, artificially created sounds (e.g., electronic music sound samples) and NASA space sounds. We expected that these sounds would not be associated with bodily movements, sensations and touch. The sounds were collected on YouTube. The full list of embodied and synthetic sounds used in the main study, as well as the recordings of sounds, can be found in the Supplemental Materials. All sounds frequency was set to 44,100 Hz.

To confirm that embodied sounds we preselected are indeed associated with bodily movements, bodily sensations, and touch, contrary to synthetic sounds, we conducted a pre-test. We also aimed to select embodied and synthetic sounds which do not differ in liking. Otherwise, the effects of sounds on aesthetic appreciation of paintings could have been attributed to aesthetic appreciation of sounds rather than to their embodied/synthetic nature. We recruited 60 participants on Prolific, a platform for online research, who took part in the pre-test (age range: 19–66 years, $M = 34.13$, $SD = 12.05$, 34 females, 25 males, 1 non-binary). They were asked to evaluate each sound on two items: embodiment ('Is this sound associated with bodily movements, bodily sensations or touch?') and liking ('Do you like this sound?') on a 4-point scale from 1 (not at all) to 4 (very much). Each sound was presented for 20 s before the participant had an opportunity to evaluate it. Participants were not told what sound they were hearing. The time of each trial was not limited. Each participant evaluated 40 sounds (20 embodied and 20 synthetic sounds).

For the main experiment, we selected 31 sounds evaluated as embodied (>2, $M = 2.84$, $SD = 0.53$) and 31 sounds evaluated as synthetic – that is, not associated with bodily movements, bodily sensations and/or touch (<2, $M = 1.34$, $SD = 0.28$). These sounds differed significantly in embodiment evaluation, $t(30) = 14.30$, $p < .0001$, $d = 2.59$. We also ensured that they did not differ in liking. There were no significant differences in liking between embodied sounds ($M = 1.97$, $SD = 0.47$) and synthetic sounds ($M = 2.12$, $SD = 0.62$), $t(30) = -0.99$, $p = .326$, $d = 0.17$.

Paintings. We selected 40 figurative paintings of various artists (e.g., *The garden* by Emile Friant, *Etoile* by Edgar Degas) and 40 abstract paintings (e.g., *Sky above clouds* by Georgia O'Keeffe, *Blue and grey* by Mark Rothko). The paintings were selected by the authors of this manuscript and four experts affiliated with the Institute for Theology, Imagination and the Arts, University of St Andrews. We aimed to select an equal amount of figurative and abstract paintings to be further paired with sounds. All but three figurative paintings involved images of humans (expect for Carl Blechen's *Forest path near Spandau*, Rembrandt's *Ox*, and John Constable's *Arundel castle and gardens*). Only three abstract paintings involved images of humans: Pablo Picasso's *Guernica* and *Weeping woman* and *Femme assise*. All paintings were uploaded from the museum online collections (e.g., Tate Museum online collection) and were not adjusted for brightness. All images varied in dimensions and were adjusted with the following parameters: the average width of images was 556 ($SD = 151$), the average height was 509 ($SD = 82$). Abstract and figurative images did not differ in their dimensions, ps>.408. See Supplemental Materials for the full list of paintings.

Sound-Painting Pairs. For the main study, we aimed to pair each painting with a congruent and a non-congruent embodied and synthetic sound. We selected four sounds and paired them with each painting (see full list of pairs in Supplemental Materials). The sounds consisted of two embodied sounds and two synthetic sounds. We aimed to pair each painting with a congruent embodied sound, a non-congruent embodied sound, a congruent synthetic sound, and a non-congruent synthetic sound. For this reason, we aimed to select only those paintings which are significantly differently associated with two embodied sounds (stronger with one than the other) and with two synthetic sounds.

A second pre-test was conducted to ensure that congruent embodied sounds are associated with paintings significantly stronger than non-congruent embodied sounds and that congruent synthetic sounds are associated with paintings significantly stronger than non-congruent synthetic sounds. One hundred and twenty participants took part in the second pre-test, they were recruited on Prolific (age range: 18–60 years, $M = 27.29$, $SD = 10.93$, 81 females, 27 males, 2 non-binary). They were asked to evaluate 80 painting-sound pairs ('Do this sound and this painting fit together?') on a scale from 1 (No) to 6 (Yes). Each painting was presented once, paired with either an embodied or a synthetic sound.

One-sample ANOVA was conducted for each painting to ensure that four sounds were differently associated with the painting (see Supplemental materials). Independent samples t-tests were conducted further to ensure that two embodied sounds were differently associated with the painting (i.e., one congruent and one non-congruent), as well as that two synthetic sounds were differently associated with the painting (see Supplemental materials). Only paintings with statistically significant differences between painting-colour associations were selected for the main study (see Supplemental Materials). Based on this analysis, 51 paintings (20 abstract and 31 figurative paintings), each paired with four sounds, were included in the main study. Twenty-nine paintings were excluded. Congruent synthetic sounds were associated with paintings stronger compared to non-congruent synthetic sounds, $t(50) = 12.82$, $p < .001$, $d = 1.81$. Congruent embodied sounds were associated with paintings stronger compared to non-congruent embodied sounds, $t(50) = 16.84$, $p < .001$, $d = 2.35$.

Participants

Using G*Power (Faul et al., 2007), the required minimal sample size to detect differences in repeated measures analysis of variance (number of measurements = 8) was 62 participants for a medium effect, a power of .95, and an alpha level of .05. One hundred and nineteen participants (age range: 18–73, $M = 35.68$, $SD = 13.48$, 38 males, 76 females, 3 non-binary, 2 did not report their gender) were recruited on Prolific. We selected participants with high approval ratings (>80%) to ensure high-quality data. None of these participants took part in the pre-tests we conducted.

Procedure

A painting and a sound appeared simultaneously. After 20 s, the response buttons were enabled, and participants were asked to evaluate the painting-sound on a 6-point scale (from 1 – No, to 6 – Yes) for each of four questions:

1. Do this sound and this painting fit together?
2. Do you feel immersed?
3. Is this painting meaningful?
4. Does this painting have aesthetic value?

The painting and the sound were presented until participants answered all four questions. The order of questions was sequential. The time was not limited. Each participant evaluated 51 paintings. Each painting was paired with either a congruent embodied sound, a non-congruent embodied sound, a congruent synthetic sound, or a non-congruent synthetic sound. Each participant was presented with an equal number of trials involving the eight types of response: abstract art + congruent embodied sounds (5 trials), abstract art + non-congruent embodied sounds (5 trials), abstract art + congruent synthetic sounds (5 trials), abstract art + non congruent synthetic sounds (5 trials), figurative art + congruent embodied sounds (8 trials), figurative art + non-congruent embodied sounds (8 trials), figurative art + congruent synthetic sounds (8 trials), figurative art + non congruent synthetic sounds (8 trials).

For the main analysis, we averaged eight types of response for each participant:

1. Figurative paintings paired with congruent embodied sounds
2. Figurative paintings paired with non-congruent embodied sounds
3. Figurative paintings paired with congruent synthetic sounds
4. Figurative paintings paired with non-congruent synthetic sounds
5. Abstract paintings paired with congruent embodied sounds
6. Abstract paintings paired with non-congruent embodied sounds
7. Abstract paintings paired with congruent synthetic sounds
8. Abstract paintings paired with non-congruent synthetic sounds

Results[2]

Associations Between Paintings and Sounds

Means and standard deviations of associations between paintings and sounds were calculated (see Table 1[3]). Repeated measures $2 \times 2 \times 2$ ANOVA (sound congruency, sound embodiment, type of painting) indicated that the main effect of sound congruency on sound-painting associations was significant (see Table 2). Paired samples *t*-tests were conducted to compare associations between paintings and congruent sounds and paintings and non-congruent sounds. As expected, congruent embodied sounds were associated more strongly with abstract paintings than non-congruent embodied sounds were, $t(118) = 21.03$, $p < .001$, $d = 1.92$. Congruent synthetic sounds were associated more strongly with abstract paintings stronger than non-congruent synthetic sounds were, $t(118) = 8.36$, $p < .001$, $d = 0.77$. Congruent embodied sounds were associated more strongly with figurative paintings than non-congruent embodied sounds were, $t(118) = 32.20$, $p < .001$, $d = 2.96$. Finally, congruent synthetic sounds were associated more strongly with figurative paintings than non-congruent synthetic sounds were, $t(118) = 19.74$, $p < .001$, $d = 1.80$ (see Table 1).

Paired samples *t*-tests were conducted to compare the strength of associations between abstract paintings and congruent synthetic and embodied sounds. Embodied sounds were associated more strongly with figurative paintings than with abstract paintings, $t(119) = -8.79$, $p < .001$, $d = 0.81$. Synthetic sounds were associated more strongly with abstract paintings than with figurative paintings, $t(119) = 4.17$, $p < .001$, $d = 0.38$.

Table 1. Descriptive statistics.

		Associations		Meaningfulness		Aesthetic value		Immersive experience	
Painting	Sound	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Abstract	Embodied (congruent)	3.70	(0.97)	2.92	(0.88)	3.37	(0.93)	2.98	(0.92)
Abstract	Embodied (non-congruent)	1.90	(0.71)	2.62	(0.88)	3.29	(1.01)	1.97	(0.72)
Abstract	Synthetic (congruent)	3.88	(1.01)	3.02	(1.03)	3.45	(1.00)	3.39	(1.06)
Abstract	Synthetic (non-congruent)	3.15	(1.01)	2.75	(0.86)	3.25	(0.96)	2.75	(0.90)
Figurative	Embodied (congruent)	4.43	(0.80)	3.99	(0.81)	4.29	(0.94)	3.61	(0.91)
Figurative	Embodied (non-congruent)	1.94	(0.63)	3.69	(0.91)	4.04	(1.04)	2.14	(0.78)
Figurative	Synthetic (congruent)	3.50	(0.99)	4.03	(0.77)	4.25	(0.92)	3.40	(0.96)
Figurative	Synthetic (non-congruent)	2.03	(0.76)	3.74	(0.89)	4.05	(1.05)	2.23	(0.81)

Table 2. Repeated measures ANOVA results.

	Associations	Perceived meaningfulness	Perceived aesthetic value	Immersive experience
Main effects				
Embodiment	$F(1,118) = 6.71$, $p = .011$, $\eta^2 = .054$	$F(1,118) = 7.94$, $p = .006$, $\eta^2 = .063$	$F(1,118) = 0.02$, $p = .886$, $\eta^2 = .000$	$F(1,118) = 32.42$, $p < .001$, $\eta^2 = .215$
Congruency	$F(1,118) = 1013.74$, $p < .001$, $\eta^2 = .896$	$F(1,118) = 53.43$, $p < .001$, $\eta^2 = .312$	$F(1,118) = 19.73$, $p < .001$, $\eta^2 = .143$	$F(1,118) = 420.73$, $p < .001$, $\eta^2 = .781$
Art (figurative vs. abstract)	$F(1,118) = 11.99$, $p = .001$, $\eta^2 = .092$	$F(1,118) = 246.35$, $p < .001$, $\eta^2 = .676$	$F(1,118) = 88.54$, $p < .001$, $\eta^2 = .429$	$F(1,118) = 1.46$, $p = .230$, $\eta^2 = .012$
Interactions				
Embodiment*Congruency	$F(1,118) = 197.62$, $p < .001$, $\eta^2 = .626$	$F(1,118) = 0.25$, $p = .620$, $\eta^2 = .002$	$F(1,118) = .60$, $p = .439$, $\eta^2 = .005$	$F(1,118) = 21.14$, $p < .001$, $\eta^2 = .152$
Embodiment*Art	$F(1,118) = 214.58$, $p < .001$, $\eta^2 = .645$	$F(1,118) = 1.87$, $p = .174$, $\eta^2 = .016$	$F(1,118) = 0.38$, $p = .536$, $\eta^2 = .003$	$F(1,118) = 92.62$, $p < .001$, $\eta^2 = .440$
Congruency*Art	$F(1,118) = 87.80$, $p < .001$, $\eta^2 = .427$	$F(1,118) = 0.08$, $p = .778$, $\eta^2 = .001$	$F(1,118) = 2.16$, $p = .145$, $\eta^2 = .018$	$F(1,118) = 42.55$, $p < .001$, $\eta^2 = .265$
Embodiment*Congruency*Art	$F(1,118) = 0.05$, $p = .824$, $\eta^2 < .001$	$F(1,118) = 0.02$, $p = .890$, $\eta^2 < .001$	$F(1,118) = 2.27$, $p = .134$, $\eta^2 = .019$	$F(1,118) = 0.25$, $p = .619$, $\eta^2 = .002$

Perceived Meaningfulness and Sound Effects

Repeated measures 2 × 2 × 2 ANOVA (sound congruency, sound embodiment, type of painting) indicated that the main effect of sound congruency on the perceived meaningfulness of the paintings was significant (see Table 2). As expected, congruent embodied sounds increased the perceived meaningfulness of abstract paintings more than non-congruent embodied sounds did, $t(118) = 4.73$, $p < .001$, $d = 0.43$. Congruent synthetic sounds increased the perceived meaningfulness of abstract paintings more than non-congruent synthetic sounds did, $t(117) = 3.76$, $p < .001$, $d = 0.35$. Congruent embodied sounds increased the perceived meaningfulness of figurative paintings more than non-congruent embodied sounds did, $t(117) = 5.49$, $p < .001$, $d = 0.51$. Finally, congruent synthetic sounds increased the perceived meaningfulness of figurative paintings more than non-congruent synthetic sounds did, $t(117) = 5.46$, $p < .001$, $d = 0.5$ (see Table 1).

Paired sample *t*-tests were conducted to compare the effects of sound embodiment on the perceived meaningfulness of abstract and figurative paintings. Congruent embodied sounds did not increase the perceived meaningfulness of abstract paintings nor figurative paintings more compared to congruent synthetic sounds, *p*s>.110.

Perceived Aesthetic Value and Sound Effects

Repeated measures 2 × 2 × 2 ANOVA (sound congruency × sound embodiment × type of paintings) indicated that the main effect of sound congruency on perceived aesthetic value of the paintings was significant (see Table 2). As expected, congruent embodied sounds increased the perceived aesthetic value of figurative paintings more than non-congruent embodied sounds did, $t(118) = 4.55$ $p < .001$, $d = 0.25$. Congruent synthetic sounds increased the perceived aesthetic value of figurative paintings more than non-congruent synthetic sounds did, $t(117) = 3.65$, $p < .001$, $d = 0.33$. Congruent synthetic sounds increased the perceived aesthetic value of abstract paintings more than non-congruent synthetic sounds did, $t(117) = 2.72$, $p = .008$, $d = 0.25$. However, congruent embodied sounds did not increase the perceived aesthetic value of abstract paintings more than non-congruent embodied sounds did, $p = .265$ (see Table 1).

Paired sample *t*-tests were conducted to compare the effects of sound embodiment on abstract and figurative paintings. Congruent embodied sounds did not increase the perceived aesthetic value of abstract paintings nor figurative paintings more compared to congruent synthetic sounds, *p*s>.192.

Immersive Experience and Sound Effects

Repeated measures 2 × 2 × 2 ANOVA (sound congruency, sound embodiment and type of painting) indicated that the main effect of sound congruency on immersive experience was significant (see Table 2). As expected, congruent embodied sounds increased the immerstive experience of abstract paintings more than non-congruent embodied sounds did, $t(118) = 11.97$ $p < .001$, $d = 1.10$. Congruent synthetic sounds increased the immersive experience of abstract paintings more than non-congruent synthetic sounds did, $t(118) = 7.27$, $p < .001$, $d = 0.67$. Congruent embodied sounds increased the immersive experience of figurative paintings more than non-congruent synthetic sounds did, $t(118) = 17.58$, $p < .001$, $d = 1.16$. Congruent synthetic sounds increased the immersive experience of figurative paintings more than non-congruent embodied sounds did, $t(118) = 17.19$, $p < .001$, $d = 1.58$ (see Table 1).

Paired sample *t*-tests were conducted to compare the effects of sound embodiment on the immersive experience of abstract and figurative paintings. Embodied sounds increased the immersive experience of figurative paintings more compared to synthetic sounds, $t(118) = 2.52$, $p = .013$,

$d = 0.23$. Synthetic sounds increased the immersive experience of abstract paintings more compared to embodied sounds, $t(118) = -5.34$, $p < .001$, $d = 0.48$ (see Table 1).

Discussion

This study investigated whether the general population has cross-modal associations between embodied and synthetic sounds and figurative and abstract paintings. We found that embodied sounds are associated more strongly with figurative paintings, and synthetic sounds with abstract paintings. To our knowledge, this is the first study which tested the role of sound embodiment in sound-painting associations. It showed that cross-modal associations between visual and auditory stimulation can be partly explained by embodiment effects.

It is possible that associations between embodied sounds and figurative paintings are stronger because these sounds are related to a concrete context or a certain situation (e.g., the sound of nails being hammered is associated with a situation where someone is hammering nails). Synthetic sounds, however, are less concrete (it is harder to link them to a concrete context) and thus are associated more strongly with abstract paintings. Indeed, artists who make abstract art suggest that it is context-invariant, free from association with any context, contrary to figurative art (MacAgy, 1945; Matisse, 1978).

This finding is also in line with the CLT, which bridges art theory and cognitive research (Trope & Liberman, 2010). CLT characterizes differences in abstract and concrete mindsets, which can be applied to perception of abstract and figurative paintings. Low level of abstraction involves representation of object's concrete, contextual features, while high level of abstraction: essential and decontextualized features. In line with this theory, it was found that abstract paintings are associated with distant situations more (temporal and spatial placement in the world), compared to figurative paintings (Durkin et al., 2020). It is possible that the judged distance of synthetic (non-embodied) sounds, associated with abstract paintings, will be longer than judged distance of embodied sounds, associated with figurative paintings. Thus, future research may investigate if sound embodiment affects auditory-distance estimation.

To our knowledge, this study is the first study to establish the role of embodiment in complex cross-modal associations between sounds and paintings. It contributes to our understanding of mechanisms which underly complex cross-modal associations (Spence, 2020b). Previous research investigated other mechanisms which may underly these associations: semantic congruency (Albertazzi et al., 2020), emotion (Di Stefano et al., 2022; Spence, 2020a), stylistic similarities (Hasenfus *et al.*, 1983). These mechanisms can partly account for cross-modal associations we described in this study, because semantically congruent sounds and paintings, similarly valanced sounds, and paintings, sounds and paintings with similar stylistics can be associated with each other. Not denying the role of these possible mechanisms in crossmodal associations, present study extends our knowledge by establishing the role of embodiment.

The second aim of this study was to investigate the effects of sounds on painting perception (perceived meaningfulness, aesthetic value and immersive experience). We found that congruent embodied and synthetic sounds increased the perceived meaningfulness, immersive experience and aesthetic value of both abstract and figurative paintings. This finding confirms that cross-modal associations between sounds and paintings have perceptual and hedonic effects on paintings. Interestingly, perceived meaningfulness and immersive experience may indicate potential mechanisms which underly the effects of congruent embodied and synthetic sounds on aesthetic appreciation of paintings.

Firstly, increased perceived meaningfulness may increase aesthetic appreciation. Some researchers suggest that an artwork is a message from the artist and needs to be understood by the viewers in order to be appreciated (Konečni, 1984). Others suggest that when interpreting the meaning of the artworks, the viewers derive pleasure from it (Berlyne, 1974). Empirical research found that perceived meaningfulness of an artwork is related to its aesthetic value: preference for paintings is positively related to their perceived meaningfulness (Martindale et al., 1990). Moreover, manipulating perceived meaningfulness of a painting (by presenting information along with the painting) increased painting's pleasingness (Russell, 2003). Thus, congruent embodied and synthetic sounds may increase aesthetic appreciation by increasing perceived meaningfulness of paintings.

Secondly, the feeling of immersion is also related to aesthetic pleasure. Researchers assume that immersion in visual art and in literature is related to aesthetic pleasure (Douglas & Hargadon, 2001). Congruent embodied and synthetic sounds may increase aesthetic appreciation of paintings by increasing the feeling of immersion. This effect of embodied and synthetic sounds on paintings perception and appreciation can be taken into account in design.

It should be mentioned that, contrary to our expectations, the effects of embodied sounds on perceived meaningfulness and aesthetic value were not greater compared to the effects of synthetic sounds. This indicates that sound congruency rather than sound embodiment has stronger effect on perceptual and hedonic aspects of painting perception. With one exception: the effects of sound embodiment on immersive experience of paintings. Embodied sounds increased immersive experience of figurative paintings more than synthetic sounds. This finding can be interpreted in line with the embodied simulation theory: embodiment plays an important role in art perception. Observing figurative paintings, individuals simulate what they depict, and figurative paintings in present study often depicted human actors. This effect of embodied sounds may be similar to the one described in Heimann et al. (2014). In this study, the perception of videos zooming in on a scene (thus imitating movements) increased the feeling that the viewer is involved in the scene and the feeling that the viewer is more like the actor. These feelings are related to the sensation of being immersed. It is thus possible that both embodied sounds (vs. synthetic sounds) and video techniques approaching the object (vs. still camera) have similar effects on immersive experience. Interestingly, this effect of embodied sounds was not confirmed for abstract paintings. Synthetic sounds increased immersive experience of abstract paintings more than embodied sounds. This might indicate that sound embodiment is not the only path to increase immersive feeling while observing a painting.

There are important limitations to consider. Firstly, since the study was conducted online and not in the lab, we were unable to control the volume of sounds which were presented to our participants. Of course, all participants were asked to turn their audio on, but we did not control the volume level. We informed our participants that the study involves listening to and evaluating sounds, but we did not account for participants hearing problems. Secondly, the effects of sounds on perceived meaningfulness, aesthetic value, and immersive experience are limited by the fact that our participants' attention was directed to associations between sounds and paintings. Future studies may investigate whether the perceptual effects of cross-modal associations are preserved when participants are not explicitly asked to evaluate how well the sound and the paintings fit each other. Thirdly, earlier researched showed that valence mediates cross-modal associations between music and colour (Whiteford et al., 2018). Since we did not measure the valence of stimuli we used, it is possible that similar valence of sounds and paintings played a role in cross-modal associations we found. Finally, we investigated associations between sounds and paintings: it is not clear whether our findings can be generalized to non-art stimuli. Although cross-modal associations research often involves artworks as stimuli (e.g., music, paintings, cinema; Albertazzi et al., 2015; Iosifian, 2020; Palmer et al., 2013), art is not an everyday stimulus. Future research may test whether embodied sounds will be similarly associated with perception of photographed and blurred images of objects.

Overall, this study detected cross-modal associations between sounds and paintings: embodied sounds were associated with figurative paintings, and synthetic sounds with abstract paintings.

Congruent embodied and synthetic sounds increased the perceived meaningfulness, immersive experience and aesthetic value of paintings. Embodied sounds also increased immersive experience of figurative paintings, while synthetic sounds – of abstract paintings. These findings indicate that embodied representations play a role in cross-modal associations between sounds and paintings.

Acknowledgements

We thank Derek Keefe for his valuable comments on this manuscript.

Data Availability

The data that support the findings of this study are openly available: https://doi.org/10.17605/osf.io/jzak3.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval

All studies described in this manuscript were approved by the School of Psychology & Neuroscience Ethics Committee, University of St Andrews, approval Code: PS15316.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Templeton Religion Trust (grant number TRT0354).

ORCID iD

Marina Iosifyan https://orcid.org/0000-0002-6617-5116

Supplemental Material

Supplemental material for this article is available online.

Notes

1. When people are asked to choose a name for two round tables, the larger one is often called 'mal', while the smaller one is more often called 'mil' (Sapir, 1929).
2. To take into account the trial variance as a repeated measure, we conducted multiple linear analysis (see Supplemental Materials: Supplementary analysis).
3. Analysis was conducted in IBM SPSS Statistics 26 (IMB Corp 2017).

References

Actis-Grosso, R., Daneyko, O., Cattaneo, Z., & Zavagno, D. (2016). What shall we listen to: abstract or figurative music? *Visual Science of Art Conference*. Barcelona. Abstracts, 17.

Albertazzi, L., Canal, L., & Micciolo, R. (2015). Cross-modal associations between materic painting and classical Spanish music. *Frontiers in Psychology*, *6*, 1644–1078. https://doi.org/10.3389/fpsyg.2015.00424

Albertazzi, L., Canal, L., Micciolo, R., & Hachen, I. (2020). Cross-modal perceptual organization in works of art. *i-Perception*, *11*, 2041669520950750. doi:10.1177/2041669520950750

Ardizzi, M., Ferroni, F., Siri, F., Umiltà, M. A., Cotti, A., Calbi, M., Fadda, E., Freedberg, D., & Gallese, V. (2020). Beholders' sensorimotor engagement enhances aesthetic rating of pictorial facial expressions of pain. *Psychological Research*, *84*, 370–379. https://doi.org/10.1007/s00426-018-1067-7

Aviv, V. (2014). What does the brain tell us about abstract art? *Frontiers in Human Neuroscience*, *8*, 85. https://doi.org/10.3389/fnhum.2014.00085

Berlyne, D. E. (Ed.). (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation*. Hemisphere.

Blakemore, S. J., Bristow, D., Bird, G., Frith, C., & Ward, J. (2005). Somatosensory activations during the observation of touch and a case of vision–touch synaesthesia. *Brain*, *128*, 1571–1583. https://doi.org/10.1093/brain/awh500

Burke, M. (2015). The neuroaesthetics of prose fiction: Pitfalls, parameters and prospects. *Frontiers in Human Neuroscience*, *9*, 442. https://doi.org/10.3389/fnhum.2015.00442

Crisinel, A. S., Cosser, S., King, S., Jones, R., Petrie, J., & Spence, C. (2012). A bittersweet symphony: Systematically modulating the taste of food by changing the sonic properties of the soundtrack playing in the background. *Food Quality and Preference*, *24*, 201–204. https://doi.org/10.1016/j.foodqual.2011.08.009

Deroy, O., & Spence, C. (2013). Are we all born synaesthetic? Examining the neonatal synaesthesia hypothesis. *Neuroscience & Biobehavioral Reviews*, *37*, 1240–1253. https://doi.org/10.1016/j.neubiorev.2013.04.001

Di Dio, C., Ardizzi, M., Massaro, D., Di Cesare, G., Gilli, G., Marchetti, A., & Gallese, V. (2016). Human, nature, dynamism: The effects of content and movement perception on brain activations during the aesthetic judgment of representational paintings. *Frontiers in Human Neuroscience*, *9*, 705. https://doi.org/10.3389/fnhum.2015.00705

Di Stefano, N., Murari, M., & Spence, C. (2022). Crossmodal correspondences in art and science: Odours, poetry, and music. In *Olfaction: An interdisciplinary perspective from philosophy to life sciences* (pp. 155–189). Springer.

Douglas, J. Y., & Hargadon, A. (2001). The pleasures of immersion and engagement: Schemas, scripts and the fifth business. *Digital Creativity*, *12*, 153–166. https://doi.org/10.1076/digc.12.3.153.3231

Durkin, C., Hartnett, E., Shohamy, D., & Kandel, E. R. (2020). An objective evaluation of the beholder's response to abstract and figurative art based on construal level theory. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 19809–19815. https://doi.org/10.1073/pnas.2001772117

Ebisch, S. J. H., Perrucci, M. G., Ferretti, A., Del Gratta, C., Romani, G. L., & Gallese, V. (2008). The sense of touch: Embodied simulation in a visuotactile mirroring mechanism for observed animate or inanimate touch. *Journal of Cognitive Neuroscience*, *20*, 1611–1623. https://doi.org/10.1162/jocn.2008.20111

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601. https://doi.org/10.1038/33402

Fan, K., Izumi, H., Sugiura, Y., Minamizawa, K., Wakisaka, S., Inami, M., et al. (2013). "Reality jockey: Lifting the barrier between alternate realities through audio and haptic feedback,". In W. Mackay, S. Brewster, & S. Bødker (Eds.), *The SIGCHI conference* (pp. 2557). ACM SIGCHI.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191. https://doi.org/10.3758/bf03193146

Forster, M., Leder, H., & Ansorge, U. (2013). It felt fluent, and I liked it: Subjective feeling of fluency rather than objective fluency determines liking. *Emotion*, *13*, 280–289. https://doi.org/10.1037/a0030115

Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, *11*, 197–203. doi:10.1016/j.tics.2007.02.003

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*, 361–377. https://doi.org/10.3758/BF03193857

Gallese, V., & Guerra, M. (2020). *The emphatic screen: Cinema and neuroscience*. Oxford University Press.

Giannos, K., Athanasopoulos, G., & Cambouropoulos, E. (2021). Cross-Modal Associations Between Harmonic Dissonance and Visual Roughness. *Music & Science*, 4, 1–15. https://doi.org/10.1177/20592043211055484

Hasenfus, N., Martindale, C., & Birnbaum, D. (1983). Psychological reality of cross–media artistic styles. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(6), 841–863. https://doi.org/10.1037/0096–1523.9.6.841

Heimann, K., Umiltà, M. A., Guerra, M., & Gallese, V. (2014). Moving mirrors: A high density EEG study investigating the effects of camera movements on motor cortex activation during action observation. *Journal of Cognitive Neuroscience*, *26*, 2087–2101. https://doi.org/10.1162/jocn_a_00602

Holle, H., Banissy, M. J., & Ward, J. (2013). Functional and structural brain differences associated with mirror-touch synaesthesia. *Neuroimage*, *83*, 1041–1050. https://doi.org/10.1016/j.neuroimage.2013.07.073

Iosifian, M. (2020). Crossmodal associations between cinema with elements of comic and tragic and texture touch. *Art & Perception*, *8*, 175–187. https://doi.org/10.1163/22134913-bja10004

Iosifyan, M., Korolkova, O., & Vlasov, I. (2017). Emotional and semantic associations between cinematographic aesthetics and haptic perception. *Multisensory Research*, *30*, 783–798. https://doi.org/10.1163/22134808-00002597

Kandinsky, W. (1926/1994). *Point and line to plane* (P. Vergo, trans.). In K. C. Lindsay & P. Vergo (Eds.), *Kandinsky: Complete writings on art* (pp. 524–700). Da Capo Press.

Kawabata, H., & Zeki, S. (2004). Neural correlates of beauty. *Journal of Neurophysiology*, *91*, 1699–1705. https://doi.org/10.1152/jn.00696.2003

Keysers, C., Kaas, J., & Gazzola, V. (2010). Somatosensation in social perception. *Nature Reviews Neuroscience*, *11*, 417–428. https://doi.org/10.1038/nrn2833

Keysers, C., Kohler, E., Umilta, M. A., Nanetti, L., Fogassi, L., & Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research*, *153*, 628–636. doi:10.1007/s00221-003-1603-5

Keysers, C., Wicker, B., Gazzola, V., Anton, J. L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, *42*, 335–346. https://doi.org/10.1016/S0896-6273(04)00156-4

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*, 846–848. doi:10.1126/science.1070311

Konečni, V. J. (1984). Elusive effects of Artists' 'messages'. *Advances in Psychology*, *19*, 71–93. doi:10.1016/S0166-4115(08)62346-8

Koning, A., & van Lier, R. (2013). Effects of music on visual art: An eye movement study. *Perception*, *42*, 103.

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments : The aesthetic episode - developments and challenges in empirical aesthetics. *British Journal of Psychology (London, England : 1953)*, *105*, 443–464. https://doi.org/10.1111/bjop.12084

Levitan, C. A., Charney, S. A., Schloss, K. B., & Palmer, S. E. (2015). The smell of jazz: Crossmodal correspondences between music, odor, and emotion. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 1326–1331.

Limbert, W. M., & Polzella, D. J. (1998). Effects of music on the perception of paintings. *Empirical Studies of the Arts*, *16*, 33–39. https://doi.org/10.2190/V8BL-GBJK-TLFP-R321

Linkovski, O., Katzin, N., & Salti, M. (2017). Mirror neurons and mirror-touch synesthesia. *The Neuroscientist*, *23*, 103–108. https://doi.org/10.1177/1073858416652079

Ludwig, V. U., Adachi, I., & Matzuzawa, T. (2011). Visuo-auditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*, *108*, 20661–20665. https://doi.org/10.1073/pnas.1112605108

MacAgy, D. (1945). Review of the book plastic art and pure plastic art 1937 and other essays, 1941–1943, by P. Mondrian. *The Journal of Aesthetics and Art Criticism*, *4*, 120–121. https://doi.org/10.2307/426088

Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology*, *14*, R990–R991. https://doi.org/10.1016/j.cub.2004.11.018

Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 384–394. https://doi.org/10.1037//0096-1523.13.3.384

Marks, L. E. (2017). Synesthesia, then and now. In O. Deroy (Ed.), *Sensory blending: On synaesthesia and related phenomena* (pp. 13–44). Oxford University Press.

Martindale, C., Moore, K., & Borkum, J. (1990). Aesthetic preference: Anomalous findings for Berlyne's psychobiological theory. *The American Journal of Psychology*, *103*, 53–80. https://doi.org/10.2307/1423259

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science*, *10*, 61–65. https://doi.org/10.1111/1467-8721.00116

Matisse, H. (1978). *Notes d'un peinture*. In J. T. Flam (Ed.), *Matisse on art* (pp. 2–3). Oxford University Press.

Mok, P., Li, G., Li, J. J., Ng, H., & Cheung, H. (2019). Cross-modal association between vowels and colours: A cross-linguistic perspective. *The Journal of the Acoustical Society of America*, *145*, 2265. https://doi.org/10.1121/1.5096632

Motoki, K., Saito, T., Nouchi, R., & Sugiura, M. (2020). Cross-modal correspondences between temperature and taste attributes. *Frontiers in Psychology*, *11*, 571852. https://doi.org/10.3389/fpsyg.2020.571852

Nell, V. (1988). *Lost in a Book: The Psychology of Reading for Pleasure*. Yale University Press.

Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. *Perception & Psychophysics*, *65*, 123–132. https://doi.org/10.3758/BF03194789

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

Parise, C., & Spence, C. (2013). Audiovisual cross-modal correspondences in the general population. In J. Simner & E. M. Hubbard (Eds.), *The Oxford handbook of synesthesia* (pp. 790–815). Oxford University Press.

Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research*, 134, 427–445. https://doi.org/10.1016/s0079-6123(01)34028-1

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2011). The mechanics of embodiment: A dialog on embodiment and computational modelling. *Frontiers in Psychology*, *2*, 5. https://doi.org/10.3389/fpsyg.2011.00005

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*, 364–382. https://doi.org/10.1207/s15327957pspr0804_3

Rizzolatti, G., & Fogassi, L. (2014). The mirror mechanism: Recent findings and perspectives. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 20130420. https://doi.org/10.1098/rstb.2013.0420

Russell, P. A. (2003). Effort after meaning and the hedonic value of paintings. *British Journal of Psychology*, *94*, 99–110. https://doi.org/10.1348/000712603762842138

Salgado-Montejo, A., Marmolejo-Ramos, F., Alvarado, J. A., Arboleda, J. C., Suarez, D. R., & Spence, C. (2016). Drawing sounds: Representing tones and chords spatially. *Experimental Brain Research*, *234*, 3509–3522. doi:10.1007/s00221-016-4747-9

Saluja, S., & Stevenson, R. J. (2018). Cross-Modal associations between real tastes and colors. *Chemical Senses*, *43*, 475–480. https://doi.org/10.1093/chemse/bjy033

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*, 225–239. https://doi.org/10.1037/h0070931

Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Experimental Psychology*, *55*, 121–132. doi:10.1027/1618-3169.55.2.121

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., Scott, K., & Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception*, *35*, 1024–1033. https://doi.org/10.1068/p5469

Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology*, *17*, 1680–1685. https://doi.org/10.1016/j.cub.2007.08.043

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*, 971–995. https://doi.org/10.3758/s13414-010-0073-7

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

Spence, C. (2020b). Simple and complex crossmodal correspondences involving audition. *Acoustical Science and Technology*, *41*, 6–12. doi:10.1250/ast.41.6

Spence, C., & Deroy, O. (2012). Hearing mouth shapes: Sound symbolism and the reverse McGurk effect. *I-Perception*, 3, 550–552. https://doi.org/10.1068/i0558ic

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, *8*, 497–506. https://doi.org/10.1162/jocn.1996.8.6.497

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2020). Multisensory integration and the society for neuroscience: Then and now. *Journal of Neuroscience*, *40*, 3–11. https://doi.org/10.1523/JNEUROSCI.0737-19.2019

Sweeny, T. D., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2012). Sounds exaggerate visual shape. *Cognition*, *124*, 194–200. https://doi.org/10.1016/j.cognition.2012.04.009

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*, 440–463. https://doi.org/10.1037/a0018963

Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others [formula: See text]. *Quarterly Journal of Experimental Psychology (2006)*, *70*, 597–609. https://doi.org/10.1080/17470218.2016.1181768

Umiltà, M. A., Berchio, C., Sestito, M., Freedberg, D., & Gallese, V. (2012). Abstract art and cortical motor activation: An EEG study. *Frontiers in Human Neuroscience*, *6*, 311. https://doi.org/10.3389/fnhum.2012.00311

Vainio, L., & Vainio, M. (2021). Sound-action symbolism. *Frontiers in Psychology*, *12*, 718700. doi:10.3389/fpsyg.2021.718700

Vallet, G., Brunel, L., & Versace, R. (2010). The perceptual nature of the cross-modal priming effect: Arguments in favor of a sensory-based conception of memory. *Experimental Psychology*, *57*, 376–382. https://doi.org/10.1027/1618-3169/a000045

Wang, Q. (Janice), & Spence, C. (2015). Assessing the effect of musical congruency on wine tasting in a live performance setting. *I-Perception*, *6*, 1–13. https://doi.org/10.1177/2041669515593027

Ward, J., & Banissy, M. J. (2015). Explaining mirror-touch synaesthesia. *Cognitive Neuroscience*, *6*, 118–133. https://doi.org/10.1080/17588928.2015.1042444

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

Yago, E., & Ishai, A. (2006). Recognition memory is modulated by visual similarity. *NeuroImage*, *31*, 807–817. https://doi.org/10.1016/j.neuroimage.2005.12.014

Zangemeister, W. H., Sherman, K., & Stark, L. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia*, *33*, 1009–1025. https://doi.org/10.1016/0028-3932(95)00014-t