RUNNING HEAD: ARE CROSSMODAL CORRESPONDENCES RELATIVE OR ABSOLUTE?

On the relative nature of (pitch-based) crossmodal correspondences

Prof. Charles Spence

Crossmodal Research Laboratory

Oxford University

RESUBMMITED TO: *MULTISENSORY RESEARCH*

WORD COUNT: 12,750

DATE: FEBRUARY, 2019

CORRESPONDENCE TO: Charles Spence, Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, Anna Watts Building, Oxford, OX2 6GG, U.K. email: charles.spence@psy.ox.ac.uk

ABSTRACT

This review deals with the question of the relative versus absolute nature of crossmodal correspondences, with a specific focus on those correspondences involving the auditory dimension of pitch. Crossmodal correspondences have been defined as the often-surprising crossmodal associations that people experience between features, attributes, or dimensions of experience in different sensory modalities, when either physically present, or else merely imagined. In the literature, crossmodal correspondences have often been contrasted with synaesthesia in that the former are frequently said to be relative phenomena (e.g., it is the higher-pitched of two sounds that is matched with the smaller of two visual stimuli, say, rather than there being a specific one-to-one crossmodal mapping between a particular pitch of sound and size of object). By contrast, in the case of synaesthesia, the idiosyncratic mapping between inducer and concurrent tends to be absolute (e.g., it is a particular sonic inducer that elicits a specific colour concurrent). However, a closer analysis of the literature soon reveals that the distinction between relative and absolute in the case of crossmodal correspondences may not be as clear-cut as some commentators would have us believe. Furthermore, it is important to note that the relative vs. absolute question may receive different answers depending on the particular (class of) correspondence under empirical investigation.

KEYWORDS: CROSSMODAL CORRESPONDENCES; RELATIVE; ABSOLUTE; PITCH; SYNAESTHESIA.

Introduction

The last few years have seen an explosion of scientific interest in the field of crossmodal correspondences (see Marks, 2004; Spence, 2011; Spence & Sathian, in press; Walker, 2016, for reviews). Crossmodal correspondences have been defined as the sometimes surprising associations that people experience between seemingly unrelated perceptual stimuli, attributes, or dimensions in different sensory modalities, no matter whether physically present, or else merely imagined (see Spence, 2011). For whatever reason, the auditory dimension of pitch turns out to have been one of the most widely studied in the growing body of crossmodal correspondences research published to date (see Deroy, Fernández-Prieto, Navarra, & Spence, 2018; Spence, 2011, for reviews).[1] The question to be addressed in this review concerns the extent to which such pitch-based correspondences are relative or absolute in nature.

Over the years, auditory pitch[2] has been shown to correspond with a wide range of different stimulus dimensions, including elevation (assessed by measuring people's responses to auditory, visual, and/or tactile stimuli; e.g., Ben-Artzi & Marks, 1995; Bernstein & Edelstein, 1971; Braaten, 1993; Carnevale & Harris, 2016; Chiou & Rich, 2012, 2015; Dolscheid, Hunnius, Casasanto, & Majid, 2014; Getz & Kubovy, 2018; Jamal, Lacey, Nygaard, & Sathian, 2017; Lidji, Kolinsky, Lochy, & Morais, 2007; McCormick, Lacey, Stilla, Nygaard, & Sathian, 2018; Miller, 1991; Mudd, 1963; Occelli, Spence, & Zampini, 2009; Pedley & Harper, 1959; Pratt, 1930; Roffler & Butler, 1968; Stumpf, 1883; Trimble, 1934; Widmann, Kujala, Tervaniemi, Kujala, & Schröger, 2004; see also Sonnadara, Gonzalez, Hansen, Elliott, & Lyons, 2009), size (e.g., Bien, ten Oever, Goebel, & Sack, 2012; Evans & Treisman, 2010; Fernández-Prieto, Navarra, & Pons, 2015; Gallace & Spence, 2006; Getz & Kubovy, 2018; Jonas, Spiller, & Hibbard, 2017; Melara & O'Brien, 1987; Mondloch & Maurer, 2004; Parise & Spence, 2009, 2012; Walker & Smith, 1985; Walker, Walker, & Francis, 2012; see also Lowe & Haws, 2017), visual lightness (Brunel, Carvalho, & Goldstone, 2015; Hubbard, 1996;

[1] One can only speculate as to the importance of research on (e.g., the connotative meaning of) music to this tendency (see Eitan, 2017; Spence, submitted; Walker, 2016, for reviews of the literature on crossmodal correspondences specifically related to music).

[2] Here one might pause to consider the assertion, unquestioningly made in much of the literature in this area, that these correspondences are between pitch and whatever other attribute or dimension in another sensory modality that happens to be the object of study. Presumably, an alternative possibility might be to suggest that it is the frequency of the sound that matters. For, as Stevens and Volkmann (1939) noted long ago, while the two dimensions, one perceptual (pitch), the other physical (frequency), are undoubtedly closely related, they are by no means identical. However, having raised this concern, it is also true to say that a broad literature appears to support the claim that it is the perceived quality of the stimulus, rather than its physical properties, that matter most as far as understanding/predicting the crossmodal correspondences are concerned.

Marks, 1974, 1975, 1987; Marks, Hammeal, & Bornstein, 1987; Martino & Marks, 1999; Melara, 1989a; Wicker, 1968), visual brightness (Getz & Kubovy, 2018; Klapetek, Ngo, & Spence, 2012; Ludwig, Adachi, & Matzuzawa, 2011; Marks, 1987; Mondloch & Maurer, 2004; Wicker, 1968), visual hue (e.g., Caivano, 1994; Goethe, 1810; Melara, 1989a, b; Melara & O'Brien, 1987; Newton, 1730/1952), visual shape/angularity (Getz & Kubovy, 2018; Marks, 1987; O'Boyle & Tarte, 1980; Parise & Spence, 2012; see also Parise & Spence, 2009), spatial frequency (Evans & Treisman, 2010; Getz & Kubovy, 2018; see also Guzman-Martinez, Ortega, Grabowecky, Mossbridge, & Suzuki, 2012; and Spence, 2011, 2018, for reviews). The auditory dimension of pitch has also been found to share a number of other crossmodal correspondences with the chemical senses, e.g., with olfactory stimuli (e.g., Belkin, Martin, Kemp, & Gilbert, 1997; Crisinel & Spence, 2010b; see Deroy, Crisinel, & Spence, 2013, for a review), with gustatory stimuli (Crisinel & Spence, 2010a; see Knöferle & Spence, 2012, for a review), as well as with a variety of flavourful stimuli (e.g., Holt-Hansen, 1968, 1976; Rudmin & Cappelli, 1983).

Crossmodal correspondences involving auditory pitch have been demonstrated to influence people's behaviour across a wide range of behavioural tasks (and corresponding dimensions), including (probably most frequently studied) speeded classification (e.g., Gallace & Spence, 2006; Getz & Kubovy, 2018; Jamal et al., 2017; Melara & O'Brien, 1987; see Marks, 2004, for a review), variants of the Implicit Association Test (e.g., Lacey, Martinez, McCormick, & Sathian, 2016; Occelli et al., 2009; Parise & Spence, 2012), the redundant targets paradigm (Miller, 1991; see also Makovac & Gerbino, 2010), the pip-and-pop visual search task (Klapetek et al., 2012), as well as a range of other behavioural tasks, both speeded (e.g., Chiou & Rich, 2012, 2015; Pitteri, Marchetti, Priftis, & Grassi, 2017; Rusconi, Kwana, Giordano, Umiltà, & Butterworth, 2006; Stekelenburg & Keetels, 2016; Walker & Smith, 1984, 1985, 1986) and unspeeded (e.g., Marks, Ben-Artzi, & Lakatos, 2003; Orchard-Mills, Van der Burg, & Alais, 2016; Parise & Spence, 2008, 2009).

Accounting for the correspondences

Several classes of explanation have been put forward over the years in order to try to account for the existence of crossmodal correspondences. Importantly, all (or, at the very least, several) of these explanations may well play some role in helping to account for those correspondences involving auditory pitch (e.g., Sadaghiani, Maier, & Noppeney, 2009; see Spence, 2011, for a review). In particular, the four main classes of explanation that are frequently mentioned in the literature are the semantic/linguistic (Martino & Marks, 1999), the statistical (Spence, 2011), the structural (e.g., Getz & Kubovy, 2018; Marks, 1978; Spence, 2011), and the emotional mediation account (e.g., Palmer, Schloss, Xu, & Prado-León, 2013; Spence, submitted; cf. Collier & Hubbard, 2001, 2004).

According to the semantic/linguistic account (Martino & Marks, 1999; Occelli et al., 2009), the fact that we use the same linguistic/lexical terms to describe different kinds of sensory stimuli provides one plausible explanation for why it is that people should experience a connection between them. In the English language, for example, tones are described as 'high' and 'low' in pitch (i.e., using the same terms used to describe elevation might help to explain the widely-documented connection between pitch and elevation (cf. Stumpf, 1888, for an analogous early argument in/concerning German). Support for the semantic/linguistic account comes from the work of Lawrence Marks and his collaborators at the John B. Pierce Laboratory at Yale University (e.g., Martino & Marks, 1999; Melara & Marks, 1990). Others have shown that words can be used to replace sensory stimuli and elicit congruency effects that are similar in magnitude (e.g., Gallace & Spence, 2006; Walker & Smith, 1984). That said, it is important to note that not all languages use the same terms in order to describe the pitch of a sound. Some languages, for instance, use the words 'thick' and 'thin' instead (e.g., Shayan, Ozturk, & Sicoli, 2011).

If the language-based account were to provide the sole/main explanation for the existence of the pitch-elevation correspondence, then one might have expected it to vary by language (or, more precisely, as a function of the different descriptors for pitch used in different languages). However, that does not appear to be the case (e.g., Fernández-Prieto, Spence, Pons, & Navarra, 2017). For instance, a remote Cambodian tribe who do not use spatial language to describe pitch also show the pitch-elevation mapping (see Parkinson, Kohler, Sievers, & Wheatley, 2012). Furthermore, linguistic stimuli do not always influence participants' performance in the same way that perceptual stimuli have been shown to do in certain more perceptual tasks (e.g., Maeda, Kanai, & Shimojo, 2004; see also Sadaghiani et al., 2009). Note here that McCormick et al (2018) also failed to find any evidence for semantic mediation in the crossmodal correspondences that they studied. And finally, it turns out that even pre-linguistic infants are sensitive to the pitch-elevation mapping (Dolscheid et al., 2014; Walker, Bremner, Mason, Spring, Mattock, Slater, & Johnson, 2010). Taken together, therefore, while the evidence does indeed support there being some role for shared language-based descriptions in accounting for crossmodal correspondences (at least under certain conditions and/or for certain correspondences), it clearly can't be the whole story (Fernández-Prieto et al., 2017; Gallace & Spence, 2006; Getz & Kubovy, 2018).

According to the statistical account, many correspondences reflect the internalization of the natural statistics of the environment (e.g., Parise, Knorre, & Ernst, 2014; Spence, 2011; see also Ernst, 2007). So, for example, it is a basic matter of physics that the size/mass of an object is correlated with its resonant frequency, such that the larger the object, the lower the frequency (see Coward & Stevens, 2004; Grassi, 2005; Kunkler-Peck & Turvey, 2000; Lakatos, McAdams, & Caussé, 1997; McMahon & Bonner, 1983). Were humans to internalize this natural correlation then it might, for instance, help to explain the pitch-size crossmodal correspondence that has been documented so frequently in the literature (e.g., Bien et al., 2012; Evans & Treisman, 2010; Gallace & Spence, 2006; Getz & Kubovy, 2018; Lowe & Haws, 2017; Melara & O'Brien, 1987; Parise & Spence, 2012; Walker & Smith, 1985). Dogs (at least the more intelligent amongst them) have also been shown to internalize such natural statistics (see Faragó, Pongrácz, Miklósi, Huber, Virányi, & Range, 2010).[3]

Parise et al. (2014) sent a freely-moving human mounted with a pair of directional microphones out into the world to record the natural statistics of the environment, in particular, with regard to any relation that might exist between pitch and elevation. Intriguingly, the results (approximately 50,000 1-second recordings from both indoors and out, urban and rural environments) demonstrated the existence of a natural correlation between the pitch of sounds and the elevation of their sources across a wide part of the frequency spectrum (see **Figure 1**).[4] The correlation was strongest in the 1-6 kHz range. Once again, therefore, the internalization of the natural statistics of the environment, perhaps as a 'coupling prior' in terms of Bayesian decision theory (Parise, 2016; Spence, 2011; Walker, Scallon, & Francis, 2017), could help to

[3] Sensitivity to the pitch-size correspondence likely also helps to explain the widespread existence of dishonest signalling in the animal kingdom; e.g., when animals deliberately lower the pitch of their calls/growls/howls, etc. in order to give the impression of being larger than is, in fact, the case (Fitch, 2000; Morton, 1977; see Ratcliffe, Taylor, & Reby, 2016, for a review).

[4] Of course, the fact that so many languages choose the same descriptors ('high' and 'low') to refer to the pitch of a sound is presumably not coincidental, and may well, in fact, turn out to be built on the natural statistics of the environment (see Spence, 2011, on this theme).

account for the pitch-elevation correspondence. It is also worth noting that a mapping between frequency and elevation was also observed in the filtering properties of the outer ear of humans.

INSERT FIGURE 1 ABOUT HERE

While the statistical account of crossmodal correspondences would appear to have broad appeal, it does require that one can point to the natural statistics of the environment underpinning such a relationship, or correspondence. However, that is not always the case (e.g., for the brightness-elevation correspondence documented in chimpanzees by Ludwig et al., 2011). Under such conditions, especially when dealing with those participants without any language abilities, the temptation is to reach for the suggestion that there may also be structural correspondences that are, in some sense, hard-wired (Walker et al., 2010; though see also Lewkowicz & Turkewitz, 1980; Spence & Deroy, 2012). According to the structural account, there are certain commonalities in neural encoding, such as the fact that more intense sensory stimuli are represented by increased neural firing in humans (see Stevens, 1957). The suggestion from some commentators is that this commonality might provide the basis for intensity-based correspondences (e.g., Lewkowicz & Turkewitz, 1980; Marks, 1978; Spence, 2011; see also Stevens & Marks, 1965). At the same time, however, the putatively transitive nature of crossmodal correspondences makes it rather more difficult to come to any firm conclusions concerning the origins of these associations. For example, while the brightness-elevation correspondence itself has no obvious direct basis in the statistics of the natural environment (as pointed out by Ludwig et al., 2011), it is possible that this is actually a second order association (see Spence & Deroy, 2012, on this theme). That is, it might reflect a transitive association between, say, pitch-elevation and pitch-brightness correspondences, both of which are present in the statistics of the environment – the latter perhaps linked to the "light-from-above" prior (see Adams, Graf, & Ernst, 2004).

Whether or not the structural account turns out to have validity, it is worth noting that, at best, it only explains a very small number of the correspondences that have been identified to date. Furthermore, given that increasing auditory pitch is also associated with increased neural firing (see Fernández-Prieto & Navarra, 2017), increasing pitch might also be expected to correspond crossmodally with increasing stimulus intensity (e.g., brightness). And indeed such a correspondence has been documented in several studies (Klapetek et al., 2012; Ludwig et al., 2011; Marks, 1987). What is more, it has been hypothesized that the intensity of auditory and visual events might be correlated in the natural statistics of the environment, hence perhaps ultimately rendering the structural account unnecessary. However, that being said, it is nevertheless still possible that future research might reveal that structural correspondences exhibit a different neural signature than do other classes of crossmodal correspondence (e.g., statistical; cf. Sadaghiani et al., 2009). Furthermore, one might also predict that any such structural correspondences, should they exist, would be more resistant to updating following any changes in the statistics of the environment – such as if, for instance, participants were to be exposed to an experimental test environment in which loudness was inversely correlated with brightness, say (cf. Brunel et al., 2015; Ernst, 2007; Getz & Kubovy, 2018; Xu, Yu, Rowland, Stanford, & Stein, 2012).[5]

Finally here, according to the emotional mediation account of crossmodal correspondences, the suggestion is that people match disparate sensory stimuli (e.g., such as musical excerpts and colour patches; Palmer et al., 2013) because of their shared emotional associations (see Spence, submitted, for a review). However, it is worth stressing that the emotional mediation account presumably explains more of the variance for those crossmodal correspondences involving stimuli that are themselves more strongly emotionally-valenced (such as, for example, music, or basic tastes; see Spence, submitted, on this point). By contrast, the pure tones used in so much of the research on crossmodal correspondences involving the pitch dimension presumably do not engender much of an emotional response in those participants who hear them. As such, while it may well be true that emotional mediation can explain a large part of the variance in people's crossmodal matching performance for certain highly emotionally-valent stimuli (such as non-modern music), it is likely to offer a less complete explanation of any auditory pitch-based correspondences (though see also Collier & Hubbard, 2001, 2004).

Here, it is interesting to note that the different accounts make somewhat different predictions concerning the relative vs. absolute nature of those crossmodal correspondences involving the pitch dimension. According to the semantic/shared-language-based account, all that matters is the linguistic label that happens to be attached to a stimulus. As such, the appropriateness of the descriptors 'low' or 'high' are presumably determined by the context in which a given tone is presented, thus supporting a relative account of the pitch-elevation correspondence. By contrast, Parise et al.'s (2014) findings would appear to suggest that there might be some absolute contribution to this correspondence, with the natural statistics of the environment

[5] Note here that in their research, Getz and Kubovy (2018) describe such correspondence-specific difficulties in reversing the natural crossmodal mapping in terms of bottom-up influence on the correspondences.

exhibiting a correlation between increasing pitch and increasing elevation, at least for a certain range of auditory frequencies (see **Figure 1**). Should the pitch-elevation association reflect the internalization of this particular statistical correspondence, then such data might then be taken to suggest that the conventional pitch-elevation correspondence would not be documented, or might even be reversed, should the auditory stimuli fall outside the mid-range frequencies of 1-6 kHz where the pitch-elevation correspondence was observed most clearly in the environment.

On the relative vs. absolute nature of crossmodal correspondences

In print, one often comes across the claim that crossmodal correspondences are relative phenomena. What this means, for example, is that it is the higher-pitched of two sounds that is matched with the higher (or smaller) of two visual stimuli, say, rather than there being a specific one-to-one crossmodal mapping between a particular pitch (or frequency) of sound and elevation/size of object). At the outset here, one might ask whether the distinction between the relative versus absolute nature of crossmodal correspondences really matters? One might also be tempted to wonder whether there is a simple answer to the question, one way or the other. It might turn out that the answer depends on the specific (class of) crossmodal correspondence under empirical investigation; who knows, perhaps some correspondences are (more) relative, while others are more absolute.

One of the reasons as to why this distinction between relative and absolute matters is because it is one of the most obvious, and most oft-mentioned, differences between synaesthesia and the crossmodal correspondences (see Deroy & Spence 2013a, b). In particular, one of the signature features of synaesthesia is that the idiosyncratic inducer-concurrent relation is based on an absolute mapping. It is a specific pitch of sound, for instance, that gives rise to a particular visual (e.g., colour) concurrent. By contrast, pitch-based crossmodal correspondences appear to be more relative in nature (though, “Relative to what?” soon becomes a key question). Furthermore, in answer to the second question posed above, it is important to note that whether correspondences are (more) relative or absolute may depend on the particular (class of) correspondence under empirical investigation. In other words, there is presumably no reason, *a priori*, to believe that just because auditory pitch-based correspondences are relative, say (if that does indeed turns out to be the case), that all other crossmodal correspondences must necessarily be relative too. In particular, we will return later to a consideration of the role that the nature of a given class of perceptual stimuli might have as far as the relative vs. absolute question is concerned.

However, beyond the synaesthesia question, it is important to stress that the relative/absolute nature of correspondences is also an issue of relevance to those who are solely interested in the latter. Indeed, getting to the bottom of the absolute vs. relative question will likely help to shed light on a number of fundamental questions in correspondences research, such as how crossmodal correspondences are acquired in the first place, whether they differ in terms of their perceptual consequences, and/or their resistance to retraining (i.e., when exposed to multisensory environments that exhibit different statistical regularities). For now, though, I want to switch to reviewing the experimental evidence directly supporting the relative account of auditory pitch-related correspondences, in particular.

A quick analysis of previous crossmodal correspondences research that has involved auditory pitch as one of the corresponding dimensions soon reveals that what sound frequencies have counted as high/low-pitched sound has varied markedly from one study to the next (see **Table 1** for a selective summary of this literature). What is also worth noting here is the fact that the researchers concerned never seem to mention having had to do any pre-testing in order to obtain the 'right' stimulus values (i.e., to settle on the most appropriate low- and high-pitched stimuli for use in their particular study). Taken together, therefore, the wide variation in pitch values associated with so-called 'low' or 'high' tones, and the lack of any obvious need for stimulus optimization/calibration, would appear to suggest that the absolute pitch of a sound may be of little relevance to the pattern of results obtained. Rather, it would seem to be the relative pitch (i.e., merely whether a tone is 'higher' or 'lower' in pitch than any comparison tone) that matters to such pitch-based correspondences. Of course, the fact that relative pitch can be shown to drive crossmodal correspondences does not in-and-of itself rule out the possibility that there is not some absolute contribution too. That said, what is also noticeable is that the majority of researchers have tended to use only a relatively narrow range of frequencies (see **Table 1**).

INSERT FIGURE 1 ABOUT HERE

At the same time, however, it is worth noting that <u>null</u> effects of variations of auditory pitch have been reported (or mentioned) in a number of studies (e.g., see Chiou & Rich, 2015; Getz & Kubovy, 2019; Gallace & Spence, 2006; Keetels & Vroomen, 2011; Klein, Brennan, &

Gilani, 1997; C. Cinel, pers. comm., email dated 20/07/2018). What is more, there are presumably likely to be more null results out there given the so-called ‘file drawer problem’ (see Rosenthal, 1979). In such cases, it is tempting to wonder whether the frequencies of tones that just so-happened to be chosen may have had something to do with the null results obtained (as might be expected if there was an absolute component to pitch-based correspondences). However, in the above-mentioned cases, the frequencies chosen would appear either to be the same as those used in studies that have documented significant effects of variations in pitch on participants’ behaviour, or at least in the same ballpark. What is also worth stressing here is that several of the null results in the above-mentioned studies are often reported in the context of significant pitch-based crossmodal correspondence/congruency effects (subject to some carefully-controlled modification of the experimental protocol/design or other).

One line of empirical support for the claim that those crossmodal correspondences involving auditory pitch are relative phenomena comes from the work of Gallace and Spence (2006). These researchers conducted a series of speeded classification studies (of visual size) in the presence of task-irrelevant auditory stimuli. The participants had to judge whether the second of two sequentially-presented filled circles was larger or smaller than the first (standard-sized) circle. The presentation of each circle was accompanied by a task-irrelevant sound (either a 300 or 4,500 Hz pure tone, or, in another experiment, the spoken words “low” and “high”). The results revealed that participants’ speeded discrimination (or categorization) performance was influenced by the task-irrelevant tone when the latter varied randomly on a trial-by-trial basis (Gallace & Spence, 2006, Experiment 1). In particular, participants responded more rapidly when the larger/smaller of the two circles was accompanied by the lower-/higher-pitched tone than when presented together with the crossmodally incongruent tones (i.e., higher-/lower-pitched, respectively). Intriguingly, however, when each of the tones was presented in separate blocks of experimental trials, no such pitch-size crossmodal congruency effect was observed (Gallace & Spence, 2006, Experiment 2).[6]

Such a pattern of results suggests that it is the trial-by-trial contrast between low- and high-pitch that matters, rather than necessarily the absolute pitch of any of the sounds that are presented. In fact, several other studies have also highlighted the need for trial-by-trial variation

[6] Meanwhile, the fact that the word stimuli gave rise to crossmodal congruency effects that were as large as those seen with tone stimuli (Gallace & Spence, 2006; Experiment 3) lends weight to the semantic-mediation/shared language based account, mentioned earlier.

in the pitch of task-irrelevant auditory stimuli to the elicitation of crossmodal congruency effects involving pitch (see Bernstein & Edelstein, 1971; Orchard-Mills et al., 2016; see also Garner, 1974, 1983, on the concept of separable and integral perceptual dimensions). That said, it is important to note that it is presumably unlikely that all of the possible crossmodal correspondences that happen to be associated with a relative change in pitch will necessarily be primed whenever the pitch of a sound changes (see Eitan & Timmers, 2010; Parise, 2016; see **Figure 2**). Indeed, this point links to the ongoing debate about the automaticity (or otherwise) of crossmodal correspondences (see Spence & Deroy, 2013, for a review). At present, the evidence would appear to suggest that crossmodal correspondences are not elicited automatically, contrary to the claim that one sometimes finds in the literature (see Chiou & Rich, 2015; Evans & Treisman, 2010; Getz & Kubovy, 2018; Mossbridge, Grabowecky, & Suzuki, 2011). Rather, participants may need to have their attention drawn to the task-relevant correspondence in order for it to influence their performance (Jamal et al., 2017). So, for example, Jamal et al. found that the interaction between auditory pitch and visual elevation was susceptible to modulation by auditory elevation when the latter was dissociated from visual elevation.

INSERT FIGURE 2 ABOUT HERE

Furthermore, of course, even when presenting a particular pure tone in a study, it is worth noting that the stimulus itself will presumably also possess several potentially relevant attributes – pitch, for one, but the sound itself will also have a certain loudness,[7] duration, and also abruptness of onset. Again, therefore, there may need to be some means of drawing a participant's attention to pitch as the relevant dimension (rather than, say, a sound's intensity – i.e., loud or quiet). Hence, all this to say that varying the pitch on a trial-by-trial basis might help to elicit crossmodal congruency effects simply because it draws the participant's attention to pitch as the relevant perceptual dimension rather than because pitch-based correspondences only operate as relative phenomena. Indeed, elsewhere the experimenter has sometimes had to explicitly point out the underpinning correspondence (e.g., between pitch and brightness) to the participant in order for its effect on human performance to be observed, as in the pip and

[7] Relevant here, Lacey et al. (2016) not only matched the duration of their stimuli, but also the perceived loudness to ensure that it was only the pitch of the sound that differed between the stimuli that they presented.

pop visual search task, popularized by Van der Burg, Olivers, Bronkhorst, and Theeuwes (2008; see Klapetek et al., 2012).

The “pip-and-pop effect” refers to the facilitation of search for a visual target (a horizontal or vertical bar whose colour changes frequently) among multiple visual distractors (tilted bars also changing colour unpredictably) as a result of the presentation of a spatially-nonpredictive auditory cue synchronized with the colour change of the visual target. Intriguingly, in Klapetek et al.’s (2012) first experiment, trial-by-trial variation in the pitch of the sound synchronized (and corresponding) with the change in brightness of the visual target had the same beneficial effect on participant’s performance (relative to a silent control condition) as when the mapping was reversed (Klapetek et al., 2012, Experiment 1). However, in a second experiment, when the nature of the pitch-brightness was pointed out to participants at the start of the study, and when crossmodally congruent and incongruent trials were now presented in separate blocks of trials, a significant crossmodal congruency effect was documented.[8]

More convincing evidence in support of the claim that crossmodal correspondences involving auditory pitch are relative comes from research showing that a given stimulus may be treated as either high or low depending on the specific context in which it happens to be presented (see Brunetti, Indraccolo, Del Gatto, Spence, & Santangelo, 2018; Chiou & Rich, 2012). So, for instance, Chiou and Rich conducted a series of experiments in which their participants had to make speeded detection responses to a series of visual targets presented randomly from one of two locations, either above or below central fixation. A task-irrelevant auditory cue was presented at a variable stimulus onset asynchrony (0-900 ms) prior to the visual target on each trial. The frequency of the task-irrelevant tone, which varied randomly on a trial-by-trial basis, was entirely non-predictive with regard to the likely elevation of the target. In one study, the presentation of a 900 Hz pure tone was shown to lead to an exogenous shift of participants’ visuospatial attention upward (i.e., speeding responses to upper as compared to lower targets) when the other tone in the same block of trials had a frequency of 100 Hz (Chiou & Rich, 2015, Experiment 3). However, in other blocks of trials in which the 900 Hz tone alternated with a 1700 Hz tone instead, it led to an exogenous shift of participants’ spatial attention downward

[8] Note here though that even though congruency was blocked, the pitch of the sound still alternated on each and every trial, presumably drawing attention to pitch as the relevant dimension. Note that an extensive cognitive psychology literature supports the claim that our attention is primarily drawn to change in the environment.

instead (i.e., speeding responses to the lower targets; see also Chiou & Rich, 2015; Mossbridge et al., 2011).

Meanwhile, elsewhere, Brunetti et al. (2018) presented low- (300 Hz), medium- (1,200 Hz), and high-pitched (4,500 Hz) task-irrelevant sounds in a random sequence on successive trials. The participants had to make speeded classification responses concerning the size of a visually-presented circle. The 1,200 Hz tone was shown to act like a 'low' tone if it followed a 4,500 Hz tone on the preceding trial, but like a 'high' stimulus if it followed a 300 Hz tone instead. Such results once again support the view that the pitch-based correspondences (both pitch-elevation and pitch-size) are relative in nature; Or, at the very least, they show that under the appropriate conditions, the relative pitch of a sound can dictate the pattern of results observed.[9]

A final piece of evidence relevant to the relative account of crossmodal correspondences (at least those involving the auditory dimension of pitch) comes from Chiou and Rich's (2012; Experiment 2) observation that the absolute magnitude of the pitch difference between the low- and high-pitched sound modulated the magnitude of the crossmodal influences that are observed (e.g., when assessing the impact of presenting a task-irrelevant low- vs. high-pitched tones) on the exogenous vertical orienting of visual spatial attention. In particular, and as shown in **Figure 3**, these researchers demonstrated crossmodal spatial cuing effects in those blocks of trials where the pair of tones were widely separated in pitch (300 and 1500 Hz) but not when they were separated by only 100 Hz (300 and 400 Hz; Chiou & Rich, 2012; Experiment 2; see Melara & Mounts, 1994, for a similar pattern of results in an intramodal auditory compatibility study involving frequency and loudness). Note here also that McCormick et al. (2018) did not find strong neural evidence for a magnitude-based account of crossmodal correspondences.

While such results might perhaps be taken to argue for an absolute contribution to the pitch elevation mapping, there are a number of alternative explanations that can be forwarded. These include the suggestion that reducing the pitch difference simply reduced the salience of this stimulus contrast. Alternatively, however, the smaller pitch difference might simply have led to a much smaller vertical shift in participants' attention, hence failing to influence performance at those locations at which target detection was assessed. It is also possible that two pitches corresponding to approximately the same elevation may just be too close to

[9] Note here that Walker and Walker (2016) have also provided evidence showing that the crossmodal correspondence between visual brightness and haptic size is more relative than absolute, using a similar experimental procedure.

separate spatially. Finally here, returning to Parise et al.'s (2014) results shown earlier, one might also be tempted to wonder whether the narrow pitch separation condition relied on presenting two tones that both fell in a part of the frequency spectrum where the natural correlation between pitch and elevation doesn't exist (in fact, the inverse relationship holds). Hence, in order to be certain that it is not merely the part of the frequency spectrum that matters to the null results reported by Chiou and Rich in the narrow frequency separation condition one would also want to know whether a similar null result would also have been documented if tones of 1,400 Hz and 1,500 Hz had been used instead. (The non-monotonic nature of the frequency-elevation mapping potentially adds a layer of complexity to research here.)

INSERT FIGURE 3 ABOUT HERE

It is worth stressing that Chiou and Rich's (2012) results stand in contrast with other results demonstrating little effect of the magnitude of any pitch difference on performance in various crossmodal correspondence tasks (see Ben-Artzi & Marks, 1995; Getz & Kubovy, 2018). So, for instance, Ben-Artzi and Marks reported similar magnitude congruency effects in a speeded visual elevation classification task no matter the pitch separation involved, at least when the auditory and visual stimulus values varied independently. Similarly, as far as it is possible to tell (given the lack of key details in the published short report), Getz and Kubovy also failed to document any influence of the pitch range chosen on the pattern of audiovisual crossmodal correspondences reported (see **Table 1**).[10] That all being said, it is certainly true to say that a few researchers have, over the years, pondered whether the presentation of stimuli at the extreme end of some perceptual continua or other might not give rise to correspondences that can be reliably ranked in the absence of any comparison stimulus (e.g., see Marks, Szczesiul, & Ohlott, 1986). However, here it is sufficient to note that as yet little (or rather no) evidence has been provided in support of such intuitive claims. What is more, at least for the pitch-elevation correspondence, the evidence (from Parise et al.'s, 2014, study of the natural statistics of the environment) actually suggests that very different results might be obtained at the extremes of the frequency spectrum than for frequencies in the mid-range. Though here it should be remembered that the natural statistics potentially underpinning other

[10] See Fernández-Prieto and Navarra (2017) for a similar consideration of whether the breadth of the rising or falling tone sweep affects the spatial orienting of visual orienting in the elevation dimension.

correspondences, such as those linking pitch and size, for instance, might well be expected to exhibit more of a monotonic relationship.

Interim summary

Taken together, the evidence reported to date clearly supports the view that crossmodal correspondences involving the auditory dimension of pitch are relative (e.g., Brunetti et al., 2018; Chiou & Rich, 2012; Gallace & Spence, 2006; Melara & O'Brien, 1987; though see also Mossbridge et al., 2011, Experiment 2; Pedley & Harper, 1959; Roffler & Butler, 1968). That is, no strong evidence has been published in support of the existence of absolute mappings involving pitch outside, that is, the world of those non-synaesthetes with absolute pitch.[11] Of course, the fact that crossmodal correspondences involving auditory pitch in those without perfect pitch (i.e., the majority of the population) should be relative might have no necessary implications for the relative versus absolute nature of other crossmodal correspondences. Intriguingly, elsewhere in the world of crossmodal correspondences research, researchers have argued both for relative and absolute correspondences. However, as we will see below, even here, closer inspection of the very limited supporting evidence that has been published to date suggests that such correspondences, at least in those involving prothetic stimulus dimensions (see Stevens, 1957), such as amplitude/intensity also exhibit properties that are more consistent with correspondences being more relative than absolute.

Evidence supporting absolute frequency-based crossmodal correspondences

Guzman-Martinez et al. (2012) argued for the absolute nature of crossmodal correspondences on the basis of a series of studies in which their participants had to adjust the auditory temporal amplitude modulation (AM) rate and pitch of a sound until it perceptually 'matched' the visual spatial frequency of Gabor displays having one of three different spatial frequencies.

[11] A separate empirical question here would be to assess how those individuals with absolute pitch perform in those tasks playing on crossmodal correspondences involving auditory pitch (such as, for example, Brunetti et al.'s, 2018, speeded classification task). At the same time, however, it is also interesting (and perhaps informative) to note that many of those with absolute pitch also have synaesthesia, although that is by no means always the case (e.g., Carroll & Greenberg, 1961; Cytowic, & Eagleman, 2009; Isbilen & Krumhansl, 2016; Loui, Zamm, & Schlaug, 2012; Petroviv, Antovic, Milankovic, & Acic, 2012).

Interestingly, the participants consistently matched a specific auditory AM rate to each of the visual spatial frequencies (with a relationship that was approximately linear). What is more, these matches persisted even when the observer had to determine an auditory AM rate that matched only a single visual spatial frequency (Experiment 2). The same pattern of results was also obtained in another study when the participants matched tactile AM rate and visual spatial frequency instead (Experiment 3). The consistent results of four crossmodal matching experiments led Guzman-Martinez et al. to suggest that there might be an absolute crossmodal mapping between visual spatial frequency and auditory/tactile temporal AM rate (features coded in primary sensory cortex). Subsequent experiments demonstrated that these seemingly-absolute correspondences could be used to direct attention/awareness to the visual Gabor whose spatial frequency matched the auditory temporal AM rate (Experiments 5 and 6). As mentioned earlier, perhaps the degree of absoluteness of any given correspondence will depend on the extent to which the sensory estimates pick-up on the same, or analogous, property out there in the environment (see also Guzman et al., 2012; Yau, Olenczak, Dammann, & Bensmaia, 2009). (Indeed, perhaps the degree of which a given crossmodal correspondence is absolute vs. relative might turn out to correlate with how surprising people find it on first learning about it.)

Importantly, however, such claims have only been partially confirmed by subsequent empirical research (Orchard-Mills, Van der Burg, & Alais, 2013). Specifically, while Orchard-Mills and her colleagues also obtained a linear relationship between AM auditory noise and visual spatial frequency (Experiment 1), other data from their study (see their Experiments 2-5) led the latter authors to conclude that this particular crossmodal correspondence/interaction is, in fact, more flexible and based on a relatively broadly-tuned, rather than absolute, frequency mapping instead. The latter experiments involved the participants performing a visual search task for a specific grating in search displays containing multiple Gabors, having different spatial frequencies. The research question addressed was whether crossmodally-matching auditory stimuli would direct participants' visual attention to the appropriate Gabor.

The fact that the crossmodal correspondence effects in Orchard-Mills et al.'s (2013) study could be demonstrated with variable pairs of visual and auditory frequencies certainly leaves room for there being both absolute constraints as well as some degree of relative context-dependent flexibility (at least within those constraints). Orchard-Mills et al. themselves chose to emphasize the top-down knowledge component of this mapping over the stimulus-driven and automatic capture of attention (e.g., given the fact that informative but non-matching auditory stimuli also facilitated performance while an uninformative but matching auditory signal did not). They are perhaps worth quoting at length here to emphasize the conclusion they come to on the relative vs. absolute (or specific) issue: *"Lastly, Experiment 5 suggests that there is no specific mapping of auditory to visual frequencies: It is a relative and variable mapping. Whether a particular auditory frequency guides search to a given visual frequency or not depends on what other frequencies are present in the stimulus set. Taken together, these experiments show that the conditions under which visual selection of spatial frequency can be enhanced by ''matched'' amplitude-modulated auditory stimuli are both limited and variable. This is incompatible with a specific mapping of auditory and visual frequencies that occurs early and automatically and instead points to a late, flexible, and strategic mapping."* (Orchard-Mills et al., 2013, p. 12).

Of course, convincing though Orchard-Mills et al.'s (2013) arguments/data may be, we are still left with the (rather more) challenging question of how, in principle, one could ever hope to provide convincing evidence for the existence of absolute crossmodal correspondences involving the auditory dimension of pitch. Perhaps one has to admit that whether or not there is an absolute contribution to such correspondences, it is hard to design the appropriate experiments so as to rule out a relative contribution. That said, the contrast with synaesthesia, where the vast majority of the research supports the existence of absolute correspondences stands out as being all the more surprising. It is perhaps, then, a challenge best left to future researchers to try and come up with a set of necessary and sufficient criteria that would provide unequivocal support for an absolute account of (e.g.,) pitch-based crossmodal correspondences

Polarity-based correspondences

It is important to stress that no matter whether or not there is any kind of absolute correspondence, or contribution to a particular correspondence (perhaps based on the natural correlation between two stimulus dimensions in the environment), there is presumably always the possibility that participants can establish some kind of 'polarity-based' correspondence between disparate stimulus dimensions. According to one version of the polarity-correspondence principle, emerging originally from the field of human factors research (e.g., Chang & Cho, 2015; Proctor & Cho 2006; though see also Lakens, 2012), when response alternatives are laid out spatially (e.g., up vs. down; or left vs. right response keys) then perceptual/conceptual dimensions that also exhibit some kind of inherent spatial mapping, or correspondence (e.g., pitch-elevation; good-up/bad-down; Meier, Hauser, Robinson, Kelland Friesen, & Schjeldahl, 2007; Meier & Robinson, 2004) may well give rise to congruency effects in response selection (hence the relevance to human factors researchers in the context of interface design).

Note here that the conflict in polarity codes is normally between a pair of perceptual stimuli and a pair of responses that are thought to activate opposite ends of a polarity spectrum, thus slowing down response-selection processes when the polarity of the response is incongruent with that associated with the perceptual stimulus. So, for example, a high tone and an upper visual target would be deemed congruent with an upper response alternative (e.g., an upper response button) because all of the stimuli would be coded as positive, rather than negative, polarity (as would a positively-valenced stimulus such as a smiling face, or the word happy). Note that such response-linked polarity codes have been associated with both physical/perceptual and conceptual dimensions (see also Lynott & Coventry, 2014). And crucially, in contrast to other accounts of the correspondences (including the other polarity-based account mentioned below), this approach tends to stress the benefits to performance when multiple positive pole stimuli/dimensions overlap while suggesting little or no effect for those combinations of stimuli at the negative polarity. (Note that any such asymmetry in the data may be missed as the ‘positive’ and ‘negative’ polarity stimuli are typically averaged in traditional crossmodal correspondences research.)

Somewhat confusingly, however, separate from this spatial-response-based polarity correspondence literature, there are also a number of researchers out there who have looked at the tendency, amongst both adults and children, to align (or experience a correspondence between) seemingly disparate polar perceptual dimensions (e.g., Gardner, 1974; Smith & Sera, 1992). Note that while evidence in support of the existence of the latter polar perceptual correspondences often comes from studies involving spatially-arrayed response keys, the existence of the correspondence is thought to be independent of any response component (thus contrasting with the response-linked polarity correspondences outlined in the preceding paragraph).

The fact that no correspondence has been observed between certain pairs of stimulus dimensions – such as between pitch and contrast (Evans & Treisman, 2010), pitch and hue

(Bernstein, Eason, & Schurman, 1971), and auditory loudness and visual lightness (Marks, 1987) – might be taken to argue against the ubiquity of polarity-based correspondences. That said, as noted by Marks, what looks like a null result at a population level might, in certain cases at least represent a bimodal split in the way in which the ends of perceptual continua are assigned to the positive and negative poles (see also Schietecat, Lakens, IJsselsteijn, & de Kort, 2018a, b). And here it is perhaps worth stressing that while pitch may be a metathetic dimension, according to Stevens (1957), its close correspondence with both elevation and size affords opposing prothetic (or polarity) tendencies. High pitch is small, but also high, while low pitch is large, but also low. Hence depending on the context, or corresponding dimension that is activated/attended/primed, high/low pitch might be associated with the 'more' end of a prothetic continuum (or the positive end of the polarity correspondence). Nevertheless, the key point remains that there are presumably normally multiple factors/influences giving rise to any given crossmodal correspondence, and that these accounts vary in the extent to which they stress the importance of relative vs. absolute mappings.

On perceptual continua/dimensions

S. S. Stevens (1957, p. 154) highlighted a potentially important distinction between two kinds of perceptual continua, namely prothetic (magnitude) and metathetic (qualitative). The former are concerned with those continua having to do with quantity (i.e., how much), while the latter describes continua where the stimuli are arranged in qualitative terms (i.e., what kind or where). Loudness, size, duration, and rate of change are all thought to constitute prothetic dimensions with a clear "more than" end (e.g., loud, bright, big, etc.) and a "less than" end (e.g., quiet, dark, small, etc.; see also Walsh, 2003). According to Panek and Stevens (1966), the saturation of colour constitutes another prothetic dimension. By contrast, according to Stevens (1957), pitch constitutes a metathetic dimension, since a high-pitched tone is different in kind from a low-pitched tone, without necessarily being meaningfully related in a more than/less than way. Note here that hue and shape are also described as metathetic dimensions (Smith & Sera, 1992).

However, in addition to the distinction between prothetic and metathetic stimulus dimensions, there are also certain perceptual continua or dimensions that can be arranged in a circular manner. Examples of the latter include both hue and pitch (see Gilbert, Fridlund, Lucchina, 2016). Indeed, over the years, the fact that both hue and auditory pitch can be organized as circular dimensions has led a number of researchers to assume/try to uncover the nature of the 'structural' correspondence between these seemingly similarly-organized perceptual continua (e.g., Caivano, 1994; Garner, 1978; Pridmore, 1992; Wells, 1980; though see also Davis, 1979). It will therefore be an intriguing question for future research to determine whether the kind of perceptual quality/continua has any relevant implications concerning the relative vs. absolute nature of any crossmodal correspondences involving that stimulus dimension. Though perhaps the more relevant question here may turn out to be one of whether the classes of stimuli involved afford categorical labelling (such as hue, basic taste, and, for that matter, gender; Smith, Grabowecky, & Suzuki, 2007). The ability to assign a categorical label to a perceptual stimulus, especially in the case of metathetic continua/dimensions, would seem to bias a correspondence more toward the absolute end of the spectrum. That said, any categorically-labelled stimulus can presumably also be turned into an intensity-based prothetic dimension – so while hue might be a categorical label permitting of absolute correspondences (such as 'yellow = sour') the intensity of yellowness, or the intensity of sourness can be varied prothetically (cf. Panek & Stevens, 1966).

It is also important to note that there are other classes of perceptual stimuli/qualities that cannot be organized onto any kind of meaningful continuum. Here one might think of basic tastes (e.g., sweet, sour, bitter, salty, umami, etc) or the timbre of sounds. Despite the best attempts of certain researchers (e.g., think only of Henning's taste tetrahedron; Henning, 1927), there is simply no commonly agreed on means of organizing the four or five basic tastes (e.g., bitter, sour, sweet, salty, and nowadays, also umami), say. As to which one belongs in the middle of the series, there is, as yet, simply no answer. And yet perceptual qualities such as basic tastes do correspond to different values on regularly organized perceptual continua (be they prothetic, metathetic, and/or circular). It is, though, worth considering, given the purported transitivity of the correspondences (e.g., Deroy & Spence, 2013b; Spence & Deroy, 2012), whether the basic tastes would align up in the same order regardless of the correspondence under consideration (e.g., taste-pitch, taste-elevation, taste-hue, etc. see Velasco, Adams, Petit, & Spence, 2019). As such, while it may well be true that the basic tastes themselves might have no clear metathetic organization, they could potentially come to have such a systematic organization by reference to the neatly-organized stimulus dimensions with which they correspond. However, that does not necessarily appear to be the case.

Conclusions

As this summary of the crossmodal correspondences literature involving pitch has hopefully made clear, while auditory pitch-based crossmodal correspondences appear to be amongst the most robust/commonly studied of all correspondences (e.g., see Deroy et al., 2018; Spence, 2011, for reviews), there are still a number of fundamental questions awaiting an answer concerning the underpinning nature of the effect being studied, specifically in terms of just how relative vs. absolute pitch-based correspondences are. That said, the majority of the evidence on pitch-based correspondences that has been published to date would appear to be more supportive of the relative than the absolute (or specific) account, despite occasional claims to the contrary. While the majority of the studies reported in this review have involved static stimuli, it is worth noting that many natural stimuli exhibit a dynamic change (e.g., in pitch or position)[12] and hence in future research it will be interesting to understand better the ways in which such correspondences involving such dynamic stimuli are similar to/different from what is seen in the case of unvarying (i.e., steady) stimuli (e.g., see Deroy et al., 2018; Fernández-Prieto & Navarra, 2017; Maeda et al., 2004; Sadaghiani et al., 2009). One relevant question here might be to consider whether those stimuli that incorporate a dynamic change of pitch might give rise to stronger (or more automatic) correspondences, as suggested by the work of Mossbridge et al. (2011; see also Eitan, Schupak, Gotler, & Marks, 2014).[13]

As forcefully highlighted by Parise (2016), it is worth noting that the majority of crossmodal correspondences research has typically only used a pair of stimulus values in each dimension. Such an experimental approach appears to be based on the (implicit) assumption of there being a monotonic mapping underlying the association between such dimensions. While this is likely sometimes the case (as, perhaps, in the mapping between the rate of amplitude modulation of sounds, and spatial frequencies in vision; Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013), at other times it turns out that sensory cues (at least in the environment) are associated by means of mappings that are distinctly non-monotonic in nature (e.g., as in the case of the pitch–elevation mapping, Parise et al., 2014). Given such results, one can certainly understand why Parise (2016) was so keen to suggest that cognitive psychologists/scientists should therefore not just limit their research to merely a couple of stimulus values along some

[12] Indeed, as Mossbridge et al. (2011) note, the majority of natural sounds are, after all, frequency-modulated.

[13] There might also be the possibility of using auditory illusions such as the Deutsch staircase (Deutsch, Dooley, & Henthorn, 2008; Shepard, 1964) – a sound sequence that permanently sounds as if to be ascending in pitch to permanently convey a corresponding attribute. If it is relative, then how long does a particular tone remain 'high'.

perceptual continua, such as, for example, pitch. However, while Parise's (2016) suggestion undoubtedly makes a lot of sense at one level, it is worth noting that polarity-based (or linguistic) correspondences can presumably superimpose in many testing situations. As such, simply picking a couple of values might be considered sufficient if one merely wants to make recommendations for interface design, say. Ultimately, therefore the question of how to assess the correspondences may, to a certain extent at least, may determine what one finds. And given that there are presumably a number of reasons as to why researchers, interface designers, commentators, and artists (to name but a few disciplines) have, over the centuries, been interested in the correspondences, there may be no fundamental truth, or right way of going about their assessment. Nevertheless, in the short-term it will likely be helpful to determine whether the putatively different classes of crossmodal correspondence behave differently (e.g., are they equally susceptible to any updating of the statistics of the environment), and/or the extent to which they can be associated with distinctively different neural substrates.

REFERENCES

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience*, **7**, 1057-1058.

Belkin, K., Martin, R., Kemp, S. E., & Gilbert, A. N. (1997). Auditory pitch as a perceptual analogue to odor quality. *Psychological Science*, **8**, 340-342.

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

Bernstein, I. H., & Eason, T. R., & Schurman, D. L. (1971). Hue-tone interaction: A negative result. *Perceptual and Motor Skills*, **33**, 1327-1330.

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: A combined TMS, EEG, and psychophysics study. *NeuroImage*, **59**, 663-672.

Braaten, R. (1993). *Synesthetic correspondence between visual location and auditory pitch in infants*. Paper presented at the Annual Meeting of the Psychonomic Society.

Brunel, L., Carvalho, P. F., & Goldstone, R. L. (2015). It does belong together: Cross-modal correspondences influence cross-modal integration during perceptual learning. *Frontiers in Psychology*, **6**:358. doi:10.3389/fpsyg.2015.00358

Brunetti, R., Indraccolo, A., Del Gatto, C., Spence, C., & Santangelo, V. (2018). Are crossmodal correspondences relative or absolute? Sequential effects on speeded classification. *Attention, Perception, & Psychophysics*, **80**, 527-534.

Caivano, J. L. (1994). Color and sound: Physical and psychophysical relations. *Color Research and Application*, **19(2)**, 126-133.

Carnevale, M. J., & Harris, L. R. (2016). Which direction is up for a high pitch? *Multisensory Research*, **29**, 113-132.

Carroll, J. B., & Greenberg, J. H. (1961). Two cases of synesthesia for color and musical tonality associated with absolute pitch ability. *Perceptual and Motor Skills,* **13***,* 48.

Chang, S., & Cho, Y. S. (2015). Polarity correspondence effect between loudness and lateralized response set. *Frontiers in Psychology*, **6**:683.

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

Chiou, R., & Rich, A. N. (2015). Volitional mechanisms mediate the cuing effect of pitch on attention orienting: The influences of perceptual difficulty and response pressure. *Perception*, **44**, 169-182.

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.

Coward, S. W., & Stevens, C. J. (2004). Extracting meaning from sound: Nomic mappings, everyday listening, and perceiving object size from frequency. *Psychological Record*, **54**, 349-364.

Crisinel, A. S., & Spence, C. (2010a). A sweet sound? Exploring implicit associations between basic tastes and pitch. *Perception*, **39**, 417-425.

Crisinel, A.-S., & Spence, C. (2010b). As bitter as a trombone: Synesthetic correspondences in non-synesthetes between tastes and flavors and musical instruments and notes. *Attention, Perception, & Psychophysics*, **72**, 1994-2002.

Cytowic, R. E., & Eagleman, D. M. (2009). *Wednesday is indigo blue: Discovering the brain of synesthesia*. Cambridge, MA: MIT Press.

Davis, J. W. (1979). A response to W. Garner's observation on the relationship between colour and music. *Leonardo*, **12**, 218.

Deroy, O., Crisinel, A.-S., & Spence, C. (2013). Crossmodal correspondences between odors and contingent features: Odors, musical notes, and geometrical shapes. *Psychonomic Bulletin & Review,* **20**, 878-896.

Deroy, O., Fernandez-Prieto, I., Navarra, J., & Spence, C. (2018). Unravelling the paradox of spatial pitch. In T. L. Hubbard (Eds.), *Spatial biases in perception and cognition* (pp. 77-93). Cambridge, UK: Cambridge University Press.

Deroy, O., & Spence, C. (2013a). Learning 'arbitrary' crossmodal correspondences: Staying away from neonatal synaesthesia. *Neuroscience & Biobehavioral Reviews,* **37**, 1240-1253.

Deroy, O., & Spence, C. (2013b). Weakening the case for 'weak synaesthesia': Why crossmodal correspondences are not synaesthetic. *Psychonomic Bulletin & Review*, **20**, 643-664.

Deutsch, D., Dooley, K., & Henthorn, T. (2008). Pitch circularity from tones comprising full harmonic series. *Journal of the Acoustical Society of America*, **124**, 589-597.

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science,* **25**, 1256-1261.

Eitan, Z. (2017). Musical connections: Cross-modal connections. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp. 213-224). London, UK: Taylor & Francis.

Eitan, Z., Schupak, A., Gotler, A., & Marks, L. E. (2014). Lower pitch is larger, yet falling pitches shrink. *Experimental Psychology*, **61**, 273-284.

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.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, **7/5/7**:1-14.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, **10(1)**:6, 1-12.

Faragó, T., Pongrácz, P., Miklósi, Á., Huber, L., Virányi, Z., & Range, F. (2010). Dogs' expectation about signalers' body size by virtue of their growls. *PLoS ONE*, **5(12)**:e15175.

Fernández-Prieto, I., & Navarra, J. (2017). The higher the pitch the larger its crossmodal influence on visuospatial processing. *Psychology of Music*, **45(5)**, 713-724.

Fernández-Prieto, I., Navarra, J., & Pons, F. (2015). How big is this sound? Cross-modal association between pitch and size in infants. *Infant Behavior and Development*, **38**, 77-81.

Fernandez-Prieto, I., Spence, C., Pons, F., & Navarra, J. (2017). Does language influence the vertical representation of auditory pitch and loudness? *i-Perception*, **May-June**, 1-11.

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences,* **4**, 258-267.

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

Gardner, H. (1974). Metaphors and modalities: How children project polar adjectives onto diverse domains. *Child Development*, **45**, 84-91.

Garner, W. R. (1974). *The processing of information and structure*. Lawrence Erlbaum Associates: Potomac, Maryland.

Garner, W. (1978). The relationship between colour and music. *Leonardo*, **11**, 225-226.

Garner, W. R. (1983). Asymmetric interactions of stimulus dimensions in perceptual information processing. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition, and development: Interactional analyses* (pp. 1-37). Hillsdale, NJ: Erlbaum.

Getz, L. M., & Kubovy, M. (2018). Questioning the automaticity of audiovisual correspondences. *Cognition*, **175**, 101-108.

Gilbert, A. N., Fridlund, A. J., Lucchina, L. A. (2016). The color of emotion: A metric for implicit color associations. *Food Quality & Preference*, **52**, 203-210.

Goethe, J. W. von (1810). *Zur Farbenlehre* (*Theory of colours*; trans. Charles Locke Eastlake). London, UK: John Murray.

Grassi, M. (2005). Do we hear size or sound: Balls dropped on plates. *Perception & Psychophysics*, **67**, 274-284.

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. *Current Biology*, **22**, 383-388.

Henning, H. (1927). Psychologische Studien am Geschmackssinn. In F. Aberholden (Ed.), *Handbuch der biologischen arbeitsmethoden* (pp. 627-740). Berlin: Urban and Schwarzenberg.

Holt-Hansen, K. (1968). Taste and pitch. *Perceptual and Motor Skills*, **27**, 59-68.

Holt-Hansen, K. (1976). Extraordinary experiences during cross-modal perception. *Perceptual and Motor Skills*, **43**, 1023-1027.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *American Journal of Psychology*, **109**, 219-238.

Isbilen, E. S., & Krumhansl, C. L. (2016). The color of music: Emotion-mediated associations to Bach's Well-Tempered Clavier. *Psychomusicology: Music, Mind, and Brain*, **26**, 149-161.

Jamal, Y., Lacey, S., Nygaard, L., & Sathian, K., (2017). Interactions between auditory elevation, auditory pitch and visual elevation during multisensory perception. *Multisensory Research*, **30**, 287-306.

Jonas, C., Spiller, M. J., & Hibbard, P. (2017). Summation of visual attributes in auditory–visual crossmodal correspondences. *Psychonomic Bulletin & Review*, **24**, 1104-1112.

Keetels, M., & Vroomen, J. (2011). No effect of synesthetic congruency on temporal ventriloquism. *Attention, Perception, & Psychophysics*, **73**, 209-218.

Klapetek, A., Ngo, M. K., & Spence, C. (2012). Do crossmodal correspondences enhance the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics,* **74**, 1154-1167.

Klein, R. M., Brennan, M., & Gilani, A. (1987, November). *Covert cross-modality orienting of attention in space*. Paper presented at the annual meeting of the Psychonomics Society, Seattle.

Knöferle, K. M., & Spence, C. (2012). Crossmodal correspondences between sounds and tastes. *Psychonomic Bulletin & Review,* **19**, 992-1006.

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 279-294.

Lacey, S., Martinez, M., McCormick, K., & Sathian, K. (2016). Synesthesia strengthens sound-symbolic cross-modal correspondences. *European Journal of Neuroscience*, **44**, 2716-2721.

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception & Psychophysics*, **59**, 1180-1190.

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **38**, 726-736.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Developmental Psychology*, **16**, 597-607.

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance*, **33**, 1189-1207.

Loui, P., Zamm, A., & Schlaug, G. (2013). Absolute pitch and synesthesia: Two sides of the same coin? Shared and distinct neural substrates of music listening. *ICMPC*, **2012**, 618-623.

Lowe, M. L., & Haws, K. L. (2017). Sounds big: The effects of acoustic pitch on product perceptions. *Journal of Marketing Research*, **54(2)**, 331-346.

Ludwig, V. U., Adachi, I., & Matzuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences USA*, **108**, 20661-20665.

Lynott, D., & Coventry, K. (2014). On the ups and downs of emotion: Testing between conceptual-metaphor and polarity accounts of emotional valence–spatial location interactions *Psychonomic Bulletin & Review*, **21**, 218-226.

Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology*, **14**, R990-R991.

Makovac, E., & Gerbino, W. (2010). Sound–shape congruency affects the multisensory response enhancement. *Visual Cognition*, **18**, 133-137.

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. (1978). *The unity of the senses: Interrelations among the modalities*. New York, NY: Academic Press.

Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception & 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, MA: MIT Press.

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *International Journal of Psychophysiology*, **50**, 125-145.

Marks, L. E., Szczesiul, R., & Ohlott, P. (1986). On the cross-modal perception of intensity. *Journal of Experimental Psychology: Human Perception and Performance*, **12**, 517-534.

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.

McCormick, K., Lacey, S., Stilla, R., Nygaard, L. C., & Sathian, K. (2018). Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia,* **112**, 19-30.

McMahon, T. A., & Bonner, J. T. (1983). *On size and life*. New York, NY: Scientific American.

Meier, B. P., Hauser, D. J., Robinson, M. D., Kelland Friesen, C., & Schjeldahl, K. (2007). What's "up" with God? Vertical space as a representation of the divine. *Journal of Personality and Social Psychology*, **93**, 699-710.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science*, **15**, 243-247.

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

Melara, R. D. (1989b). Similarity relations among synesthetic stimuli and their attributes. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 212-231.

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

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

Melara, R. D., & Mounts, J. R. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Perception & Psychophysics*, **56**, 73-90

Miller, J. O. (1991). Channel interaction and the redundant targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, **17**, 160-169.

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

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, **111**, 855-869.

Mossbridge, J. A., Grabowecky, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual-spatial attention. *Cognition*, **121**, 133-139.

Mudd, S. A. (1963). Spatial stereotypes of four dimensions of pure tone. *Journal of Experimental Psychology*, **66**, 347-352.

Newton, I. (1952). *Opticks or a treatise on the reflections, refractions, inflections & colours of light*. New York, NY: Dover Publications. (Original work published 1730).

O'Boyle, M. W., & Tarte, R. D. (1980). Implications for phonetic symbolism: The relationship between pure tones and geometric figures. *Journal of Psycholinguistic Research*, **9**, 535-544.

Occelli, V., Spence, C., & Zampini, M. (2009). Compatibility effects between sound frequencies and tactually stimulated locations on the hand. *Neuroreport*, **20**, 793-797.

Orchard-Mills, E., Van der Burg, E., & Alais, D. (2013). Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies. *Journal of Vision*, **13**:6.

Orchard-Mills, E., Van der Burg, E., & Alais, D. (2016). Crossmodal correspondence between auditory pitch and visual elevation affects temporal ventriloquism. *Perception*, **45**, 409-424.

Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music-color associations are mediated by emotion. *Proceedings of the National Academy of Sciences of the USA,* **110**, 8836-8841.

Panek, W., & Stevens, S. S. (1966). Saturation of red: A prothetic continuum. *Perception & Psychophysics*, **1(1)**, 59-66.

Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research*, **29**, 7-28.

Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences of the USA*, **111**, 6104-6108.

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

Parise, C., & Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE,* **4(5):**e5664.

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: An IAT study. *Experimental Brain Research,* **220**, 319-333.

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception,* **41**, 854-861.

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

Pedley, P. E., & Harper, R. S. (1959). Pitch and the vertical localization of sound. *The American Journal of Psychology*, **72**, 447-449.

Petrovic, M., Antovic, M., Milankovic, V., & Acic, G. (2012). Interplay of tone and color: Absolute pitch and synesthesia. *Proceedings in the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 799-806). July 23-28th. Thessaloniki, Greece.

Pitteri, M., Marchetti, M., Priftis, K., & Grassi, M. (2017). Naturally together: Pitch-height and brightness as coupled factors for eliciting the SMARC effect in non-musicians. *Psychological Research*, **81**, 243-254.

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, **13**, 278-285.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, **132**, 416-442.

Pridmore, R. W. (1992). Music and color: Relations in the psychophysical perspective. *Color Research & Application*, **17**, 57-61.

Ratcliffe, V. F., Taylor, A. M., & Reby, D. (2016). Cross-modal correspondences in non-human mammal communication. *Multisensory Research*, **29**, 49-91.

Roffler, S. K., & Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *Journal of the Acoustical Society of America*, **43**, 1255-1259.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, **86**, 638-641.

Rudmin, F., & Cappelli, M. (1983). Tone-taste synesthesia: A replication. *Perceptual & Motor Skills,* **56**, 118.

Rusconi, E., Kwana, B., Giordano, B. L., Umiltà, 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 direction signals have distinct influences on visual motion processing. *Journal of Neuroscience*, **29**, 6490-6499.

Schietecat, A. C., Lakens, D., IJsselsteijn, W. A., & de Kort, Y. A. (2018). Predicting context-dependent cross-modal associations with dimension-specific polarity attributions Part 1–Brightness and aggression. *Collabra: Psychology*, **4(1)**:14.

Schietecat, A. C., Lakens, D., IJsselsteijn, W. A., & de Kort, Y. A. W. (2018). Predicting context-dependent cross-modal associations with dimension-specific polarity attributions. Part 2: Red and valence. *Collabra: Psychology*, **4(1)**:21.

Shayan, S., Ozturk, O., & Sicoli, M. A. (2011). The thickness of pitch: Crossmodal metaphors in Farsi, Turkish, and Zapotec. *The Senses and Society*, **6(1)**, 96-105.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, **36**, 2345-2353.

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, **24**, 99-142.

Sonnadara, R. R., Gonzalez, D. A., Hansen, S., Elliott, D., & Lyons, J. L. (2009). Spatial properties of perceived pitch: Influence on reaching movements. *Annals of the New York Academy of Sciences*, **1169**, 503-507.

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

Spence, C. (2018). Crossmodal correspondences: A tutorial review. In D. Howes (Ed.), *Senses and sensation: Critical and primary sources* (Vol. III; pp. 91-125). London, UK: Bloomsbury Academic. https://www.bloomsbury.com/uk/senses-and-sensation-9781474274050/.

Spence, C. (submitted). Crossmodal correspondences involving musical stimuli. *Psychological Bulletin*.

Spence, C., & Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *i-Perception,* **3**, 316-318.

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition,* **22**, 245-260.

Spence, C., & Sathian, K. (in press). Audiovisual crossmodal correspondences: Behavioural consequences & neural underpinnings. To appear in K. Sathian & V. S. Ramachandran (Eds.), *Multisensory perception: From laboratory to clinic*. Elsevier.

Stekelenburg, J. J., & Keetels, M. (2016). The effect of synesthetic associations between the visual and auditory modalities on the Colavita effect. *Experimental Brain Research*, **234**, 1209-1219.

Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences*, **54**, 407-411.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**, 153-181.

Stevens, S. S., & Volkmann, J. (1940). The relation of pitch to frequency: A revised scale. *American Journal of Psychology*, **53**, 329-353.

Stumpf, K. (1883). *Tonpsychologie I* [Psychology of the tone]. Leipzig: Hirzel.

Trimble, O. C. (1934). Localization of sound in the anterior posterior and vertical dimensions of auditory space. *British Journal of Psychology*, **24**, 320-334.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Non-spatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, **34**, 1053-1065.

Velasco, C., Adams, C., Petit, O., & Spence, C. (2019). On the localization of tastes and tasty products in 2D space. *Food Quality & Preference*, **71**, 438-446.

Walker, L., & Walker, P. (2016). 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*, **42**, 138-150.

Walker, P. (2016). Cross-sensory correspondences: A theoretical framework and their relevance to music. *Psychomusicology: Music, Mind, and Brain*, **26**, 103-116.

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

Walker, P., Scallon, G., & Francis, B. (2017). Cross-sensory correspondences: Heaviness is dark and low-pitched. *Perception,* **46**, 772-792.

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., & Smith, S. (1986). The basis of Stroop interference involving the multimodal correlates of auditory pitch. *Perception*, **15**, 491-496.

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

Walsh, V. (2003). A theory of magnitude: Common cortical metrices of time, space and quality. *Trends in Cognitive Sciences*, **7**, 483-488.

Wells, A. (1980). Music and visual color: A proposed correlation. *Leonardo*, **13**, 101-107.

Wicker, F. W. (1968). Mapping the intersensory regions of perceptual space. *American Journal of Psychology*, **81**, 178-188.

Widmann, A., Kujala, T., Tervaniemi, M., Kujala, A., & Schröger, E. (2004). From symbols to sounds: Visual symbolic information activates sound representations. *Psychophysiology*, **41**, 709-715.

Xu, J., Yu, L., Rowland, B. A., Stanford, T. R., & Stein, B. E. (2012). Incorporating cross-modal statistics in the development and maintenance of multisensory integration. *Journal of Neuroscience*, **32**, 2287-2298.

Yau, J. M., Olenczak, J. B., Dammann, J. F., & Bensmaia, S. J. (2009). Temporal frequency channels are linked across audition and touch. *Current Biology,* **19**, 561-566.

FIGURE LEGENDS

Figure 1. Results highlighting the correlation between the pitch of a sound and the elevation of its source in the natural environment. Natural scene statistics and perceptual mappings. (A) External signals have their source in the environment, and then they are filtered by the transfer function of our sensory organs (the pinnae in this case) before being converted into neural activity in the brain. Any correlation between cues may already be present in the external signals or else might be introduced by the transfer functions of our sensory organs. The plots at the bottom of the figure represent the frequency–elevation mapping measured from natural scene statistics (B), the head-related transfer function (C), and the Bayesian priors representing the brain's belief about the mapping between frequency and elevation in head-centred and world-centred reference frames (D; for details, see Parise *et al*., 2014). This figure is published in colour in the online version. [Reprinted from Parise (2016), with permission.]

Figure 2. Schematic representation of a putative associative network of interconnected sensory cues. The thicker the connecting line, the stronger/more robust the correspondence. [Reprinted from Parise (2016), with permission.]

Figure 3. Mean reaction times (RTs) from Chiou and Rich's (2012, Experiment 2) speeded visual target detection task as a function of Delay Time (or stimulus onset asynchrony) and crossmodal congruency in the (a) 100 Hz difference and the (b) 1200 Hz difference conditions. Error bars represent $\pm$1 SEM. [Reprinted from Chiou & Rich (2012), Figure 3, with permission.]

Figure 1.

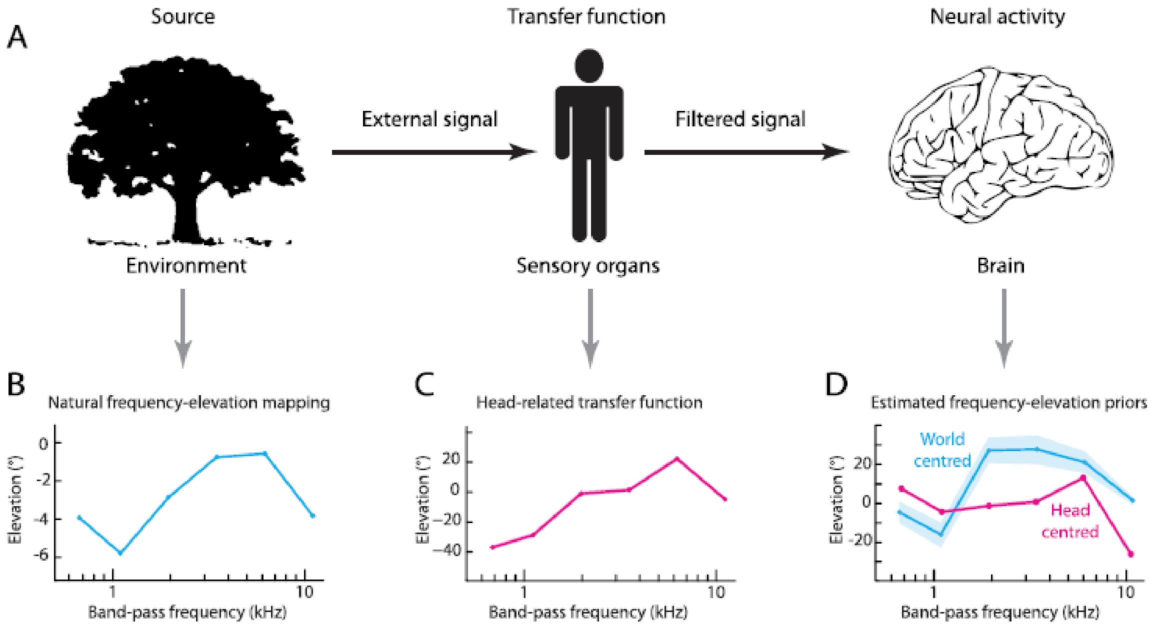

A
Source
Transfer function
Neural activity
External signal
Filtered signal
Environment
Sensory organs
Brain
B
Natural frequency-elevation mapping
Elevation (°)
Band-pass frequency (kHz)
C
Head-related transfer function
Elevation (°)
Band-pass frequency (kHz)
D
Estimated frequency-elevation priors
World centred
Head centred
Elevation (°)
Band-pass frequency (kHz)

Figure 2.

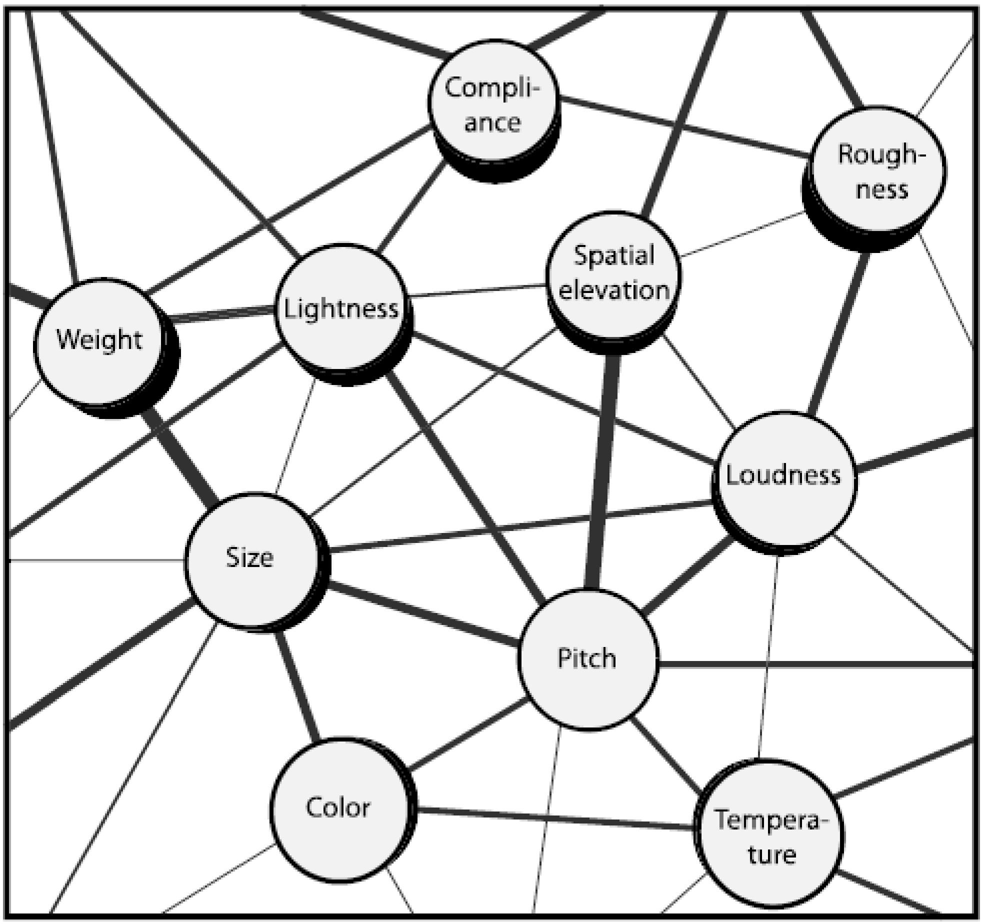
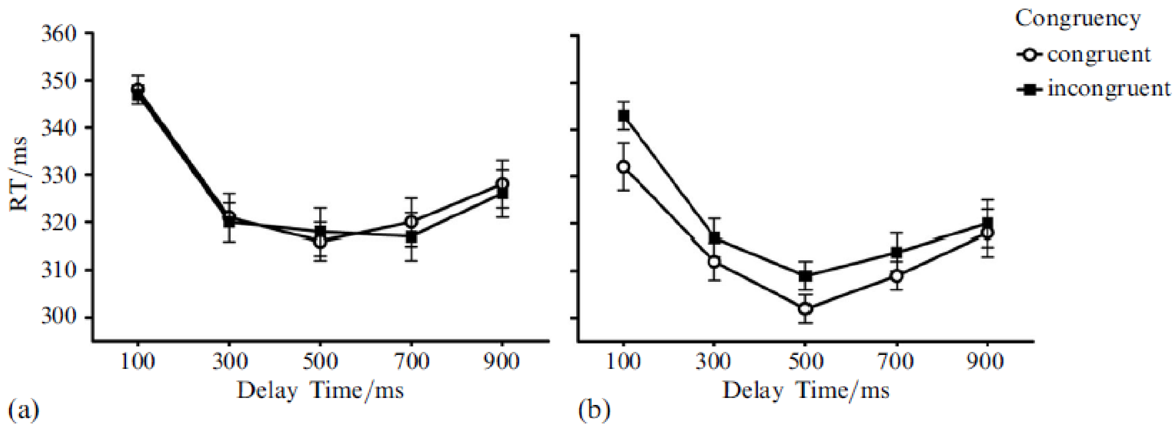

Compli-
ance
Rough-
ness
Spatial
elevation
Weight
Lightness
Loudness
Size
Pitch
Color
Tempera-
ture

Figure 3.


Congruency
congruent
incongruent
RT/ms
360
350
340
330
320
310
300
100
300
500
700
900
Delay Time/ms
(a)
(b)