Cross-Sensory Correspondences: A Theoretical Framework and Their Relevance to Music

Peter Walker [1, 2]

1 Department of Psychology, Lancaster University, UK

2 Department of Psychology, Sunway University, Malaysia

Author Note

Peter Walker, Department of Psychology, Lancaster University and Department of Psychology, Sunway University, Malaysia

Correspondence concerning this article should be addressed to Peter Walker, Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK. Email: p.walker@lancaster.ac.uk

Abstract

Each of our senses is 'blind' to some features of objects and events (e.g., hearing can tell us little or nothing about the shape, colour, and weight of an object, or about how it might taste or smell). When we listen to sounds without support from other sensory modalities, such as when listening to recorded music, how do will fill-in these blind spots? Evidence identifying a core set of cross-sensory correspondences among basic stimulus features is reviewed, and it is proposed that they offer a potential basis for the filling-in of information that is missing when one or more sensory systems is not available. An emerging theoretical framework for understanding correspondences and their impact on behaviour is presented. Evidence pertaining to key features of the framework is reviewed, including that cross-sensory correspondences are based on cross-talk among conceptual representations of aligned feature dimensions, are bi-directional in their effects, obey transitivity in the feature associations they support, involve the relative (context-sensitive) coding of stimulus features, and can be accessed through the verbal specification of feature values. After illustrating how cross-sensory correspondences are able to embrace basic features of bodily actions, gestures, and vocalisations, their potential for exploitation in the communication of ideas is explained. The relevance of cross-sensory correspondences to musical sounds, and their potential to enhance the composition, performance, and appreciation of music, are discussed.

Cross-Sensory Correspondences: A Theoretical Framework and Their Relevance to Music

In our everyday interactions with the environment we gather information about objects and events simultaneously through different sensory modalities, with the acquisition of information about some features of objects and events being duplicated across modalities. For example, vision and touch can both discern the size and shape of a tangible object, though touch normally requires the object to be within reach. In such cases of duplication it need not be a problem if, for some reason, one of the modalities cannot be used, because the same information can be gathered using a modality that is available (e.g., when vision confirms the size and shape of a tangible object that is out of reach).

Information about some features, however, is available only through a single modality. A novel object's brightness and colour, for example, can be identified only through vision, the pitch of the sound it is making only through audition, and its weight only through dynamic touch. In such cases problems do arise when the modality best placed to provide information about a particular feature is not available, and there is considerable interest in how the 'blind spots' this creates are filled-in. How is it that in everyday life we readily refer to, for example, the *brightness* of a sound that cannot be seen, the *loudness* of a shirt that cannot be heard, and the *thickness* and *heaviness* of an aroma that cannot be seen or grasped? Perhaps such features need not be restricted to a single sensory channel, but instead can be shared by stimuli encoded in different sensory channels. This would be possible if the features are amodal and conceptual in nature, allowing them to transcend the individual modality with which they are normally most directly associated. In this way sounds will share their brightness and loudness with visual stimuli, and odours will share their thickness and weight with objects seen and felt.

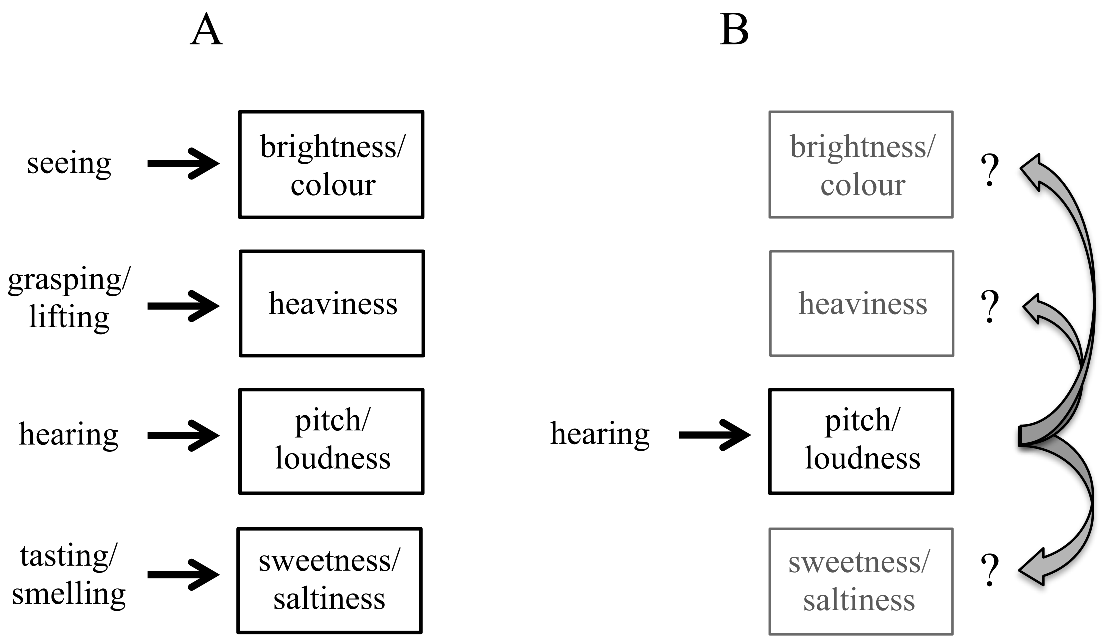

Figure 1. A. How some features of objects and events are available only through a single modality. B. When we hear a novel sound in the absence of any relevant concomitant information from another sensory modality, what sense do we have of the features normally conveyed by vision, dynamic touch, gustation and olfaction?

Sounds are often encountered without support from other sensory modalities, such as when listening to recorded music where the music cannot be seen, touched, lifted, or tasted (see Figure 1). Where this is the case, does the listener try to fill-in the missing information? What colours should they imagine seeing, what shapes and textures should they imagine seeing and touching, and what sense of heaviness should they feel? Can the basic features of simple and complex sounds, such as their pitch and the abruptness of their amplitude envelope, guide the filling-in of missing information and, if so, according to what rules? Evidence from various sources, starting with visual-hearing synaesthesia, suggests they do and that cross-sensory correspondences provide at least some of the rules for doing so.

After reviewing evidence identifying a core set of cross-sensory correspondences among basic stimulus features, a theoretical framework is introduced for understanding correspondences and their impact on behaviour. Evidence supporting key features of the framework is reviewed, including that cross-sensory correspondences embrace basic features of bodily actions, gestures, and vocalisations. Their potential for exploitation in the communication of ideas is then explained. The review concludes by highlighting the relevance of cross-sensory correspondences to musical sounds and their potential to enhance the composition, performance, and appreciation of music.

The Cross-Sensory Features of Sounds

Cross-sensory features in sound-induced imagery. When visual images are induced by non-speech sounds in auditory-visual synaesthesia, higher pitch sounds tend to induce images that are brighter, higher in their spatial elevation, lighter in weight, more likely to be moving, sharper, and smaller than those induced by lower pitch sounds (e.g., Chiou, Stelter & Rich, 2013; Karwoski & Odbert, 1938; Marks, 1974, 1975, 1978; Ward, Huckstep & Tsakanikos, 2006). This indicates that there are some systematic cross-sensory associations underlying visual-hearing synaesthesia. Other evidence, in various forms, confirms the same cross-sensory associations involving auditory pitch in people who are not regarded as being synaesthetes (i.e., the general population). For example, when typical English speaking adults draw music they are listening to, they draw lines and forms that are more angular (sharper), brighter, higher on the page, smaller, and thinner, the higher in pitch and/or faster in tempo is the music (Karwoski, Odbert, & Osgood, 1942; Kussner & Leech-Wilkinson, 2013).[1]

Judging the cross-sensory features of sounds. When they are explicitly asked to indicate what cross-sensory features are possessed by simple sounds (typically pure tones) differing in pitch (preferably matched for perceived loudness, see P. Walker & Smith, 1984; L. Walker, P. Walker, & Francis, 2012), non-synaesthetes confirm the same associations, judging higher pitch sounds to be, among other things, more active, brighter, faster, higher in space, lighter in weight, shallower, sharper (more angular/pointier), smaller, and thinner than lower pitch sounds (Boltz, 2011; Collier & Hubbard, 2001, 2004; Eitan & Timmers, 2010; Marks, 1974, 1975, 1978; Perrott, Musicant, & Schwethelm, 1980; Tarte, 1982; L. Walker, P. Walker, & Francis, 2012; P. Walker & Smith, 1984). Whether they are also harder than lower pitch sounds is not always clear, though they are consistently judged to be more feminine (Eitan & Timmers, 2010; Tarte, 1982). The same associations also emerge with indirect questioning. For example, when young children indicate which of two bouncing objects they think is making a higher pitch impact sound, they point to the smaller or brighter of the two objects (Mondloch & Maurer, 2004).

Cross-sensory features of sounds in speeded classification. Because they avoid inducing people to deliberately generate systematic associations on the fly, associations that might otherwise not be in place, studies in which cross-sensory features influence

1 The term *cross-sensory* is preferred over *cross-modality* because, as Karwoski, Odbert, and Osgood (1942) point out (*op. cit.,* p. 213), although the cross-modality association of feature values is most obvious, similar systematic associations occur within modalities, such as between the brightness of a visual form and its angularity (see, for example, P. Walker, 2012a).

behaviour automatically are especially important. Speeded classification tasks are relevant in this regard (see Marks, 2004). Thus, when people classify stimuli on the basis of a criterial feature (e.g., classify a visual stimulus according to whether its surface is bright or dark), they are influenced by whether an accompanying incidental stimulus has associated (congruent) or non-associated (incongruent) features (e.g., whether an accompanying sound is high in pitch or low in pitch). More specifically, people respond more quickly and accurately when the criterial and incidental feature values are congruent with each other, rather than when they are incongruent. Where the incidental feature concerns the pitch of a sound, the congruity effect it induces confirms that the criterial feature is a cross-sensory feature associated with auditory pitch (e.g., if the congruity in the pitch of an incidental sound were to influence the fluency with which the brightness of a visual stimulus was confirmed, then brightness would be deemed to be a cross-sensory feature associated with pitch).

People are faster to classify a visual stimulus as bright when it is accompanied by a high-pitched sound (a bright sound) rather than a low-pitched sound (a dark sound) (Marks, 1987; Melara, 1989; Martino & Marks, 1999), and this confirms the associations observed elsewhere linking higher pitch sounds to brighter visual stimuli. There is equivalent evidence confirming the links between higher pitch sounds and smaller visual stimuli (Evans & Treisman, 2010; Gallace & Spence, 2006), pointier visual stimuli (Marks, 1987; P. Walker 2012a), thinner visual stimuli (Evans & Treisman, 2010), and spatially higher visual stimuli (Ben-Artzi & Marks, 1995; Bernstein & Edelstein, 1971; Chiou & Rich, 2012; Evans & Treisman, 2010; Melara & O'Brien, 1987; Patching & Quinlan, 2002; Sadaghiani et al., 2009), to tactile and haptic sensations located higher in space (Occelli, Spence, & Zampini, 2009; Rusconi et al., 2006), and to vibrotactile stimulation at higher rates of vibration (Ro, Hsu, Yasar, Elmore, & Beauchamp, 2009).

Cross-Sensory Correspondences and the Alignment of Feature Dimensions

How should we think about cross-sensory feature associations, such as those involving auditory pitch? We can begin by acknowledging that all the features enjoying an association with auditory pitch are dimensional in nature (i.e., their feature values lie on a continuum). We might then contemplate that it is the relative positioning of feature values on these dimensions that is shared with the pitch of a sound. The resulting systematicity in the cross-sensory associations (i.e., their conforming to a 'rule' reflecting the alignment of different dimensions) is what the term *correspondence* is intended to capture.

Karwoski, Odbert, and Osgood (1942) elaborate on how we might think about cross-sensory correspondences. They propose that elementary stimulus features (e.g., visual surface brightness, visual angularity, auditory pitch) are rich in conceptual connotations, and that the conceptual dimensions along which their values lie are aligned in ways that determine the correspondences evident in cross-sensory induced imagery. With regard to how the alignment of these dimensions shapes such imagery, Karwoski et al. propose that:

> *The synesthetic or analogical process appears to be the parallel alignment of two gradients in such a way that the appropriate extremes are related, followed in some cases by translation in terms of equivalent parts of the two gradients thus paralleled (op. cit., p. 217).*

In this way, Karwoski et al. anticipate claims that cross-sensory correspondences involve the modality-independent conceptual representation of elementary stimulus features (see

Martino & Marks, 1999; Melara & Marks, 1990; P. Walker & Smith, 1984). In addition, their notion of 'translation' anticipates recent claims that such correspondences involve crosstalk (cross-activation) between correspondingly positioned feature values on different dimensions, with the dimensions including those evident in the correspondences emerging when contrasting levels of auditory pitch are explored (see Figure 2).

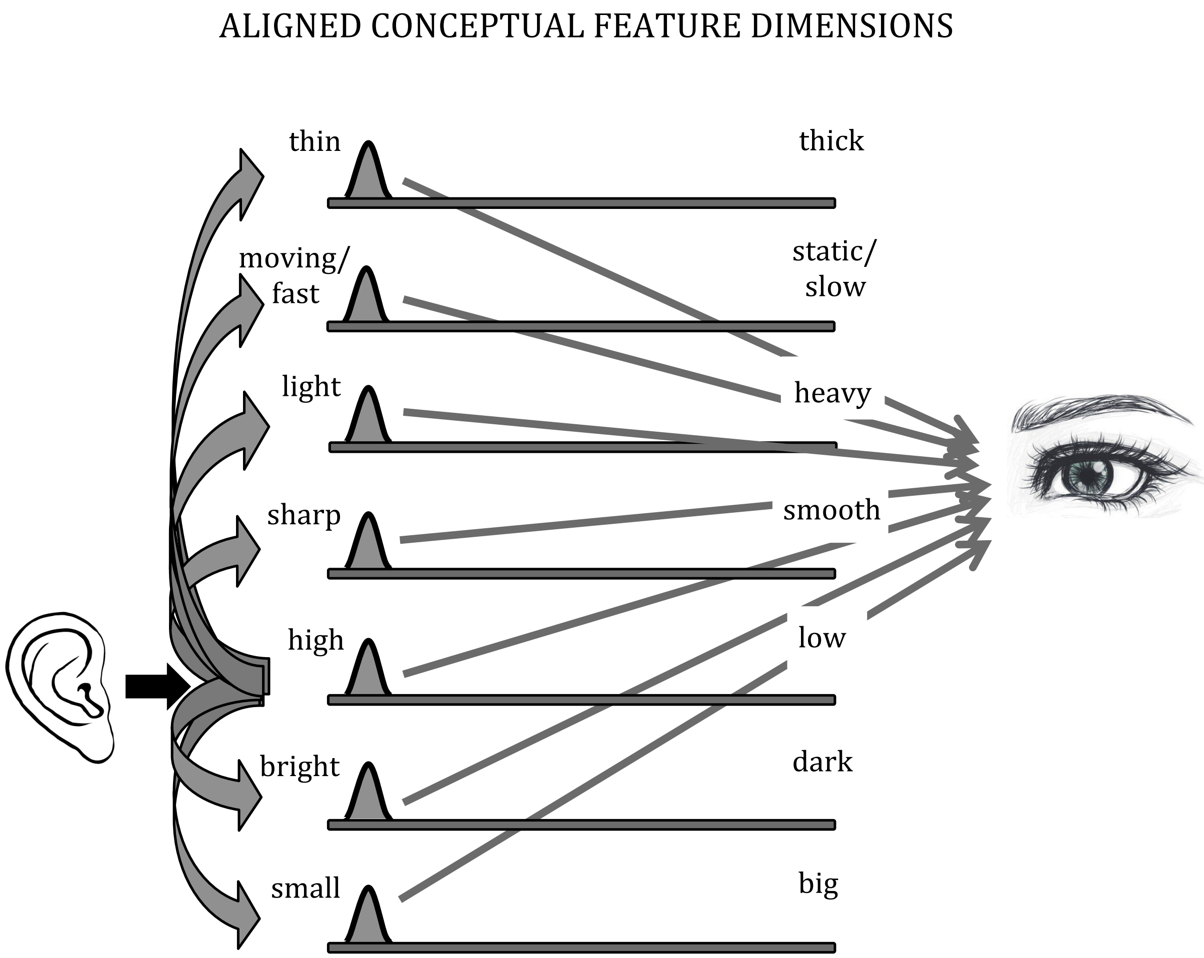

Figure 2. Cross-sensory correspondences evident in the visual imagery induced by sounds of contrasting pitch are thought to arise from the alignment, *en bloc*, of several conceptual dimensions (based on Karwoski, Odbert, & Osgood, 1942). Here, a relatively high-pitched sound induces visual images that, amongst other things, are relatively bright, high in space, light in weight, moving/fast, sharp, small, and thin.

Though not shown here, it is assumed that extensive bi-directional activation occurs between corresponding places across all the dimensions (see Figures 4 & 5 and the later section dealing with transitivity).

The core set of cross-sensory correspondences linking auditory pitch to other cross-sensory features offers a potential basis for the filling-in of information that occurs when novel sounds are heard in the absence of any concomitant information from other sensory channels. For example, because auditory pitch and visual brightness enjoy a corresponding relationship, novel sounds that are high-pitched normally ‘feel’ as though they are emanating from bright objects when the source of the auditory information cannot be seen.

A Common Scheme for the Cross-Sensory Mapping of Features

If the same core set of correspondences were to link features across all sensory channels, then they would offer a very powerful basis for the filling-in of missing information more generally, beyond the filling-in that is observed when people hear sounds of different pitch in isolation. But this requires the aligned dimensions to be conceptual and amodal in nature, so that the same correspondences will emerge whatever contrasting sensory features are used to probe them (e.g., whether correspondences are probed by visual stimuli contrasting in brightness, aromas contrasting in heaviness, or haptic objects and sounds contrasting in sharpness). The evidence indicates that this is the case.[2]

---

[2] The claim is not that interactions among sensory-perceptual representations are unable to support cross-sensory correspondences. Instead it is that interactions among conceptual representations can do so and in ways that explain aspects of correspondences

Judging the mapping of cross-sensory features beyond auditory pitch. When people indicate the cross-sensory features linked to the brightness and size of objects they can only see, they indicate that darker and bigger objects are heavier and make lower pitch sounds than brighter and smaller objects (P. Walker 2012b; P. Walker, Francis, & L. Walker, 2010; L. Walker et al., 2012).[3] And when people are asked to indicate the cross-sensory features possessed by pointy versus curved visual shapes, pointier shapes are judged to be brighter, faster, lighter in weight, higher in space, and to make higher pitch sounds than curved shapes (P. Walker, 2012a; L. Walker et al., 2012). Likewise, sequences of sounds at faster tempi are judged to be brighter than sequences at slower tempi (Collier & Hubbard, 2001). And when hidden objects varying in size are explored by touch alone, smaller objects are judged to be brighter, faster, harder, higher in space, thinner, sharper, and to make higher pitch sounds compared to bigger objects (P. Walker

---

that are not amenable to explanations based entirely on sensory-perceptual representations. These aspects are reviewed in the present paper and include the bi-directionality and transitivity of correspondences, their potential to be engaged with feature values that are specified verbally, and their sensitivity to the relative/context-sensitive values of features rather than to their absolute values (see L. Walker & P. Walker, 2015, and P. Walker, L. Walker, & Francis, 2015, for recent discussion of these issues). Note also that a conceptual basis for correspondences is not incompatible with their appearance in preverbal infants (e.g., P. Walker, Bremner, et al., 2010). Preverbal infants, including neonates, are known to be capable of acquiring and utilising abstracted categories of stimuli (i.e., concepts) (e.g., Bomba & Siqueland, 1983).

3 And in the case of the correspondence between the brightness of objects and their weights, an illusion of perceived heaviness is induced when the objects are lifted. That is, when otherwise identical objects varying in visual brightness are lifted, the brighter objects are perceived to be heavier, rather than lighter, than the darker objects (an illusion analogous to the classic size-weight illusion) (P. Walker, L. Walker, & Francis, 2010).

& Smith, 1985; P. Walker & L. Walker, 2012; L. Walker et al., 2012). And, as a final example, when people lift unseen objects differing in weight they judge heavier objects to be lower in pitch, bigger, darker, less sharp, slower, and thicker than objects that are lighter in weight (P. Walker, Scallon, & Francis, 2015). In summary, when cross-sensory associations are probed with a wide range of contrasting sensory features, other than auditory pitch, the same core correspondences continue to be observed.

The cross-sensory mapping of features in speeded classification. Converging evidence that the same correspondences emerge whatever feature contrast is used to probe them comes from the congruity effects observed in speeded classification tasks. Thus, higher frequency tactile vibrations on the hand are congruent both with higher levels of visual brightness (white shapes rather than black shapes) (Martino & Marks, 2000) and with higher levels of auditory pitch (Ro, Hsu, Yasar, Elmore & Beauchamp, 2009). Tactile sensations on the hand that are located higher in space also prove to be congruent with higher pitch sounds (Occelli, Spence, & Zampini, 2009). Similarly, visual shapes that are more angular are congruent with higher levels of surface brightness (P. Walker, 2012a). And, as a final example, visual stimuli with brighter surfaces are observed to be congruent with smaller tactile objects (i.e., objects that are hidden from view) (P. Walker & L. Walker, 2012).

Summary. The consistent appearance of the same core set of cross-sensory correspondences, regardless of the sensory channel through which they are probed, suggests two things. First, the aligned dimensions on which correspondences are based are modality-independent and conceptual in nature (e.g., it is an amodal concept of brightness, rather than specifically visual brightness, that is aligned with an amodal concept of elevation, not specifically auditory pitch or visuo-spatial elevation). Second, the conceptual feature dimensions remain aligned with each other in the same way whatever stimulus contrast is explored. An important principle follows from these two points, namely that cross-sensory correspondences are bi-directional.

The Bi-Directionality of Cross-Sensory Correspondences

Associations among modality-specific, rather than amodal, representations could reasonably be expected to be uni-directional. For example, we often hear something before we see what is making the sound, in part because the 'field of view' for hearing is not restricted in the same way as it is for vision (e.g., we can hear sounds originating from behind us but cannot see what is making the sound). Therefore, when we hear a high pitch sound we might generate an expectation that we will see a small object. When we see a small object, however, we are much less likely to generate equivalent expectations regarding what we might hear because in the majority of cases the object will not make a sound at all. This will result in an association between auditory pitch and visual size that is largely uni-directional. Similarly, because we normally see objects before we lift them, the modality-specific association between the brightness of an object and its felt heaviness also is likely to be uni-directional. When we see a dark object we might expect it to feel heavy, but when we feel something heavy there is normally no incentive to generate any expectations about its visual appearance because this has already been observed.

The opening discussion of cross-sensory associations focused on the properties people judge sounds contrasting in pitch to possess, and it was the consistent associations emerging from these judgments that first indicated the nature of the dimensions underlying correspondences and the manner of their alignment. In other studies, however, the same cross-sensory associations with pitch have been probed in the opposite direction.

That is, other feature contrasts have been presented to people and one of the judgments they have had to make has concerned the levels of auditory pitch associated with each contrast. What emerges is evidence for just the same cross-sensory correspondences, confirming their bi-directionality. Thus, bigger visual stimuli (L. Walker et al., 2012), bigger haptic stimuli (L. Walker et al., 2012; P. Walker & L. Walker, 2012; P. Walker & Smith, 1985), darker visual stimuli (Marks, 1974; L. Walker et al., 2012; P. Walker, 2012b; L. Walker et al., 2012; P. Walker & L. Walker, 2012), more curved (less pointy) visual stimuli (L. Walker et al., 2012; P. Walker, 2012a), slower tempo musical sequences (Collier & Hubbard, 2001), and heavier (unseen) objects (P. Walker, Scallon, & Francis, 2015) have all been associated with lower pitch sounds, and their opposites with higher pitch sounds.

Finally, several studies of speeded classification also have confirmed the bi-directional nature of cross-sensory correspondences, sometimes with the same stimuli in the same task situation. For example, when concurrent visual and auditory stimuli vary independently in their 'height' (spatial elevation and pitch, respectively), equivalent congruity effects are observed whether it is the visual stimuli that are being classified for their height, or the auditory stimuli (Ben-Artzi & Marks, 1995; Evans & Treisman, 2010; Melara & O'Brien, 1987; Patching & Quinlan, 2002). The same bi-directionality is observed when tactile and auditory stimuli both vary in height (spatial elevation on a touched object and auditory pitch, respectively) (Occelli, Spence, & Zampini, 2009). Though correspondences involving 'height' might be driven by the shared verbal labeling applied to contrasting values across the two domains (i.e., *high* and *low*), this cannot be the case with similar demonstrations in which contrasting values of auditory pitch have been combined with contrasting values of visual size (Evans & Treisman, 2010), visual thickness (Evans & Treisman, 2010), and visual brightness (Marks, 1987; Melara, 1989). Neither can it be applied to equivalent bi-directional congruity effects involving vibrotactile frequency and visual surface brightness (Martino & Marks, 2000). In summary, there is very direct evidence in these studies of speeded classification for the bi-directionality of the cross-sensory correspondences between auditory pitch and each of visuo-spatial elevation, tactile elevation, vibrotactile frequency, visual size, visual thickness, and visual surface brightness.

Cross-Sensory Correspondences and Transitivity

Transitivity of implication in logic refers to a rule governing the relationships linking different material properties (also known as *material conditional*). In general, the logic of material implication is: *If A implies B, and B implies C, then A implies C.* To indicate its relevance to correspondences, the same logic might be exemplified as: *If high is bright, and bright is thin, then high will be thin*.

P. Walker and L. Walker (2012) assumed, as had others (e.g., Hornbostel 1931), that cross-sensory associations will display this type of transitivity. On this basis they predicted the existence of a correspondence between size and brightness. They reasoned that it was already known that higher pitch sounds are bi-directionally associated with both smaller and brighter things, as compared with lower pitch sounds. Though there was very little evidence available for a correspondence between size and brightness (see P. Walker & Smith, 1985), they predicted one on the basis of transitivity, reasoning that: *If brighter is higher, and higher is smaller, then brighter will be smaller.*

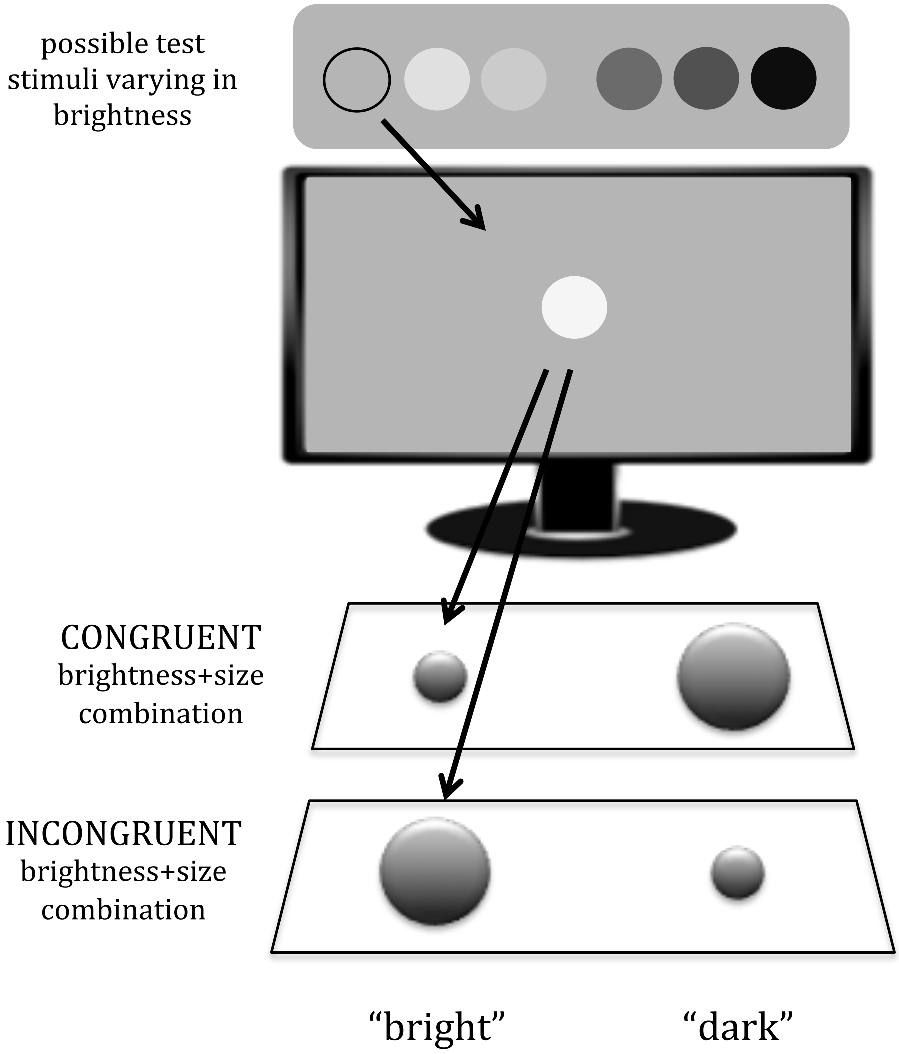

Figure 3. The speeded brightness classification task in which individual circles varying in brightness appear at the centre of a screen and participants decide as quickly as possible if each one is brighter or darker than the mid-grey background. Participants register their decision by pressing the left or right of two keys which are always hidden from view and which, as an incidental feature of the task, differ in size (see P. Walker & L. Walker, 2012).

To test their prediction, P. Walker & L. Walker (2012) used a speeded classification task in which participants were presented with individual circles at one of six levels of brightness on a mid-grey background (Figure 3). Three levels were brighter than the background, and three were darker than the background, and participants had to classify each circle as quickly as possible according to whether it was brighter or darker than the background. Participants confirmed their decision by pressing one of two hidden response keys with their left or right hand. As a task irrelevant aspect of the situation, the response keys differed in size, so that on any trial the key needing to be pressed was either the smaller or larger of the two keys. P. Walker and L. Walker observed the congruity effect they predicted on the basis of a correspondence between size and brightness, with participants classifying brighter (darker) circles more quickly when the key needing to be pressed happened to be the smaller (bigger) of the two. In light of these results, P. Walker and L. Walker confirmed their commitment to incorporating the transitivity of cross-sensory feature associations in their theoretical framework. It is worth exploring the issue of transitivity in cross-sensory feature associations in more detail.

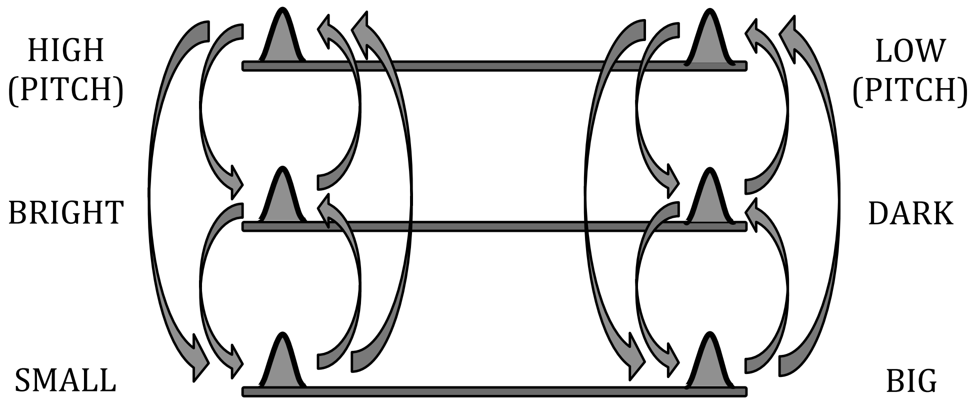

Figure 4. The assumed transitivity among bi-directional cross-sensory correspondences ensures the same mapping between pitch and size whether this mapping is direct, or is indirectly mediated through their associations with brightness.

Sampling just the three core dimensions of height, brightness, and size (any three dimensions suffice), Figure 4 illustrates the proposed functional organisation of the cross-sensory associations involved in correspondences. Depicted in this figure are the indirect associations between pitch and size that are mediated by brightness, that is, the intermediate associations from higher (lower) to brighter (darker), and then from brighter (darker) to smaller (bigger). On the understanding that cross-sensory associations are bi-directional, the same indirectly mediated associations in reverse also are depicted. Figure 4 incorporates an additional feature central to the property of transitivity, that is, the same cross-sensory feature values are also associated with each other directly.[4] For the three dimensions illustrated, smaller (bigger) and higher (lower) also activate each other in a direct (i.e., unmediated) manner. This aspect of the theoretical framework accommodates the transitivity intended by P. Walker and L. Walker (2012), in this instance the logic of transitivity appearing as: *If higher is brighter, and brighter is smaller, then higher will be smaller.* Without overall transitivity among cross-sensory correspondences, contradictory indications regarding the values of features would arise from different correspondences, leading to incoherence in the network of associations. This point is illustrated in Figure 5, where a different way of depicting transitivity among these three feature dimensions is adopted.

---

[4] To reiterate, the claim is not that some correspondences are direct and others indirect, but rather that the same correspondence can be both directly and indirectly mediated.

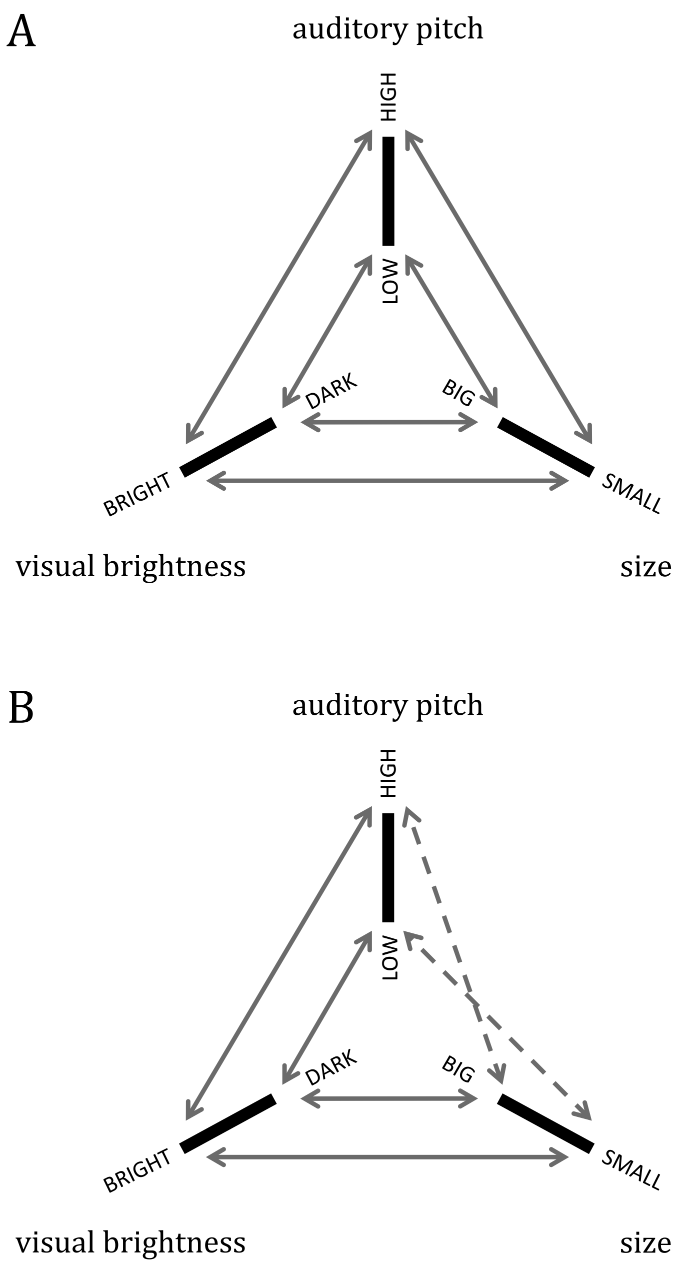

Figure 5. A: Illustrated in a different way, how the functional organisation of the cross-sensory associations between height, brightness, and size, incorporates the principle of transitivity. B: In the absence of transitivity, the direct association between height and size might contradict the indirect association between them (i.e., with bigger now mapping directly on to higher, rather than lower).

Evidence for transitivity. Evidence can be gleaned from several studies for overall transitivity among cross-sensory correspondences, including the correspondences illustrated in Figure 5A. There is ample evidence for a consistent, bi-directional correspondence between auditory pitch and visual brightness (more specifically, visual surface brightness), wherein higher levels of auditory pitch and increasing levels of surface brightness map onto each other (Collier & Hubbard, 2001; Eitan & Timmers, 2010; Marks, 1974, 1987; Martino & Marks, 1999; Melara, 1989; Mondloch & Maurer, 2004; Tarte, 1982; L. Walker et al., 2012; P. Walker, 2012a; P. Walker, Francis, & L. Walker, 2010; P. Walker & Smith, 1984). There is also evidence for a bi-directional correspondence between surface brightness and size, with higher levels of surface brightness and smaller size mapping onto each other (L. Walker et al., 2012; P. Walker & Smith, 1985; P. Walker & L. Walker, 2012). Confirming transitivity, therefore, is evidence for a direct and bi-directional correspondence between auditory pitch and size, wherein higher auditory pitch maps onto smaller size (Bien et al., 2012; Eitan & Timmers, 2010; Evans & Treisman, 2010; Gallace & Spence, 2006; Mondloch & Maurer, 2004; Parise & Spence, 2008, 2009; Tarte, 1982; L. Walker et al., 2012; P. Walker & Smith, 1984, 1985; P. Walker & L. Walker, 2012). Furthermore, confirming the amodal and conceptual nature of the features being associated, the same direct and bi-directional cross-sensory associations occur whether size is manifest as visual size (Bien et al., 2012; Evans & Treisman, 2010; Gallace & Spence, 2006; Mondloch & Maurer, 2004; Parise & Spence, 2008, 2009; L. Walker et al., 2012) or as haptic size (L. Walker et al., 2012; P. Walker & Smith, 1985; P. Walker & L. Walker, 2012). For the three core dimensions of height (auditory pitch), brightness, and size, therefore, there is evidence confirming the coherent transitivity of the associations among them (as illustrated in Figure 5A), and for the amodal and conceptual nature of the representations being linked.

There is also evidence, albeit less extensive, confirming the transitivity of the cross-sensory correspondences among other sets of three feature dimensions. Specifically, equivalent evidence emerges when auditory pitch and visual surface brightness together are considered in conjunction with other feature dimensions, including pointiness (Marks, 1987; Parise & Spence, 2009; Tarte, 1982; L. Walker et al., 2012; P. Walker, 2012a; P. Walker, Bremner et al., 2010; P. Walker & Smith, 1984; P. Walker & L. Walker, 2012), the heaviness of an unseen object (P. Walker, Scallon, & Francis, 2015; Tarte, 1982; L. Walker et al., 2012; P. Walker, 2012b; P. Walker, Francis, & L. Walker, 2010), and the vibrotactile frequency of a stimulus applied to the hand (wherein higher frequencies of vibration correspond with higher levels of auditory pitch and increasing levels of surface brightness) (Martino & Marks, 2000; Rho, Hsu, Yasar, Elmore, & Beauchamp, 2009). There is also modest evidence confirming the transitivity of cross-sensory correspondences involving the three dimensions of size, visual surface brightness, and an abstract notion of motion/speed (Collier & Hubbard, 2001; L. Walker et al., 2012).

To summarise, there is emerging evidence supporting the general principle of transitivity among cross-sensory correspondences and, therefore, theoretical frameworks predicting such transitivity, including the framework being promoted here.

Relative Coding of Stimulus Features in Cross-Sensory Correspondences

Hornbostel (1931) famously claimed to have evidence that the transitivity among cross-sensory feature associations served to link absolute values for stimulus features. He reported that when values for auditory frequency (in Hz) and surface brightness (as %

white/black on a colour wheel) were separately identified as matching the same odour, the same values were also judged to match each other when directly compared. In the logic of transitivity, it seemed that: *If an absolute value for auditory frequency matches a particular odour, and the same odour matches a specific absolute value for visual brightness, then the same absolute value for auditory frequency will match the same absolute value for visual brightness.*

The claim that cross-sensory correspondences link absolute values of features is noteworthy in three respects. First, it incorporates the unlikely notion that the whole network of cross-sensory correspondences is precisely tuned to ensure the same absolute values for a feature are indicated by any of the many indirect associative pathways in the network of cross-sensory associations (e.g., the absolute value for visual brightness that is indicated directly from auditory frequency, will also be indicated indirectly via any combination of, for example, size, weight, and sharpness as intermediate steps). Second, it assumes that the features whose values are being matched are sensory-perceptual (modality-specific) in nature, because it is only these that have absolute values. But this assumption contradicts claims that the feature dimensions underlying correspondences are amodal in nature. Third, because sensory-perceptual features are thought to map onto each other in a context-insensitive manner,[5] the claim also assumes that the same absolute feature values will show transitivity regardless of the range of feature values from which people are able to select corresponding values. However, and directly

[5] It is generally accepted that, whereas absolute feature values associate with each other in a context-insensitive manner, the (relative) values of more conceptual features associate with each other in a largely context-sensitive manner (Marks, 1987; Martino & Marks, 2001; L. Walker & P. Walker, 2015).

counter to Hornbostel's claim, there is evidence that the mapping of features in cross-sensory correspondences is context-sensitive (i.e., it is the value of a feature relative to the set of values available that gets matched to feature values in other domains).

When Cohen (1934) repeated Hornbostel's study he found that the value of auditory frequency judged to match either a specific odour, or a particular level of brightness, varied according to the range of values from which a match could be selected. That is, it was the relative position of a feature value along the range of available values that was matched to feature values in different domains, not its absolute value. More recently, Marks (1987) confirmed a substantial context-sensitive component to a correspondence-induced congruity effect involving visual surface brightness and auditory pitch. He asked participants to classify a single shape according to whether it appeared in black or white, while ignoring a concurrent task-irrelevant tone of either 220 or 360 Hz. He observed a significant interaction between pitch and brightness, with responses being relatively fast to the black shape when it was accompanied by the 220 Hz tone, rather than the 360 Hz tone, but to the white shape when it was accompanied by the 360 Hz tone, rather than the 220 Hz tone. Of particular interest, the magnitude of this interaction was reduced when two more extreme tones of 100 Hz and 800 Hz were added to the mix of incidental sounds. Marks argued that this was likely due to the fact that the 220 and 360 Hz tones were now no longer the lowest and highest pitched sounds being presented in the task, but instead had *relatively* intermediate values.

Gallace and Spence (2006) also demonstrated relative mapping in the automatic induction of a congruity effect deriving from the cross-sensory correspondence between auditory pitch and visual size. They asked participants to classify the second of two successively presented circles according to whether it was bigger or smaller than the first (the size of which was fixed). As a task-irrelevant stimulus, an auditory tone of either 300 or 4500 Hz accompanied the second circle. In some blocks of trials the incidental sound was always either the 300 Hz tone or the 4500 Hz tone, but not both. In other blocks of trials, however, the presentation of both tones was mixed, with one of the tones being selected for a trial independently of the relative size of the second circle. For neither type of trial block, therefore, was the pitch of the tone informative about the correct classification of the circle. Gallace and Spence observed a congruity effect induced by the correspondence between pitch and size only when the two tones appeared mixed within the same block of trials, and because of this they reasoned that the mapping of feature values across the two dimensions is relative in nature. That is, for the correspondence between pitch and size to induce a congruity effect, there needs to be a context provided by having multiple feature values presented in both stimulus domains.

The most compelling evidence for the relative coding of feature values in cross-sensory correspondences is provided by L. Walker and P. Walker (2015). Following up on their earlier demonstration of a size-brightness congruity effect, in which circles were classified according to their brightness using two response keys that differed in size, they show how the same circle can interact with key size as either a bright circle or a dark circle depending on the brightness of the other visual stimuli with which it appears (see Figure 6).

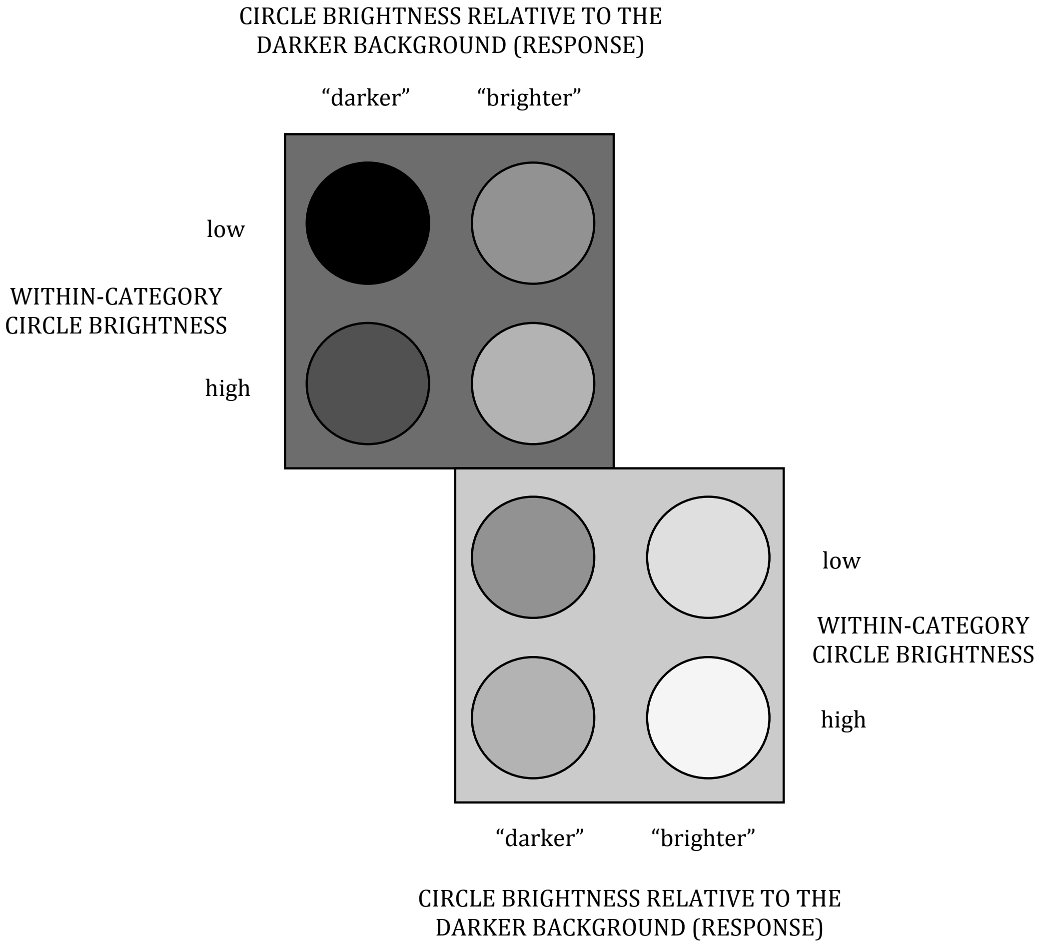

Figure 6. The different brightness levels for the six visual test stimuli and the background against which they appeared, together with an indication of the conditions these created.

In the first of two experiments, L. Walker and P. Walker (2015) arranged for three ranges of brightness level to be available for the circles, with two levels of brightness in each range. However, only brightness values from the two higher ranges, or the two lower ranges, were presented to any individual participant. Whichever two ranges of brightness were used, the brightness of the background was set to fall between them, ensuring that a circle was equally likely to be brighter or darker than the background against which it appeared. The two response keys again differed in size as a task irrelevant feature. What transpired was that, regardless of the levels of absolute brightness selected for the circles, the same size-brightness congruity effect appeared. Most compelling, however, was the observation that when circles of intermediate absolute brightness appeared against the darker background they formed a congruent relationship with the smaller of the two keys, whereas when they appeared against the brighter background they formed a congruent relationship with the bigger key.

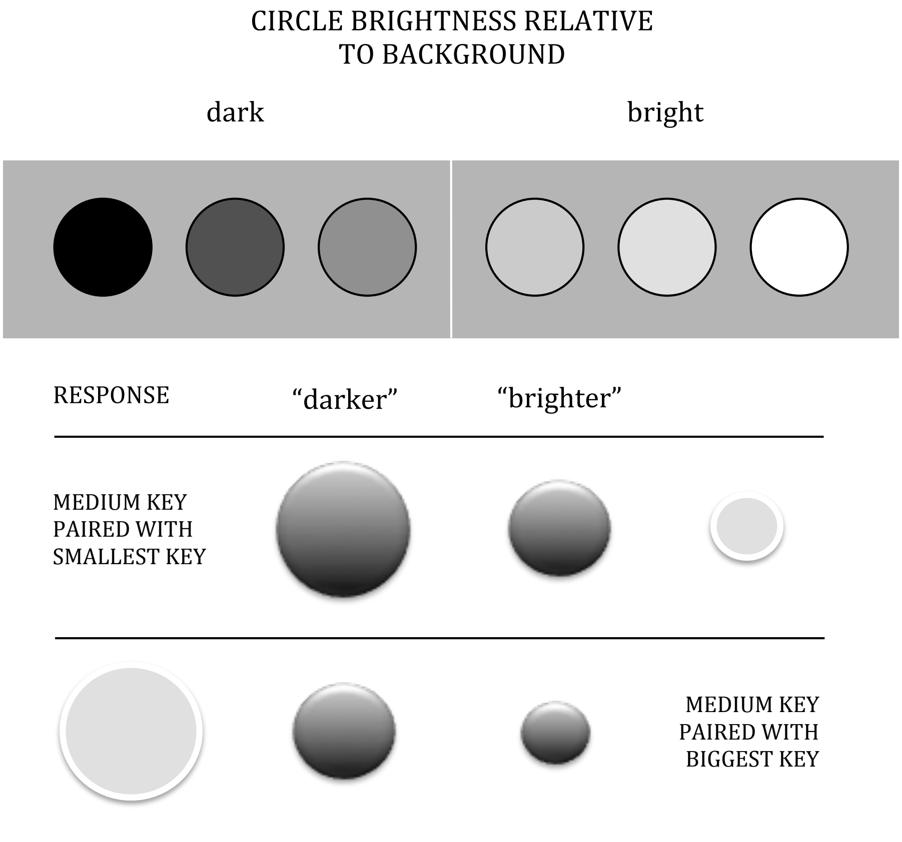

Figure 7. The different brightness levels for the six visual test stimuli that appeared against a medium grey background, together with the two alternative pairs of response keys with which participants responded.

In the second follow-up experiment the focus switched to the sizes of the keys and L. Walker and P. Walker (2015) revealed how the same response key can interact with brightness either as a small key or as a big key depending on the size of the other key with which it is paired (see Figure 7). Though only two ranges of brightness for the circles were utilised, three sizes of key were available for use in the study. However, only the bigger two keys, or the smaller two keys, were used together as the alternative response keys for a participant to use. The *relative* size of the medium response key was manipulated across separate blocks of trials by pairing it with either the smallest key, or the biggest key, from the three keys available, thereby ensuring that it was variously either the bigger key, or the smaller key, being used by participants, respectively. Confirming that the cross-sensory mapping of size to brightness can depend on the relative size of a key, rather than solely on its absolute size, an equivalent size-brightness congruity effect was observed regardless of whether the two smaller keys were used, or the two bigger keys. Specifically, regardless of their absolute size, the smaller key formed a congruent relationship with brighter circles, and the bigger key formed a congruent relationship with darker circles. Most compelling, therefore, the medium size key behaved as a small key when it was paired with the biggest of the three keys (forming a congruent relationship with brighter circles), but as a big key when, in separate blocks of trials, it was paired with the smallest of the three keys (forming a congruent relationship with darker circles).

In line with Martino & Marks' (2001) claim that the coding of stimulus features in cross-sensory correspondences is generally context-sensitive, L. Walker and P. Walker concluded from these two experiments that there can be a largely relative, context-sensitive aspect to the coding of feature values in the correspondence between size and brightness.

Specifying Feature Values Verbally Confirms a Conceptual Basis for Correspondences

Evidence that it is relative, rather than absolute feature values that are being mapped in cross-sensory correspondences helps confirm the significance of conceptual levels of representation. There are converging lines of evidence for this, some based on secondary aspects of the results from studies reviewed already (see P. Walker & L. Walker, 2012, for details), and some coming from studies in which feature values have been specified verbally. Before reviewing the latter studies, it is worth acknowledging that many of the studies exposing the nature of cross-sensory correspondences have required participants to indicate what these are by responding on verbally-specified rating scales. For example, participants have indicated if a simple tone is *very heavy in weight, quite heavy in weight, slightly heavy in weight, slightly light in weight, quite light in weight,* or *very light in weight* (see, for example, Collier & Hubbard, 2001, 2004; Tarte, 1982; L. Walker et al., 2012; P. Walker & Smith, 1984). It is clear that verbal labels such as these do not specify absolute values for the feature dimension to which they are being applied.

Martino and Marks (1999) claim that correspondence-induced effects can arise from conceptual representations established after cross-sensory features from different domains have been recoded into an abstract format common to perceptual and linguistic systems, a format they labelled *semantic* (*op. cit.,* p. 64). Though in his tutorial review of cross-sensory correspondences Spence (2011) elects to highlight three non-semantic bases for correspondences, at the same time he acknowledges that correspondences might sometimes be rooted in the semantic representation of basic stimulus features. As evidence for the latter, he points to demonstrations of cross-sensory correspondences, typically using speeded classification tasks, in which at least some elementary stimulus features values have been specified verbally (e.g., with the words *high* and *low* replacing high and low pitched tones) (see also Gallace & Spence, 2006; Martino & Marks, 1999; Melara & Marks, 1990; P. Walker, 2012a; P. Walker & Smith; 1984, 1985). The argument is that it is only because correspondences are based on amodal conceptual representations that cross-sensory filling-in can occur when feature values are specified verbally (i.e., the word *bright* and a bright visual stimulus can induce the same filling-in because they access the same concept of brightness).

P. Walker and Smith (1984, 1985), and later Melara and Marks (1990) and Gallace and Spence (2006), demonstrate correspondence-induced congruity interactions in situations where the values for one of the interacting features are specified verbally (e.g., the words *HI* and *LO* are presented either as printed text or as speech), whereas the values for the other interacting feature are specific non-verbally (e.g., the spatial elevation of the word on the computer screen is high or low, or the overall auditory pitch of the spoken word is high or low). For example, Gallace and Spence (2006) confirmed that their demonstration of a correspondence-induced congruity effect between visual size, as the feature being classified, and auditory pitch, as the task irrelevant feature, extended to a situation in which the high and low pitch sounds were replaced by the spoken words *high* and *low.* P. Walker (2012a) also observed correspondence-induced congruity effects with a mix of verbally and non-verbally presented feature values. He presented to-be-classified words inside novel outline shapes that were either angular or curved. The words referred to contrasting levels of auditory pitch, brightness, or hardness, and it was on the basis of each of these contrasts that participants classified the words. The congruity effects observed reflected underlying interactions between the concept of sharpness (realised through the varying angularity of the shape), and the concepts of elevation, brightness, and hardness. Specifically, the angularity of the outline geometric shape within which a to-be-classified word appeared interacted with the conceptual connotations of the word to yield a correspondence-induced congruity effect. For example, his participants found it easier to classify a word as referring to a high-pitched (sharp) sound when it appeared within an angular (sharp) shape (see Figure 8).

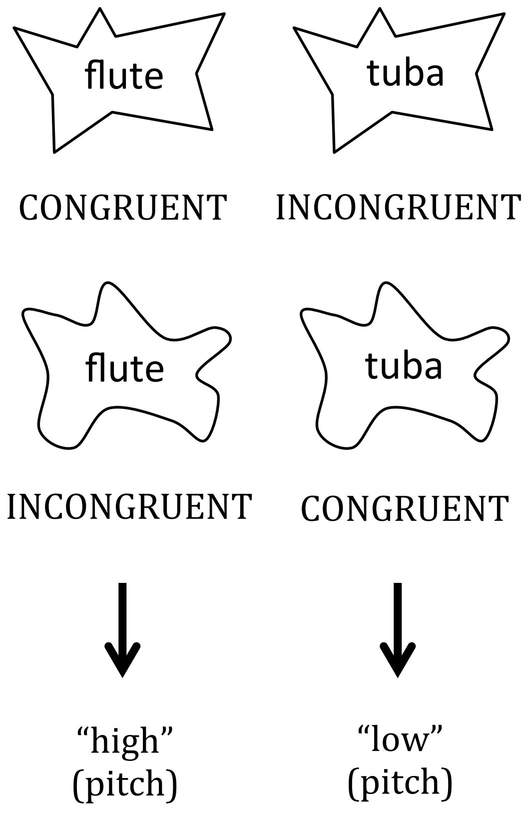

Figure 8. Congruent and incongruent combinations of a task-irrelevant novel shape, with alternative shapes contrasting in angularity-curvedness, and a to-be-classified word. Illustrated here is a situation in which words are to be classified according to whether their referents are associated with high pitch or low pitch sounds.

Finally, P. Walker, L. Walker, and Francis (2015) recently confirmed the involvement in correspondences of high level processes utilising semantic representations of the kind proposed by Martino and Marks (1999), that is, representations that are sufficiently abstract to accommodate information from perceptual and linguistic systems. They asked participants to classify the names of substances according to whether a named substance was bright (white or close to white, such as *salt* and *flour*) or dark (black or close to black, such as *ink* and *coal*). In line with the size-brightness congruity effect they had previously observed with non-verbal visual stimuli, they observed the size of the key needing to be pressed to interact with the brightness of the substance being classified, that is, participants responded more easily when bright (dark) classifications were registered with the smaller (bigger) of two response keys, rather than the reverse. It was the results of a second experiment that confirmed the high level nature of the processes and representations behind this correspondence-induced congruity effect. In this experiment participants had to classify the same substance names for the edibility of the named substances, rather than for their brightness. Because the named substances had been carefully selected to ensure that all the brighter substances were edible, and all the darker substances were inedible, participants responded to them in exactly the same way in both experiments, despite the different semantic basis for doing so. According to the theoretical framework being promoted here, P. Walker et al. expected the size-brightness congruity effect not to be in evidence in the second experiment. The reasoning behind this was that, notwithstanding the lower levels of processing being identical in both experiments, the classification decision now had to be made on the basis of semantic information separate from any of the conceptual feature dimensions underpinning correspondences. Because of this, there would be no reason for the classification decision to be influenced by any processing relating to these feature dimensions. As expected, the size-brightness congruity effect failed to appear in the second of their experiments.

Correspondences in Action and in the Communication of Ideas

Bodily actions. The motivation behind P. Walker and L. Walker's (2012) demonstration of a size-brightness congruity effect was largely to test a prediction derived from the proposed transitivity of correspondences, but also to show how correspondences can embrace a wide range of stimulus features, including the size of an object as conveyed haptically. However, a different way of viewing the study suggests that an additional feature of the task situation might have theoretical and practical significance, that is, size as a feature of participants' behavioural responses to the brightness of the visual stimuli, and not just a feature of the response key. Perhaps it was size as an aspect of the hand configuration adopted during the communication of a decision that contributed to the correspondence-induced congruity effect (i.e., hand configurations capable of grasping small and big objects). Specifically, hand actions incorporating an element of relative smallness (bigness) would be congruent with the relative brightness (darkness) of the visual stimuli to which they are a response. Perhaps this kind of congruity can influence the fluency with which the brightness (darkness) of a visual stimulus is communicated through a hand action.

The potential for correspondence-based congruity effects to extend to features of peoples' actions is made clear in several studies. For example, Eitan and Granot (2006) and Eitan and Tubul (2010) asked participants to listen to short melodic figures and visualise an animated character moving to the music. Movement in the vertical direction followed the pitch contour of the melodic figure, with relatively more downward bodily movements when pitch descended rather than ascended. Furthermore, movement in the lateral direction became relatively faster as pitch ascended, compared to when it descended, echoing the core correspondence between height and speed. In another study, Kussner, Tidhar, Prior and Leech-Wilkinson (2014) played continually sounding pure tones, each lasting 8 s, to participants. The tones varied internally in their frequency, amplitude, and tempo profiles, but in a simple way: In the first and second 4 s period of a tone, each of these features either remained constant, or changed in opposite directions (e.g., pitch rise and then fall, amplitude increase and then decrease, tempo increase and then decrease). Participants were required to represent each sound, as if to communicate the nature of the sound to someone else, by moving their hand as they held a movement sensor. Kussner et al. observed changes in the pitch of the sound to be represented as changes in the vertical elevation of the hand, in accordance with the correspondence between auditory pitch and spatial elevation. Though changes in loudness also were represented as vertical motion (louder being higher), when these changes were contradicted by changes in the direction of pitch it was the latter that dominated the vertical elevation of the hand.[6]

Vocalisations. Similar evidence is available concerning vocalisations as actions. Parise and Pavani (2011) explored the correspondences between visual sharpness and auditory pitch, and between visual size and auditory pitch, but with pitch as a feature of

---

[6] Though pitch did not appear to be represented by the speed of hand movement, in apparent contradiction of the correspondence between the two features (e.g., Walker & Smith, 1984), it is worth noting that the rate of change in the pitch and amplitude of the tones would be expected to have a much stronger influence on the judged speed of the tone sequence and, therefore, on the speed of the hand movements.

participants' vocalisations to stimuli. They presented participants with simple visual stimuli varying along individual feature dimensions, such as sharpness and size. Participants were not required to classify the stimuli, but instead simply had to respond to the presentation of each one by immediately generating the same simple vocalisation, namely, say the letter /a/ (as in "ah"). In line with the cross-sensory correspondences observed elsewhere, Parise and Pavani observed participants to raise the fundamental frequency of their utterances when a stimulus happened to be relatively sharp or relatively small.

In a related study, Dolscheid, Shayan, Majid and Casasanto (2013) asked native adult speakers of Dutch (a language using a spatial height metaphor for pitch) to reproduce a simple auditory tone presented at one of several levels of pitch. At the same time as listening to the tone, participants watched a screen on which a single task-irrelevant horizontal line appeared. The spatial height at which the line appeared on the screen was varied independently of the pitch of the tone. Despite the appearance of the line being task-irrelevant, participants were observed to raise the fundamental frequency $F_{(0)}$ of their vocal reproductions according to the spatial elevation of the line (the higher in space the line, the higher in pitch their sung reproduction). In addition, after priming their Dutch speaking participants to think about how auditory pitch and visual thinness-thickness might be associated, Dolscheid et al. observed the same correspondence-induced adaptations in participants' vocalisations in response to seeing single vertical lines varying in thinness-thickness (the thicker the line, the lower in pitch their sung reproduction). Adult speakers of Farsi, a language using a thinness-thickness metaphor for pitch, did not need to be primed for the pitch of their vocalisations to adapt to the thinness-thickness of the task-irrelevant line.

Actions in the communication of ideas. In studies where participants have rated individual stimuli for their cross-sensory features (e.g., indicating on a verbally defined scale that a simple tone is *very bright*), we can readily think of their responses as being a means of communicating an idea they have to the experimenter. Indeed, they could just as easily have provided a spoken response for the experimenter, as if in a conversation. Even in the speeded classification task it is possible to regard participants' key pressing as a means of communicating to the experimenter what they think is the nature of the test stimulus (e.g., whether they regard it as being bright or dark). Again it would be possible to have them vocalise their idea, and record the latency to the onset of their vocalisation as the dependent measure. Whatever the details of the response, the communicative nature of the participant's task in speeded classification can seem almost as transparent as in the rating task. Highlighting the communicative element of the tasks employed in studies of cross-sensory correspondences emphasises the potential relevance of cross-sensory correspondences to communication more generally, including the communication of ideas through music.

A relatively direct demonstration of the contribution cross-sensory correspondences can make to the communication of ideas is embedded in a study concerned with the semantics of prosody. Nygaard, Herold and Namy (2009) asked adults to produce infant-directed speech that, on its own, would serve to draw a child's attention to one of two items with contrasting values on a single conceptual dimension (e.g., big/small, happy/sad, hot/cold). For example, when two pictured items contrasting in relative size were viewed simultaneously, an adult might have to imagine directing a young child's attention either to the bigger item, or to the smaller item (i.e., communicate the notion of bigness or smallness). They were restricted to using a single prescribed sentence frame that included the same verbal label for whatever feature value was being targeted (e.g., "Can you get me the *blicket* one?" with *blicket* referring sometimes to the bigger item, sometimes to the smaller item). Interest focussed on how adults might moderate the basic acoustic features of their vocalisations (the prosody, or melody in their voice) to better communicate the feature value identifying the target item. Though the researchers did not frame their interests in terms of correspondences, it is interesting to see how these might explain some relevant aspects of their results. Consistent with the core correspondences identified elsewhere, Nygaard et al. observed that where an item to which a child's attention was to be drawn was relatively small, adults raised the fundamental frequency of their voice while uttering the novel label (e.g., *blicket*) for smallness. At the same time, they reduced both the loudness and duration of their vocalisation. It seems, therefore, that the idea of relative smallness (bigness) was being conveyed through the cross-sensory correspondences between size and each of higher (lower), quieter (louder), and faster (slower). The same outcome was observed with regard to the acoustic features of each vocalisation overall (i.e., with sentence-level acoustic features). Finally, and importantly, Nygaard et al. went on to confirm that adults listening to the infant-directed speech that had been produced in this way reliably picked up the acoustic cues identifying the target feature value (i.e., they were able to identify to which member of a contrasting pair of items a vocalisation was directed). In other words, use of the appropriate correspondence did facilitate communication of the idea of a relative value for an elementary stimulus feature.

Cross-Sensory Correspondences and Musical Stimuli

Reflecting the nature of the research being reviewed, much of the discussion has concerned the elementary features of simple sounds and not, for example, the features of complex sounds, such as excerpts from real music. Nevertheless, discussion has sometimes concerned the elementary features of more complex sounds, including speech, specially created sequences of musical notes, and excerpts from real music. The Nygaard et al. (2009) study, for example, focussed on correspondences involving the fundamental frequency (F_0), amplitude, and temporal duration of spoken sentences. In other studies, musical sequences were used to explore correspondences (Collier & Hubbard, 2001, 2004; Eitan & Timmers, 2010; Karwoski, Odbert, & Osgood, 1942; Kussner & Leech-Wilkinson, 2013), and it was observing the auditory-visual feature associations evident in peoples' drawings of musical excerpts that prompted Karwoski et al. to propose the existence of cross-sensory correspondences. They observed that music in higher pitch registers, just like high pitch single tones, tended to be represented by visual forms that were more angular, brighter, smaller, spatially higher, thinner, and with more movement implied than was music in lower pitch registers. Similarly, music judged to be relatively heavy tended to be represented by visual forms that were more curved, darker, larger, lower in space, thicker, and with less implication of movement, compared to less heavy music, while faster tempo music was represented by smaller, spatially higher, thinner, more angular, brighter, and with movement more likely to be implied, compared to slower tempo music.[7] More recently, Kussner & Leech-Wilkinson (2013) confirmed that

---

[7] After indicating what visual features best represented contrasts in musical sequences (e.g., that bright visual forms best represent music in a higher register), Karwoski, Odbert, and Osgood (1942) asked their participants to indicate what they thought was the general basis for their associations. Karwoski et al. comment that:

> *There was little consistency in their reports. Some subjects thought of past experiences, others applied the opposites to real objects, some thought of personal characteristics, yet others allied words in terms of their pleasantness or*

when people draw musical excerpts that are relatively high in pitch, they tend to apply less downward pressure, thereby producing thinner lines. They do the same when drawing musical excerpts that are relatively quiet.

Collier and Hubbard (2001, 2004) played musical scales to participants in one of several keys, in either ascending or descending pitch, and at either a fast or slow tempo. Their participants judged sequences in a higher register to be brighter and faster than sequences in a lower register, as they did sequences ascending rather than descending in pitch. Participants also judged sequences played at a relatively fast tempo to be brighter than those played at a relatively slow tempo. Finally, Eitan and Timmers (2010) observed the same cross-sensory feature associations when segments of music contrasting in pitch register were sampled from the second movement of *Beethoven's piano sonata, opus 111.* These sequences were rated on verbal scales anchored by pairs of antonyms and, like simple tones contrasting in acoustic frequency, music set at a higher pitch register was judged to be more active, brighter, faster, lighter in weight, sharper, smaller, and thinner than segments set at a lower pitch register.

The correspondence between auditory frequency and speed has been confirmed in three recent studies using musical stimuli (Boltz, 2011; Broze & Huron, 2013; Tamir-

---

> *unpleasantness, and so forth. The high agreement in response despite various conscious attitudes toward the items strongly suggests that these words are related in similar ways in many different types of experience, activities, objects, and situations in life. ... This diversity of explanations accompanying agreement in response suggests that a basic, common frame of reference may actually be operating, although it is not verbalized. op. cit.,* pp. 209-210.

This conclusion, that the feature dimensions underlying correspondences are fundamental by virtue of being amodal and conceptual in nature, is incorporated in the elaboration of Karwoski et al.'s theoretical stance being promoted in the present paper.

Ostrover & Eitan, 2015). Broze and Huron (2013) sampled Western musical scores and observed that parts for lower musical voices, and for instruments with lower pitch range (*tessitura*), incorporate fewer notes per part, or per unit time, than parts for higher musical voices and instruments with higher pitch range. They also observed that Baroque ornaments are more likely to appear in musical parts in a higher register. In addition, Boltz (2011) created music-like sequences of notes that varied independently in the octave pitch (high vs. low) and timbre (bright vs. dull) in which they were played, but for which the tempo was held constant. The sequencies were judged to be faster by listeners when their pitch was high, rather than low, and when their timbre was bright, rather than dull. And again, melodies were judged to have relatively faster tempi when their pitch contour ascended, rather than descended. Finally, Tamir-Ostrover and Eitan (2015) asked participants to adjust the tempi of melodic sequences set at high or low pitch registers until they felt the result was satisfactory. Melodic sequences at higher registers were set at faster tempi than were melodic sequences set at lower registers.

In summary, there is encouraging evidence that the core cross-sensory correspondences revealed with simple sounds continue to be relevant in relation to the complex sound sequences characteristic of musical compositions.

Cross-Sensory Correspondences in Music

It remains to enquire where opportunities might arise for cross-sensory correspondences to be exploited in the writing, performance, and appreciation of music. If music is regarded as an enterprise in which the composer and performer together attempt to communicate certain ideas (e.g., images) to the listener, we might ask how cross-sensory correspondences can be exploited to facilitate the listener's appreciation of the ideas behind the music (just as they facilitated participants' appreciation of certain characteristics of the test stimuli in the experimental studies reviewed above). The Nygaard et al. (2009) study illustrates this point very well: The 'melody' the adult speakers introduced to their speech served to facilitate the communication of the idea they were trying to communicate. Crucially in the context of the present review, addressing the exploitation of cross-sensory correspondences in music serves to confirm the significance of the bi-directionality and transitivity of correspondences, and of their capacity to embrace the elementary features of people's actions.

It has been noted that sounds can represent basic non-auditory aspects of things by virtue of being able to share their cross-sensory features. For example, the correspondence between brightness and pitch enables musical sequences in a relatively high register to represent visually bright things. At the risk of over-simplifying matters, but for the sake of the argument, Figure 9 illustrates a situation where a composer endeavours to convey to a listener of their music the brightness of the visual forms typical of a firework display.

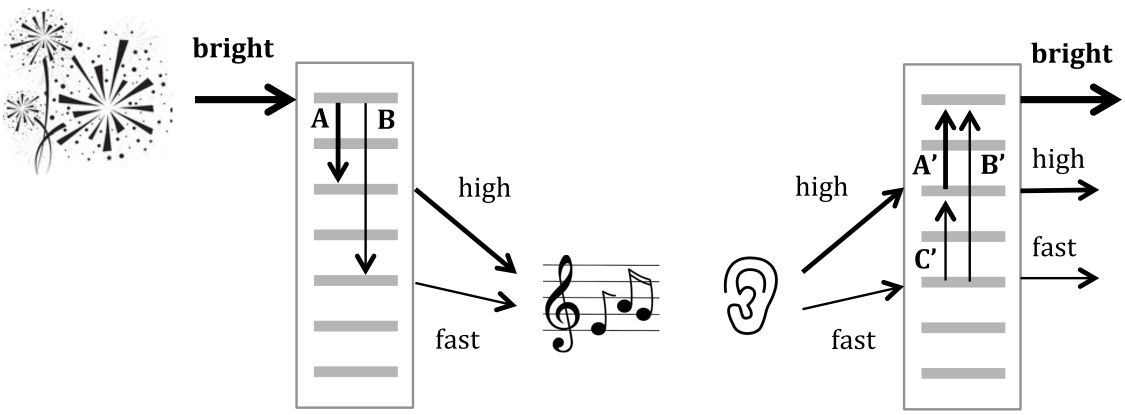

Figure 9. How cross-sensory correspondences are able to contribute to the composition and appreciation of music (see text for explanation).

The composer elects to exploit the association between visual brightness and higher pitch sounds (A in Figure 9) by choosing a higher pitch register for the most salient melodic lines in the composition. A listener to the music hears the relatively high pitch nature of these lines and immediately gets a sense of height. Critical in relation to the composer's intention, however, the correspondence between height and brightness (A' in Figure 9) ensures the listener also gets the sense of brightness intended by the composer. As straightforward as this sequence of events might seem, its success hinges on the bi-directional nature of cross-sensory correspondences. For the composer, the brightness-height correspondence takes us from brightness to height, but for the listener the same correspondence needs to take us in the reverse direction, from height to brightness (A' in Figure 9).

The composer also exploits, as a secondary tactic, the correspondence between brightness and speed (B in Figure 9), and so arranges for the relevant melodic lines to have relatively fast tempi. The relatively fast tempi are registered directly by the listener, and the correspondence between brightness and speed takes us, again in reverse direction, from speed to brightness (B' in Figure 9). But speed as a cross-sensory feature also enters into correspondence with height (C' in Figure 9), and the listener will be open to having the brightness of their imagery influenced by this correspondence also. It is here that the transitivity of correspondences becomes important, because it helps ensure that the sense of speed felt by the listener will reinforce the same sense of height that is being induced by pitch (i.e., with bi-directionality assumed, the logic of transitivity in this instance is: *If bright is high, and bright is fast, then fast will be high*). Because of this, the sense of speed induced by the tempi of the music will support, rather than contradict, the impact that height (induced by pitch) is having on the sense of brightness being induced in the listener. In this way, transitivity ensures that the sense of brightness induced by the relatively high pitch of the melodic lines is made even more salient (C' in Figure 9). What is achieved in the end is a coherent reinstatement in the listener of the full multi-modal experience of the firework display as felt by the composer, but with the brightness of the fireworks made especially salient.

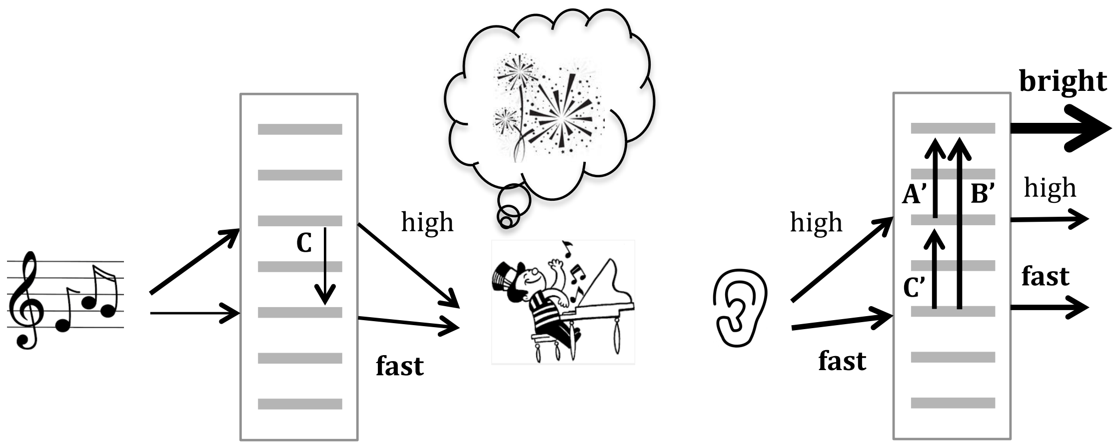

Figure 10. How cross-sensory correspondences are able to contribute to the performance of music (see text for explanation).

Aspects of musical performance provide additional opportunity for correspondences to contribute positively to the composer's aim (see Figure 10). Though the primary focus for the performer is the score, they are also likely to have knowledge of the composer's desire to capture and communicate notions of brightness, knowledge that will sit alongside their own experience of firework displays. Free to add something to the music, such knowledge will prime the performer to be especially sensitive to both the relatively high pitch of the important melodic lines and the correspondence linking this to speed (see Boltz, 2011; Broze & Huron, 2013; Tamir-Ostrover & Eitan, 2015). This should, perhaps, induce them to further increase the tempi of the melodic lines, thereby enhancing the communication of brightness, just as the adults in Nygaard et al.'s (2009) study speeded up their speech to better convey notions of smallness to a listener. Through the correspondences between speed and brightness (B' in Figure 10), and speed and height (C' in Figure 10), the additional increase in tempi introduced by the performance adds to the sense of brightness experienced by the listener based on the score alone.

Though the visual brightness typical of a firework display has been used to make these points, it is clear that other elementary features could have served the same purpose equally well. Asking how cross-sensory correspondences might be used to help music convey the thickness of the aroma of roast coffee, the heaviness and warmth of a perfume, the softness of cotton wool, or the brightness and sharpness of a toothpaste, would all provide good scenarios. Indeed, the exploitation of cross-sensory correspondences in music should be especially apparent in the music of sensory marketing (e.g., the music of TV commercials for products with distinctive non-auditory features, such as paper that is distinctively soft to touch).

Continuing to reflect on the nature of firework displays reveals a final point worth making in this review of cross-sensory correspondences. A firework display is a compound stimulus replete with features that are congruent with each other. As a visual event it incorporates brightness, smallness, and sharpness, along with fast and ascending movement at high spatial elevation. Indeed, the fast movement of the small forms is normally accompanied by high pitch sounds. Perhaps this internal congruity adds to the aesthetic appeal of a firework display. Put more generally, perhaps people will have a fundamental tendency to prefer compound stimuli that incorporate cross-sensory congruity, rather than cross-sensory incongruity. Most pertinent in the present context is the possibility that a preference for cross-sensory congruity will extend to music, in which case it would be no coincidence that faster tempi tend to be assigned to melodies at a higher pitch register (just as participants in Tamir-Ostrover & Eitan's, 2015, study found such assignment to produce a more satisfactory result). Not only will these two features work together to reinforce the ideas being communicated by the music, but the congruity between them will itself be received positively by the listener. The influence of internal cross-sensory congruity on the aesthetic appeal of stimuli, including music, deserves to be explored.[8]

[8] This kind of internal congruity within a compound stimulus will not be the only factor influencing a person's preference for a stimulus. There might even be occasions where people express a preference for being challenged by incongruity, perhaps because of its novelty.

References

Ben-Artzi, E. & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception and Psychophysics, 57,* 1151-1162.

Bernstein, I. H. & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology, 87,* 241-247.

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synesthesia. *NeuroImage, 59,* 663-672.

Boltz, M. (2011). Illusory tempo changes due to musical characteristics. *Music Perception, 28,* 367-386.

Bomba, P. C. & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology, 35,* 294-328.

Broze, Y. & Huron, D. (2013). Is higher music faster? Pitch-speed relationships in Western compositions. *Music Perception, 31,* 19-31.

Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, *41*, 339-353.

Chiou, R., Stelter, M. & Rich, A. N. (2013). Beyond colour perception: Auditory-visual synaesthesia induces experiences of geometric objects in specific locations. *Cortex, 49,* 1750-1763.

Cohen, N. E. (1934). Equivalence of brightnesses across modalities. *The American Journal of Psychology, 46,* 117-119.

Collier, W. G., & Hubbard, T. L. (2001). Judgements of happiness, brightness, speed, and tempo change of auditory stimuli varying in pitch and tempo. *Psychomusicology*, *17*, 36-55.

Collier, W. G., & Hubbard, T. L. (2004). Musical scales and brightness evaluations: Effects of pitch, direction, and scale mode. *Musicae Scientiae, VIII,* 151-173.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science, 24,* 613-621.

Eitan, Z. & Granot, R. Y. (2006). How music moves: Musical parameters and listener's images of motion. *Music Perception, 23,* 221-247.

Eitan, Z. & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition, 114,* 405-422.

Eitan, Z. & Tubul, N. (2010). Musical parameters and children's images of motion. *Musicae Scientiae, 14,* 89-111.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*, 1-12.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*, 1191-1203.

Hornbostel, E. M. von. (1931). Über Geruschshelligkeit. *Pfugers Archiv fur die Gesamte Psysiologie, 227,* 517-538.

Karwoski, T. F. & Odbert, H. S. (1938). Color-music. *Psychological Monographs, 50(2),* 1-60.

Karwoski, T. F., Odbert, H. S., & Osgood, C. E. (1942). Studies in synesthetic thinking: II. The role of form in visual responses to music. *The Journal of General Psychology, 26,* 199-222.

Kussner, M. B. & Leech-Wilkinson, D. (2013). Investigating the influence of musical training on cross-modal correspondences and sensorimotor skills in a real-time drawing task. *Psychology of Music, 42, 448-469.*

Kussner, M. B., Tidhar, D., Prior, H. M., & Leech-Wilkinson, D. (2014). Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real time. *Frontiers in Psychology, 5.* doi 10.3389/fpsyg.2014.00789

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *American Journal of Psychology*, *87*, 173-188.

Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin, 82,* 303-331.

Marks, L. E. (1978). *The Unity of the Senses: Interrelations among the modalities.* New York: Academic Press.

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.

Marks, L. E. (2004). Cross-modal interactions in speeded classification. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 85-105). Cambridge: MIT Press.

Martino, G. & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception, 28,* 903-923.

Martino, G., & Marks, L. E. (2000). Cross-modal interaction between vision and touch: The role of synesthetic correspondence. *Perception, 29,* 745-754.

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science*, *10*, 61-65.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 69-79.

Melara, R. D. & Marks, L. E. (1990) Processes underlying dimensional interactions: Correspondences between linguistic and non-linguistic dimensions. *Memory and Cognition, 18,* 477-495.

Melara, R. D. & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 116,* 323-336.

Mondloch, C. J. & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience, 4,* 133-136.

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science, 33,* 127-146.

Occelli, V., Spence, C., & Zampini, M. (2009). Compatibility effects between sound frequency and tactile elevation. *NeuroReport*, *20*, 793-797.

Parise, C. V. & Pavani, P. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research, 214,* 373-380.

Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters*, *442*, 257-261.

Parise, C. V., & Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE*, *4*, doi:10.1371/journal.pone.0005664

Patching, G. R. & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 755-775.

Perrott, D. R., Musicant, A., & Schwethelm, B. (1980). The expanding-image effect: The concept of tonal volume revisited. *Journal of Auditory Research, 20,* 43-55.

Ro, T., Hsu, J., Yasar, N. E., Elmore, L. C., & Beauchamp, M. S. (2009). Sound enhances touch perception. *Experimental Brain Research, 195,* 135-143.

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition, 99,* 113-129.

Sadaghiani, S., Maier, J. X., & Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory directions signals have distinct influences on visual motion processing. *The Journal of Neuroscience, 29,* 6490-6499.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*, 971-995.

Tamir-Ostrover, H. & Eitan, Z. (2015). Higher is faster: Pitch register and tempo preferences. *Music Perception,* in press.

Tarte, R. D. (1982). The relationship between monosyllables and pure tones: An investigation of phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior, 21,* 352-360.

Walker, L., Walker, P., & Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception, 41,* 1186-1192.

Walker, L. & Walker, P. (2015). Cross-sensory mapping of feature values in the size-brightness correspondence can be more relative than absolute. *Journal of Experimental Psychology: Human Perception and Performance,* in press.

Walker, P. (2012a). Cross-sensory correspondences and crosstalk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention, Perception, & Psychophysics*, *74*, 1792-1809.

Walker, P. (2012b). Cross-sensory correspondences and naïve conceptions of natural phenomena. *Perception, 41,* 620-622.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, *21*, 21-25.

Walker, P., Francis, B. J., & Walker, L. (2010). The brightness-weight illusion: Darker objects look heavier but feel lighter. *Experimental Psychology, 57,* 462-469.

Walker, P., Scallon, G., & Francis, B. (2015). Heaviness is dark and low-pitched. *Submitted.*

Walker, P. & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception, 13,* 75-81.

Walker, P. & Smith, S. (1985). Stroop interference based on the multimodal correlates of haptic size and auditory pitch. *Perception, 14,* 729-736.

Walker, P., & Walker, L. (2012). Size-brightness correspondence: Crosstalk and congruity among dimensions of connotative meaning. *Attention, Perception, & Psychophysics*, *74*, 1226-1240.

Walker, P., Walker, L., & Francis, B (2015). The size-brightness correspondence: Evidence for crosstalk among aligned conceptual feature dimensions. *Attention, Perception, & Psychophysics*, *77,* 2694-2710.

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex, 42,* 264-280.