

INVITED TUTORIAL

Simple and complex crossmodal correspondences involving audition

Charles Spence*

Department of Experimental Psychology, Anna Watts Building, University of Oxford, Oxford, OX2 6GG, United Kingdom

Abstract: The last few years have seen an explosion of interest from researchers in the crossmodal correspondences, defined as the surprising connections that the majority of people share between seemingly-unrelated stimuli presented in different sensory modalities. Intriguingly, many of the crossmodal correspondences that have been documented/studied to date have involved audition as one of the corresponding modalities. In fact, auditory pitch may well be the single most commonly studied dimension in correspondences research thus far. That said, relatively separate literatures have focused on the crossmodal correspondences involving simple versus more complex auditory stimuli. In this review, I summarize the evidence in this area and consider the relative explanatory power of the various different accounts (statistical, structural, semantic, and emotional) that have been put forward to explain the correspondences. The suggestion is made that the relative contributions of the different accounts likely differs in the case of correspondences involving simple versus more complex stimuli (i.e., pure tones vs. short musical excerpts). Furthermore, the consequences of presenting corresponding versus non-corresponding stimuli likely also differ in the two cases. In particular, while crossmodal correspondences may facilitate binding (i.e., multisensory integration) in the case of simple stimuli, the combination of more complex stimuli (such as, for example, musical excerpts and paintings) may instead be processed more fluently when the component stimuli correspond. Finally, attention is drawn to the fact that the existence of a crossmodal correspondence does not in-and-of-itself necessarily imply that a crossmodal influence of one modality on the perception of stimuli in the other will also be observed.

Keywords: Crossmodal correspondences, Simple/complex stimuli, Pitch, Timbre, Sonic seasoning, Auditory, Music

PACS number: 43.66.+y [doi:10.1250/ast.41.6]

1. INTRODUCTION

In recent years, there has been a rapid growth of interest in research on the crossmodal correspondences. According to Spence [1, p. 973], crossmodal correspondences can be defined as “*a compatibility effect between attributes or dimensions of a stimulus (i.e., an object or event) in different sensory modalities (be they redundant or not)*.” Crossmodal correspondences have now been documented between pretty much every pair of sensory modalities (see [2–6] for reviews). Interestingly, however, thus far, there has been far less research/evidence concerning intramodal correspondences, excluding, that is, limited work on Kandinsky’s colour-shape correspondences (see [7] for a review). In the crossmodal arena, it is noteworthy how much of the research that has been published to date has focused specifically on correspondences involving simple auditory stimuli, varying in pitch and to a lesser extent loudness (see [1,3,8] for reviews).

In this tutorial review, the focus is on those correspondences involving audition as one of the component modalities. However, in contrast to previous reviews of audiovisual correspondences involving simple stimuli (e.g., [2–4]), I also want to review the evidence concerning complex crossmodal correspondences (or rather correspondences between, or involving, stimuli that are themselves more complex). I start by briefly reviewing the evidence concerning simple crossmodal correspondences involving basic auditory features such as pitch, loudness, and timbre [9–11]. Thereafter, I move on to look at those correspondences involving more complex auditory stimuli, such as short musical excerpts [12]. Complex here being defined operationally as having multiple individuable elements or attributes. In terms of the putative consequences for perception (and ‘processing fluency’) of presenting corresponding versus non-corresponding sensory stimuli, it has been suggested that simple correspondences between different unisensory features (such as auditory pitch and visual size) may, at least under certain conditions, facilitate crossmodal binding [13]. By contrast, in the case of

*e-mail: charles.spence@psy.ox.ac.uk

combinations involving more complex stimuli, the primary consequence of ‘playing to the correspondences’ would appear to be in terms of a modulation of processing fluency (cf. [14]). In fact, while several of the same explanations (see [1,3] for reviews) can be put forward to account for both simple and complex correspondences, the relative weighting (or contribution) of the different accounts to explaining any particular crossmodal correspondence likely varies in the two cases.

2. CROSSMODAL CORRESPONDENCES

2.1. Correspondences Involving Simple Auditory Stimuli

Perhaps the single most extensively studied auditory attribute, or dimension, in the field of crossmodal correspondences research is pitch (see [8] for a review). Over the years, auditory pitch has been shown to correspond to visual size [10,15], visual angularity [9,16], brightness [16], lightness [16,17], elevation ([8,10,18–21]), and direction of movement [22]. In terms of auditory timbre-colour/shape crossmodal correspondences, Adeli *et al.* [11] concluded that their participants: “*strongly associated soft timbres with blue, green or light gray rounded shapes, harsh timbres with red, yellow or dark gray sharp angular shapes and timbres having elements of softness and harshness together with a mixture of the two previous shapes.*” It is, though, important to stress here, in passing, that no correspondence has been documented between certain pairs of sensory dimensions such as, for instance, auditory pitch and hue [23], or loudness and lightness [16].

In recent years, several researchers have attempted to investigate the automaticity of crossmodal correspondences (e.g., [24–26]). However, the evidence that has been published to date would appear to suggest that the audiovisual correspondences (or at least those that have been studied thus far) fail to meet all of the criteria necessary to classify a particular effect/phenomenon as automatic [27]. In terms of the neural substrates underlying, or associated with coding, the correspondence between auditory and visual stimuli/stimulus dimensions, parietal areas (specifically the intraparietal sulcus) looks likely to be one of the relevant candidate sites ([28,29]; see also [30,31]).

By now, crossmodal correspondences have been demonstrated between simple auditory and visual features in both chimpanzees [32] and very young infants ([33–36]; see also [37]). However, while such results have been taken by some commentators to imply that chimpanzees and neonates are synaesthetic, it should be noted that my colleague Ophelia Deroy and I have argued at length elsewhere against any such suggestion [38]. That said, the last few years has seen the publication of a few intriguing papers showing that individuals with synaesthesia may be

more susceptible to the influences of certain crossmodal correspondences than are non-synaesthetes (e.g., see [39], see also [40]). At the same time, however, it is also worth remembering that crossmodal correspondences are often characterized as being relative [41,42] — that is, it is the larger of two circles that is matched to the lower-pitched of two sounds, rather than there being any specific match between a sound having a particular pitch and an object of a specific size. By contrast, the relation between inducer and concurrent in the case of synaesthesia proper tends to be absolute. This, then, just one of the many fundamental differences between crossmodal correspondences and synaesthesia (see [41] for a fuller discussion of this issue).

Studies using a variety of behavioural tasks have demonstrated that performance on crossmodally congruent audiovisual trials tends to be significantly different from that seen on those trials where crossmodally incongruent stimulus combinations are presented instead (see [1–3,31] for reviews). However, whether performance is better or worse really does seem to depend on the task that participants have to perform. For instance, crossmodally congruent stimulus combinations typically facilitate performance on speeded discrimination ([43]; and see [2] for a review) and Implicit Association Test-type tasks [9,44]. Meanwhile, in terms of the detrimental effects of audiovisual crossmodal correspondences, Parise and Spence [13] reported a series of unspeeded psychophysical studies demonstrating that participants found it significantly harder to discriminate the relative location, or timing, of auditory stimuli (pure tones) when presented together with a corresponding as compared to an incongruent visual stimulus. Elsewhere, presenting crossmodally corresponding audiovisual stimuli has been shown to lead to an enhanced temporal ventriloquism effect [45,46]. And, beyond their effect on perception, crossmodal correspondences have, by now, also been shown to influence both working memory [47] and learning too [48].

Crossmodal correspondences involving simple auditory stimuli are thought to be bidirectional [41]. That is, larger circles are matched with lower-pitched sounds, and lower-pitched sounds are just as strongly associated with larger circles. However, given that we are all visually dominant organisms [49,50], one might reasonably well expect to find that the crossmodal influence of visual over auditory stimuli would be larger than any effects documented in the reverse direction [51]. However, at least for the audiovisual case, the senses appear to be reasonably well matched in this regard. Indeed, over the years, a number of robust crossmodal influences have been documented in both directions, at least between simple auditory and visual stimuli (e.g., see [2] for a review). It should, though, be noted that such bidirectional crossmodal influences have not always been demonstrated [51].

2.2. Correspondences Involving Complex Auditory Stimuli

In terms of crossmodal correspondences involving more complex auditory stimuli, short musical excerpts have often been used as experimental stimuli. So, for instance, in one oft-cited study, Palmer and his colleagues demonstrated that people reliably associated different pieces of classical music with different colour patches [52]. The participants in the first of their studies had to listen to 18 short pieces of classical orchestral music by Bach, Brahms, and Mozart. The musical selections varied in terms of their tempo (slow/medium/fast) and mode (major/minor). For each musical excerpt (18–50 s in duration), the participants had to pick the five best-matching and the five worst-matching colours in order from the 37 carefully-selected Berkeley Colour Project colour patches (shown simultaneously). The results revealed some surprisingly robust crossmodal correspondences between the musical selections and the colours chosen. Intriguingly, this was the case both for Californian students ($N = 48$) as well as for a group of participants from the university of Guadalajara in Mexico ($N = 49$) who took part in the same experiment. The latter finding hinting at the possible cross-cultural generalizability of the audiovisual crossmodal correspondences that were documented.

Palmer *et al.*'s [52] participants also had to rate each and every one of the musical excerpts and colour patches in terms of how strongly associated they were to each of eight emotional descriptors: happy, sad, angry, calm, strong, weak, lively, and dreary. The participants responded using line scales ranging from -100 to 100 . Generally-speaking, the results revealed that faster tempi musical pieces (as well as those musical selections played in major, as compared to minor, mode) were associated with more saturated, lighter, and yellower (i.e., warmer) colours. By contrast, musical selections played in minor mode were associated by the participants with darker, desaturated, bluer colours. Separately, in terms of the composers, Brahms's music (or at least the 6 pieces selected for use in this study) was associated with less saturated, darker, bluer colours than were the musical selections of Bach and Mozart (which did not differ significantly from one another).

Further support for the emotional mediation account of such crossmodal correspondences was provided by the results of two further experiments in which Palmer *et al.*'s [52] participants matched the colour patches (Experiment 2) or the classical music selections (Experiment 3) to one of eight faces displaying different emotional expressions. Once again, robust correlations were obtained. Moreover, the music-colour matches documented in Palmer *et al.*'s Experiment 1 could be predicted on the basis of the music-

emotion and colour-emotion results documented in Experiments 2 and 3. However, that said, one thing that remains unclear on the basis of this, and other studies like it, is whether any emotional mediation is based on the emotion experienced in response to the music, or rather, on a more cognitive assessment of the emotion that one might want to associate with the stimuli themselves [53]. Others, meanwhile, have questioned whether we should really be talking about affect rather than emotion [54].

In subsequent work, 34 musical excerpts sampled from a variety of different genres (including blues, salsa, heavy metal, etc.) were presented to participants [55]. In this case, arousal and valence were found to be key emotional attributes mediating the crossmodal association, or correspondence, between music and colour. There is a potentially intriguing link here to Osgood and colleagues' classic early work on the semantic differential technique ([56]; see also [57–59]). Selecting short musical excerpts from pre-composed music, be it classical or any other genre, is, in some sense, unconstrained/uncontrolled in terms of stimulus generation. In order to address any such criticism, Palmer and his colleagues subsequently went on to show much the same pattern of emotionally-mediated crossmodal matches when their participants matched much more precisely controlled single-line melodies to colour patches instead [60]. However, while oft-cited, it should be stressed that Palmer *et al.*'s [52] study is by no means unique in demonstrating a connection between music and visual stimuli. In fact, over the years, many other studies have also documented an emotional mediation of music-colour correspondences (e.g., see [61–67]).

Here, it is worth remembering that the musical stimuli used in much of the crossmodal correspondences research that has been published to date were presumably professionally composed in order to elicit some particular emotional response in the listener. As such, it is perhaps not so surprising to find that the emotional mediation account should prove to have so much explanatory validity for complex auditory stimuli, at least when compared to its role in explaining the correspondences that have been documented between the presumably much less emotionally-valenced simple visual and auditory stimuli used in much of the research in this area (though see [68–71]). Whatever the explanation, the emotional mediation account of crossmodal correspondences/influences has undoubtedly become an increasingly common theme (or explanatory approach) in the literature in this area in recent years (e.g., see [72–74]).

While Palmer *et al.*'s [52] work involved complex auditory stimuli, the abstract colour patches that constituted the visual stimuli were relatively simple. Elsewhere, researchers have investigated people's matching tendencies when presented with pieces of music and paintings. So, for

instance, Albertazzi and colleagues [75] demonstrated that people tend to associate Spanish music with matric painting. Furthermore, they also found that emotional adjectives such as ‘calm’ and ‘happy’ played a significant role in mediating the crossmodal associations between music and visual art that were observed. Here, it is worth highlighting the separate line of research assessing people’s sensitivity to cross-media artistic styles [76]. So, for example, Hasenfus *et al.* presented people with unfamiliar paintings, architecture, poetry, and music from several different historical epochs/styles (e.g., Baroque, Neoclassical, or Romantic). Across a series of studies, participants had to group the stimuli as they saw fit. Intriguingly, people were significantly more likely to group the stimuli from the same stylistic period together (seemingly regardless of media format) than would have been expected by chance. However, it turns out that such results can sometimes be explained more parsimoniously by the existence of basic sensory/perceptual correspondences rather than necessarily needing to postulate the extraction of common underlying stylistic qualities in the cross-media stimuli used [77,78]. (This presumably being more likely in those cases where the participants are unfamiliar with the complex stimuli that they happen to be rating.) And, once again, emotional mediation may play a significant role in people’s matching behaviour.

In closing this brief summary of audiovisual cross-modal correspondences involving complex auditory stimuli (which nearly always turn out to be excerpts from composed music), it is worth stressing, once again, that just because people experience a correspondence, or some other sort of crossmodal association between music and visual stimuli that does not necessarily mean that there will be an influence on one stimulus on the other, if presented together. This, note, seemingly being the implication of Duthie’s dissertation entitled: “*Do music and art influence one another? Measuring cross-modal similarities in music and art*” [77,78]. However, while by no means obligatory, when a crossmodal influence of music (a complex auditory stimulus) on people’s response to either simple (e.g., colour patches) or complex visual stimuli (e.g., paintings or drama [79]) is documented then several potential explanations may be viable. (Note that exactly the same mechanisms may also explain the effects of visual stimuli, either simple or complex, on people’s response to musical stimuli too [72,80]). On the one hand, there may be a role for ‘sensation transference’ [81,82]. This is the name given to the phenomenon whereby whatever a person happens to think or feel about one stimulus (the music in this case) is transferred to influence their ratings of any other stimuli (visual in this case) that happen to be presented at around the same time (cf. [83]). When one is dealing with complex stimuli, there would seem little likelihood that the sensory

inputs presented in the two modalities are going to be bound into a single multisensory object representation (as is sometimes the case for crossmodal correspondences involving simple stimuli [13]). Rather, when complex stimuli correspond crossmodally, they may simply be processed more fluently than is the case when the component stimuli do not correspond [14,84]. It is this fluency in human information processing that may lead to increased aesthetic appreciation.

3. EXPLAINING CORRESPONDENCES

One popular explanation for the existence of cross-modal correspondences is that they reflect the internalization of the statistical regularities of the environment [1,3]. For instance, when searching for an explanation for the pitch-elevation correspondence, intriguing work by Parise and his colleagues has demonstrated that higher-pitched sounds do, on average, tend to originate from higher in space, perhaps, in part, explaining this particular correspondence [85; see also 86]. Meanwhile, research reported by Ernst [87] has shown that correspondences between unrelated stimulus dimensions in different sensory modalities (i.e., haptic stiffness and visual brightness) can be acquired after only a surprisingly limited amount of experience to the artificially correlated sensory inputs. For half the participants in Ernst’s study, stiffness and brightness were correlated, while for the remainder, the reverse relationship was established.

A second account, first suggested by S. S. Stevens [88], is that certain of the correspondences may have a structural explanation, based on common principles of neural encoding (or representation) [1,3]. Although this account is unlikely to explain many of the correspondences that have been reported to date, it may provide a plausible account for intensity-based correspondences, given similar neural coding of intensity (in terms of increased neural firing) in all senses. In theoretical work on sound(pitch)-colour(hue) matching one also sometimes finds researchers talking of some sort of physical structural relationship between the stimulus dimensions themselves (e.g., [65,89]).

According to a third account, there might be a semantic, or perhaps better said linguistic [90], explanation for certain correspondences in terms of the common terms that are sometimes used to describe different kinds of stimuli [91,92]. Here, for example, one might think only of the words “high” and “low” that are used in English to describe both pitch and spatial elevation (though see also [93]). And, finally, there is also the increasingly-popular emotional mediation account of crossmodal correspondences. While this account has been used by researchers working with both simple and complex stimuli, it would seem to account for more of the variance in the data when

accounting for correspondences involving more complex (and, as it happens, typically emotionally complex/meaningful) stimuli (e.g., music, paintings).

4. CONCLUSIONS

There has been a rapid growth of interest in crossmodal correspondences research in recent years amongst experimental psychologists, cognitive neuroscientists, and practitioners (e.g., in the multimedia community) [94–96]. Much of that research interest has tended to focus on the surprising associations that people experience with simple auditory stimuli (i.e., features such as pitch and, to a lesser extent, loudness and timbre). As well as providing a brief summary of that literature, in this tutorial review, I have also attempted to summarize a number of the correspondences involving more complex auditory stimuli (typically short excerpts taken from composed music; [12,97,98]).

REFERENCES

- [1] C. Spence, “Crossmodal correspondences: A tutorial review,” *Attent. Percept. Psychophys.*, **73**, 971–995 (2011).
- [2] L. E. Marks, “Cross-modal interactions in speeded classification,” in *Handbook of Multisensory Processes*, G. A. Calvert, C. Spence and B. E. Stein, Eds. (MIT Press, Cambridge, MA, 2004), pp. 85–105.
- [3] C. Spence, “Crossmodal correspondences: A tutorial review,” in *Senses and Sensation: Critical and Primary Sources*, D. Howes, Ed., Vol. III (Bloomsbury Academic, London, 2018), pp. 91–125. <https://www.bloomsbury.com/uk/senses-and-sensation-9781474274050/> (accessed 20 Oct. 2018).
- [4] C. V. Parise and C. Spence, “Audiovisual crossmodal correspondences in the general population,” in *The Oxford Handbook of Synesthesia*, J. Simner and E. M. Hubbard, Eds. (Oxford University Press, Oxford, 2013), pp. 790–815.
- [5] K. M. Knöferle and C. Spence, “Crossmodal correspondences between sounds and tastes,” *Psychon. Bull. Rev.*, **19**, 992–1006 (2012).
- [6] O. Deroy, A.-S. Crisinel and C. Spence, “Crossmodal correspondences between odors and contingent features: Odors, musical notes, and geometrical shapes,” *Psychon. Bull. Rev.*, **20**, 878–896 (2013).
- [7] N. Dreksler and C. Spence, “A critical analysis of colour-shape correspondences: Examining the replicability of colour-shape associations,” *i-Perception*, **10**(2), pp. 1–34 (2019).
- [8] O. Deroy, I. Fernandez-Prieto, J. Navarra and C. Spence, “Unravelling the paradox of spatial pitch,” in *Spatial Biases in Perception and Cognition*, T. L. Hubbard, Ed. (Cambridge University Press, Cambridge, UK, 2018), pp. 77–93.
- [9] C. V. Parise and C. Spence, “Audiovisual crossmodal correspondences and sound symbolism: An IAT study,” *Exp. Brain Res.*, **220**, 319–333 (2012).
- [10] K. K. Evans and A. Treisman, “Natural cross-modal mappings between visual and auditory features,” *J. Vis.*, **10**(1):6, pp. 1–12 (2010).
- [11] M. Adeli, J. Rouat and S. Molotchnikoff, “Audiovisual correspondence between musical timbre and visual shapes,” *Front. Hum. Neurosci.*, **8**, 352 (2014).
- [12] P. Walker, “Cross-sensory correspondences: A theoretical framework and their relevance to music,” *Psychomusicol.: Music Mind Brain*, **26**, 103–116 (2016).
- [13] C. Parise and C. Spence, “‘When birds of a feather flock together’: Synesthetic correspondences modulate audiovisual integration in non-synesthetes,” *PLoS ONE*, **4**(5), e5664 (2009).
- [14] P. Winkielman, M. Ziembowicz and A. Nowak, “The coherent and fluent mind: How unified consciousness is constructed from cross-modal inputs via integrated processing experiences,” *Front. Psychol.*, **6**, 83 (2015).
- [15] A. Gallace and C. Spence, “Multisensory synesthetic interactions in the speeded classification of visual size,” *Percept. Psychophys.*, **68**, 1191–1203 (2006).
- [16] L. E. Marks, “On cross-modal similarity: Auditory–visual interactions in speeded discrimination,” *J. Exp. Psychol. Hum. Percept. Perform.*, **13**, 384–394 (1987).
- [17] G. Martino and L. E. Marks, “Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis,” *Perception*, **28**, 903–923 (1999).
- [18] R. D. Melara and T. P. O’Brien, “Interaction between synesthetically corresponding dimensions,” *J. Exp. Psychol. Gen.*, **116**, 323–336 (1987).
- [19] E. Ben-Artzi and L. E. Marks, “Visual-auditory interaction in speeded classification: Role of stimulus difference,” *Percept. Psychophys.*, **57**, 1151–1162 (1995).
- [20] I. H. Bernstein and B. A. Edelstein, “Effects of some variations in auditory input upon visual choice reaction time,” *J. Exp. Psychol.*, **87**, 241–247 (1971).
- [21] G. R. Patching and P. T. Quinlan, “Garner and congruence effects in the speeded classification of bimodal signals,” *J. Exp. Psychol. Hum. Percept. Perform.*, **28**, 755–775 (2002).
- [22] H. H. Clark and H. H. Brownell, “Position, direction, and their perceptual integrality,” *Percept. Psychophys.*, **19**, 328–334 (1976).
- [23] I. H. Bernstein, T. R. Eason and D. L. Schurman, “Hue-tone interaction: A negative result,” *Percept. Mot. Skills*, **33**, 1327–1330 (1971).
- [24] A. Klapetek, M. K. Ngo and C. Spence, “Do crossmodal correspondences enhance the facilitatory effect of auditory cues on visual search?” *Attent. Percept. Psychophys.*, **74**, 1154–1167 (2012).
- [25] E. Orchard-Mills, D. Alais and E. Van der Burg, “Cross-modal associations between vision, touch and audition influence visual search through top-down attention, not bottom-up capture,” *Attent. Percept. Psychophys.*, **75**, 1892–1905 (2013).
- [26] L. M. Getz and M. Kubovy, “Questioning the automaticity of audiovisual correspondences,” *Cognition*, **175**, 101–108 (2018).
- [27] C. Spence and O. Deroy, “How automatic are crossmodal correspondences?” *Conscious. Cognit.*, **22**, 245–260 (2013).
- [28] N. Bien, S. ten Oever, R. Goebel and A. T. Sack, “The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG, and psychophysics study,” *NeuroImage*, **59**, 663–672 (2012).
- [29] K. P. Revell, L. L. Namy, L. C. DeFife and L. C. Nygaard, “Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI,” *Brain Lang.*, **128**, 18–24 (2014).
- [30] S. Sadaghiani, J. X. Maier and U. Noppeney, “Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing,” *J. Neurosci.*, **29**, 6490–6499 (2009).
- [31] C. Spence and K. Sathian, “Audiovisual crossmodal correspondences: Behavioural consequences and neural underpinnings,” To appear in *Multisensory Perception: From Laboratory to Clinic*, K. Sathian and V. S. Ramachandran, Eds. (Elsevier, New York, in press).
- [32] V. U. Ludwig, I. Adachi and T. Matzuzawa, “Visuoauditory

- mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans," *Proc. Natl. Acad. Sci. USA*, **108**, 20661–20665 (2011).
- [33] D. J. Lewkowicz and G. Turkewitz, "Cross-modal equivalence in early infancy: Auditory-visual intensity matching," *Dev. Psychol.*, **16**, 597–607 (1980).
- [34] P. Walker, J. G. Bremner, U. Mason, J. Spring, K. Mattock, A. Slater and S. P. Johnson, "Preverbal infants' sensitivity to synesthetic cross-modality correspondences," *Psychol. Sci.*, **21**, 21–25 (2010).
- [35] S. Dolscheid, S. Hunnius, D. Casasanto and A. Majid, "Prelinguistic infants are sensitive to space-pitch associations found across cultures," *Psychol. Sci.*, **25**, 1256–1261 (2014).
- [36] E. Haryu and S. Kajikawa, "Are higher-frequency sounds brighter in color and smaller in size? Auditory-visual correspondences in 10-month-old infants," *Infant Behav. Dev.*, **35**, 727–732 (2012).
- [37] R. H. Simpson, M. Quinn and D. P. Ausubel, "Synaesthesia in children: Association of colors with pure tone frequencies," *J. Gen. Psychol.*, **89**, 95–103 (1956).
- [38] O. Deroy and C. Spence, "Learning 'arbitrary' crossmodal correspondences: Staying away from neonatal synaesthesia," *Neurosci. Biobehav. Rev.*, **37**, 1240–1253 (2013).
- [39] S. Lacey, M. Martinez, K. McCormick and K. Sathian, "Synesthesia strengthens sound-symbolic cross-modal correspondences," *Eur. J. Neurosci.*, **44**, 2716–2721 (2016).
- [40] D. Brang, L. E. Williams and V. S. Ramachandran, "Grapheme-color synesthetes show enhanced crossmodal processing between auditory and visual modalities," *Cortex*, **48**, 630–637 (2012).
- [41] O. Deroy and C. Spence, "Weakening the case for 'weak synaesthesia': Why crossmodal correspondences are not synaesthetic," *Psychon. Bull. Rev.*, **20**, 643–664 (2013).
- [42] R. Brunetti, A. Indraccolo, C. Del Gatto, C. Spence and V. Santangelo, "Are crossmodal correspondences absolute or relative? Context effects on speeded classification," *Attent. Percept. Psychophys.*, **80**, 527–534 (2018).
- [43] Y. Jamal, S. Lacey, L. Nygaard and K. Sathian, "Interactions between auditory elevation, auditory pitch and visual elevation during multisensory perception," *Multisens. Res.*, **30**, 287–306 (2017).
- [44] K. McCormick, S. Lacey, R. Stilla, L. C. Nygaard and K. Sathian, "Neural basis of crossmodal correspondence between auditory pseudowords and visual shapes," *bioRxiv* doi: <http://dx.doi.org/10.1101/478347> (2018).
- [45] C. Parise and C. Spence, "Synaesthetic congruency modulates the temporal ventriloquism effect," *Neurosci. Lett.*, **442**, 257–261 (2008).
- [46] E. Orchard-Mills, E. Van der Burg and D. Alais, "Crossmodal correspondence between auditory pitch and visual elevation affects temporal ventriloquism," *Perception*, **45**, 409–424 (2016).
- [47] R. Brunetti, A. Indraccolo, S. Mastroberardino, C. Spence and V. Santangelo, "The impact of cross-modal correspondences on working memory performance," *J. Exp. Psychol. Hum. Percept. Perform.*, **43**, 819–831 (2017).
- [48] L. Brunel, P. F. Carvalho and R. L. Goldstone, "It does belong together: Cross-modal correspondences influence cross-modal integration during perceptual learning," *Front. Psychol.*, **6**, 358 (2015).
- [49] M. I. Posner, M. J. Nissen and R. M. Klein, "Visual dominance: An information-processing account of its origins and significance," *Psychol. Rev.*, **83**, 157–171 (1976).
- [50] C. Spence, D. I. Shore and R. M. Klein, "Multimodal prior entry," *J. Exp. Psychol. Gen.*, **130**, 799–832 (2001).
- [51] C. Spence, "On the relationship(s) between colour and taste," *Exp. Psychol.* (in press).
- [52] S. E. Palmer, K. B. Schloss, Z. Xu and L. R. Prado-León, "Music-color associations are mediated by emotion," *Proc. Natl. Acad. Sci. USA*, **110**, 8836–8841 (2013).
- [53] A. Gabriellson, "Emotion perceived and emotion felt: Same or different?" *Musica. Sci.*, **5**, 123–147 (2002). doi: 10.1177/10298649020050S105
- [54] J. Cespedes-Guevara and T. Eerola, "Music communicates affects, not basic emotions—A constructionist account of attribution of emotional meanings to music," *Front. Psychol.*, **9**, 215 (2018).
- [55] K. L. Whiteford, K. B. Schloss, N. E. Helwig and S. E. Palmer, "Color, music, and emotion: Bach to the blues," *i-Perception*, **9**(6), pp. 1–27 (2018).
- [56] C. E. Osgood, G. J. Suci and P. H. Tannenbaum, *The Measurement of Meaning* (University of Illinois Press, Urbana, 1957).
- [57] T. F. Karwowski, H. S. Odbert and C. E. Osgood, "Studies in synesthetic thinking. II. The rôle of form in visual responses to music," *J. Gen. Psychol.*, **26**, 199–222 (1942).
- [58] D. Janković, "Cross-modal nature of evaluative meaning," in *Different Psychological Perspectives on Cognitive Processes: Current Research Trends in Alps-Adria Region*, A. Galmonte and R. Actis-Grosso, Eds. (Cambridge Scholars Publishing, Newcastle upon Tyne, 2014), pp. 58–75.
- [59] M. Suzuki, J. Gyoba and Y. Sakuta, "Multichannel NIRS analysis of brain activity during semantic differential rating of drawing stimuli containing different affective polarities," *Neurosci. Lett.*, **375**, 53–58 (2005).
- [60] S. E. Palmer, T. A. Langlois and K. B. Schloss, "Music-to-color associations of single line piano melodies in non-synesthetes," *Multisens. Res.*, **29**, 157–193 (2016).
- [61] T. F. Karwowski and H. S. Odbert, "Color-music," *Psychol. Mon.*, **50**(2, Whole No. 22) (1938).
- [62] J. M. Barbieri, A. Vidal and D. A. Zellner, "The color of music: Correspondence through emotion," *Emp. Stud. Arts*, **25**, 193–208 (2007).
- [63] A. K. Lindborg and A. K. Friberg, "Colour association with music is mediated by emotion: Evidence from an experiment using a CIE Lab interface and interviews," *PLoS ONE*, **10**(12), e0144013 (2015).
- [64] E. S. Isbilen and C. L. Krumhansl, "The color of music: Emotion-mediated associations to Bach's Well-Tempered Clavier," *Psychomusicol. Music Mind Brain*, **26**, 149–161 (2016).
- [65] R. Sebba, "Structural correspondence between music and color," *Col. Res. Appl.*, **16**, 81–88 (1991).
- [66] A. Wells, "Music and visual color: A proposed correlation," *Leonardo*, **13**, 101–107 (1980).
- [67] R. Bresin, "What is the color of that music performance?" *Proc. Int. Computer Music Conf.*, Barcelona, Spain: Int. Computer Music Assoc., pp. 367–370 (2005).
- [68] H. N. J. Schifferstein and I. Tanudjaja, "Visualizing fragrances through colors: The mediating role of emotions," *Perception*, **33**, 1249–1266 (2004).
- [69] Q. (J.) Wang, S. Wang and C. Spence, "'Turn up the taste': Assessing the role of taste intensity and emotion in mediating crossmodal correspondences between basic tastes and pitch," *Chem. Senses*, **41**, 345–356 (2016).
- [70] G. L. Collier, "Affective synaesthesia: Extracting emotion space from simple perceptual stimuli," *Motiv. Emot.*, **20**, 1–32 (1996).
- [71] W. G. Collier and T. L. Hubbard, "Judgements of happiness, brightness, speed and tempo change of auditory stimuli varying

- in pitch and tempo,” *Psychomusicology*, **17**, 36–55 (2001).
- [72] J. Bhattacharya and J. P. Lindsen, “Music for a brighter world: Brightness judgment bias by musical emotion,” *PLoS ONE*, **11**(2) (2016).
- [73] A. B. M. Gerdes, M. J. Wieser and G. W. Alpers, “Emotional pictures and sounds: A review of multimodal interactions of emotion cues in multiple domains,” *Front. Psychol.*, **5**, 1351 (2014).
- [74] C. A. Levitan, S. A. Charney, K. B. Schloss and S. E. Palmer, “The smell of jazz: Crossmodal correspondences between music, odor, and emotion,” *Cogsci Proc.*, Pasadena, CA, pp. 1326–1331 (2015).
- [75] L. Albertazzi, L. Canal and R. Micciolo, “Cross-modal association between materic painting and classical Spanish music,” *Front. Psychol.*, **6**, 424 (2015).
- [76] N. Hasenfus, C. Martindale and D. Birnbaum, “Psychological reality of cross-media artistic styles,” *J. Exp. Psychol. Hum. Percept. Perform.*, **9**, 841–863 (1983).
- [77] A. C. Duthie, “Do music and art influence one another? Measuring cross-modal similarities in music and art,” Iowa State University (2013).
- [78] C. Duthie and B. Duthey, “Do music and art influence one another? Measuring cross-modal similarities in music and art,” *Polymath Interdisciplin. Arts Sci. J.* (2015). <https://bradduthie.github.io/Duthie%26Duthie2015.pdf>.
- [79] P. H. Tannenbaum, “Music background in the judgment of stage and television drama,” *Audio-Visual Commun. Rev.*, **4**, 92–101 (1956).
- [80] M. Boltz, “Music videos and visual influences on music perception and appreciation: Should you want your MTV?” in *The Psychology of Music in Multimedia*, S. L. Tan, A. Cohen, S. Lipscomb and R. Kendall, Eds. (OUP, Oxford, 2013), pp. 217–235.
- [81] L. Cheskin, *Marketing Success: How to Achieve It* (Cahners Books, Boston, 1972).
- [82] C. Spence, “Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism,” *J. Consum. Psychol.*, **22**, 37–54 (2012).
- [83] S. Brown, X. Gao, L. Tisdelle, S. B. Eickhoff and M. Liotti, “Naturalizing aesthetics: Brain areas for aesthetic appraisal across sensory modalities,” *NeuroImage*, **58**, 250–258 (2011).
- [84] R. Reber, “Processing fluency, aesthetic pleasure, and culturally shared taste,” in *Aesthetic Science: Connecting Minds, Brains, and Experience*, A. P. Shimamura and S. E. Palmer, Eds. (Oxford University Press, Oxford, 2012), pp. 223–249.
- [85] C. V. Parise, K. Knorre and M. O. Ernst, “Natural auditory scene statistics shapes human spatial hearing,” *Proc. Natl. Acad. Sci. USA*, **111**, 6104–6108 (2014).
- [86] M. A. K. Peters, J. Balzer and L. Shams, “Smaller = denser, and the brain knows it: Natural statistics of object density shape weight expectations,” *PLoS ONE*, **10**(3), e0119794 (2015).
- [87] M. O. Ernst, “Learning to integrate arbitrary signals from vision and touch,” *J. Vis.*, **7**/5/7: 1–14 (2007).
- [88] S. S. Stevens, “On the psychophysical law,” *Psychol. Rev.*, **64**, 153–181 (1957).
- [89] J. L. Caivano, “Color and sound: Physical and psychophysical relations,” *Color Res. Appl.*, **19**, 126–133 (1994).
- [90] P. Walker, “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 (2012).
- [91] G. Martino and L. E. Marks, “Cross-modal interaction between vision and touch: The role of synesthetic correspondence,” *Perception*, **29**, 745–754 (2000).
- [92] L. E. Marks, “On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness,” *J. Exp. Psychol. Hum. Percept. Perform.*, **15**, 586–602 (1989).
- [93] C. Parkinson, P. J. Kohler, B. Sievers and T. Wheatley, “Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population,” *Perception*, **41**, 854–861 (2012).
- [94] S. L. Tan, A. Cohen, S. Lipscomb and R. Kendall, Eds., *The Psychology of Music in Multimedia* (Oxford University Press, Oxford, 2013).
- [95] M. Haverkamp, *Synesthetic Design: Handbook for a Multi-sensory Approach* (Birkhäuser, Basel, 2014).
- [96] N. Davis, “Welcome to the Tate Sensorium, where the paintings come with chocolates,” *The Guardian*, August 22nd. <http://www.theguardian.com/artanddesign/video/2015/aug/25/welcome-tate-sensorium-taste-touch-smell-art-video> (2015).
- [97] Z. Eitan, “Musical connections: Cross-modal connections,” in *The Routledge Companion to Music Cognition*, R. Ashley and R. Timmers, Eds. (Taylor & Francis, Abingdon-on-Thames, UK, 2017), pp. 213–224.
- [98] R. Walker, “The effects of culture, environment, age, and musical training on choices of visual metaphors for sound,” *Percept. Psychophys.*, **42**, 491–502 (1987).