



# Lightness/pitch and elevation/pitch crossmodal correspondences are low-level sensory effects

Mick Zeljko<sup>1</sup> · Ada Kritikos<sup>1</sup> · Philip M Grove<sup>1</sup>

Published online: 29 January 2019  
© The Psychonomic Society, Inc. 2019

## Abstract

We tested the sensory versus decisional origins of two established audiovisual crossmodal correspondences (CMCs; lightness/pitch and elevation/pitch), applying a signal discrimination paradigm to low-level stimulus features and controlling for attentional cueing. An audiovisual stimulus randomly varied along two visual dimensions (lightness: black/white; elevation: high/low) and one auditory dimension (pitch: high/low), and participants discriminated either only lightness, only elevation, or both lightness and elevation. The discrimination task and the stimulus duration varied between subjects. To investigate the influence of crossmodal congruency, we considered the effect of each CMC (lightness/pitch and elevation/pitch) on the sensitivity and criterion of each discrimination as a function of stimulus duration. There were three main findings. First, discrimination sensitivity was significantly higher for visual targets paired congruently (compared with incongruently) with tones while criterion was unaffected. Second, the sensitivity increase occurred for all stimulus durations, ruling out attentional cueing effects. Third, the sensitivity increase was feature specific such that only the CMC that related to the feature being discriminated influenced sensitivity (i.e. lightness congruency only influenced lightness discrimination and elevation congruency only influenced elevation discrimination in the single and dual task conditions). We suggest that these congruency effects reflect low-level sensory processes.

**Keywords** Crossmodal correspondence · Multisensory integration · Signal detection theory

Multisensory perception depends upon the brain solving a correspondence problem: which signals in the different sensory modalities have a common underlying cause, and so should be integrated, and which do not, and so should be segregated (Bien, ten Oever, Goebel, & Sack, 2012; Ernst & Bühlhoff, 2004; Gau & Noppeney, 2016; Parise, Harrar, Ernst, & Spence, 2013; Spence, 2011; Van Wanrooij, Bremen, & John Van Opstal, 2010). A strong indicator that cross-sensory signals likely have a common underlying cause is spatiotemporal coincidence. If signals in two or more modalities appear to come from the same location, at the same time, then they were likely caused by the same environmental event, and so should be integrated (Diaconescu, Alain, & McIntosh, 2011; Ernst & Bühlhoff, 2004; Evans & Treisman, 2010; Meredith, Nemitz, & Stein, 1987; Meredith & Stein, 1983, 1986). While

spatiotemporal coincidence is a strong integration cue, other cross-sensory stimulus features can act as subtle cues, and these may be particularly important if the spatiotemporal cues are weak or ambiguous (Bien et al., 2012; Hidaka, Teramoto, Keetels, & Vroomen, 2013; Kording et al., 2007; Parise & Spence, 2009, 2013). Crossmodal correspondences (CMCs) are considered such an integration cue (for reviews, see Parise, 2016; Spence, 2011; Spence & Deroy, 2013).

CMCs are observed relationships between specific stimulus features in different modalities, whereby certain stimulus values in each modality appear to associate preferentially compared with other stimulus values. For example, Marks (1987) demonstrated a CMC between visual brightness and auditory loudness. An audiovisual stimulus randomly varied between two values along each sensory dimension (bright or dim *and* loud or soft), and participants made a speeded response to one aspect of the stimulus (either brightness or loudness) while ignoring the other. Response times were significantly faster (and accuracy greater) when the stimulus pairs where either bright and loud *or* dim and soft compared with the alternative pairings of bright and soft *or* dim and loud. Importantly, although each stimulus value could be associated

---

✉ Mick Zeljko  
m.zeljko@uq.edu.au

<sup>1</sup> School of Psychology, The University of Queensland, St. Lucia, Queensland 4072, Australia

with both slowed and speeded response times, certain *pairings* were consistently associated with either slowed or speeded response times. The crossmodal feature pairings that elicited the behavioural benefit (i.e. bright/loud and dim/soft) Marks termed ‘matching’, and the alternative pairings (i.e. bright/soft and dim/loud) he termed ‘mismatching’.

CMCs have been demonstrated between numerous crossmodal feature pairs, in many experimental paradigms, using a range of measures, and the supporting evidence includes behavioural outcomes (e.g. Chiou & Rich, 2012; Evans & Treisman, 2010; Gallace & Spence, 2006; Marks, 1987) and electrophysiological findings (e.g. Bien et al., 2012). Typically, feature pairings that suggest a preferential association (e.g. an observable behavioural benefit) are termed congruent (rather than ‘matching’), the alternative pairings are incongruent (rather than ‘mismatching’), and the behavioural difference between congruent and incongruent pairings is termed a congruency effect. To date, audiovisual CMCs have been the most extensively investigated,<sup>1</sup> and auditory pitch has been found to correspond with several visual stimulus features. High (compared to low) pitch tones pair congruently with visually smaller (compared with larger) objects (e.g. Evans & Treisman, 2010), with visual stimuli higher (compared with lower) in the visual field (e.g. Bernstein & Edelman, 1971), and with whiter (compared with blacker) objects (e.g. Marks, 1987).

Congruency effects have been typically investigated in speeded response experiments and evidenced by faster detection, discrimination, or classification of a stimulus when it is paired with an irrelevant but congruent (compared with incongruent) crossmodal stimulus (e.g. Chiou & Rich, 2012; Evans & Treisman, 2010; Gallace & Spence, 2006; Marks, 1987). Additionally, congruency effects have been observed using signal detection theory (Marks, Ben-Artzi, & Lakatos, 2003; Mossbridge, Grabowecy, & Suzuki, 2011), in multisensory integration paradigms (Parise & Spence, 2008, 2009), in visual search (Klapetek, Ngo, & Spence, 2012), and in implicit association (e.g. Parise & Spence, 2012).

The motivation for the present study is that, despite many investigations using a range of approaches, there is no consensus about the mechanism underlying CMC congruency effects. CMCs may reflect sensory or decisional processes, crossmodally cued shifts of attention, response selection effects, or any combination of these (e.g. Gallace & Spence, 2006; Hidaka et al., 2013; Marks et al., 2003; Parise & Spence, 2012; Spence, 2011). Indeed, the characteristic congruency effect (faster response times) is potentially compatible with any of these accounts (e.g. Chiou & Rich, 2012; Evans & Treisman, 2010; Gallace & Spence, 2006; Marks,

1987). As noted in McDonald, Green, Störmer, and Hillyard (2012), “measures of response speed are inherently ambiguous in that RTs reflect the cumulative output of multiple stages of processing, including low-level sensory and intermediate perceptual stages, as well as later stages involved in making decisions and executing actions” (Chapter 26, p. 3). Consider the following:

First, multisensory integration studies provide compelling evidence that congruency effects rely, at least to some degree, on sensory/perceptual processing. Parise and Spence (2008, 2009) found evidence that congruency effects reflect enhanced multisensory integration which they suggested “reflects a genuine perceptual effect” (Parise & Spence, 2008, p. 260). Using a modified version of the ventriloquist paradigm (whereby spatially or temporally offset crossmodal stimuli are perceptually “pulled together” into a single multisensory event: e.g. Alais & Burr, 2004), Parise and Spence demonstrated that for size/pitch and shape/pitch, crossmodal congruency acted as a cue to promote multisensory integration. Presented with audiovisual stimulus combinations that were either temporally offset or spatially offset, the authors found that congruently paired stimuli were more likely to be “pulled into” a single spatiotemporal event, making discrimination of the offset (temporal or spatial) more difficult.

Next, two reliably demonstrated aspects of CMCs suggest a late cognitive influence on congruency effects. First, for congruency effects to emerge, multiple versions of each modality specific stimulus (e.g. high and low pitch, or bright and dim visual stimuli) must be presented randomly *within* a block of trials rather than separately blocked (Gallace & Spence, 2006; Klapetek et al., 2012; Melara & O’Brien, 1987). Gallace and Spence (2006) took this to reflect a postperceptual categorical process influencing congruency effects. Second, congruency effects rely on relative rather than absolute stimulus attributes (Chiou & Rich, 2012; Marks, 1987). For example, Chiou and Rich (2012), investigating the elevation/pitch CMC, found that when a 900-Hz tone was paired with a 100-Hz tone, the 900-Hz tone took on the value of ‘high’ and corresponded with high in the visual field as is usually the case. However, subsequently, when the 900-Hz tone was separately paired with a 1700-Hz tone, the 900-Hz tone took on the value of ‘low’ and corresponded with low in the visual field. Chiou and Rich took this dependence on the difference between the two tones rather than their absolute values to reflect a late, semantic influence on congruency effects (Chiou & Rich, 2012).

Finally, while one investigation using signal detection theory (SDT) concluded that congruency effects are decisional in nature (Marks et al., 2003), another has concluded that congruency effects are sensory and result from attentional cueing effects (Mossbridge et al., 2011). Marks et al. (2003) used SDT to examine the mechanisms underlying both the pitch/brightness and loudness/brightness audiovisual CMCs using

<sup>1</sup> Although audiovisual CMCs are the most extensively studied, congruency effects have been demonstrated between other crossmodal features including, for example, visual brightness and haptic size (Walker & Walker, 2016).

two sensory discrimination paradigms. Using a one-interval confidence-rating procedure, they found effects of visual stimulation on both auditory pitch and loudness discrimination, but not the reverse, and concluded that either sensory or decisional factors could account for the observed effects. Following this with a two-interval same/different procedure examining pitch/brightness, they found the opposite pattern: no effect of vision on pitch discrimination, but a small effect of audition on visual sensitivity in brightness discrimination. Although their findings were potentially compatible with either sensory or decisional accounts for congruency effects, Marks et al. suggested that when combined with other evidence, they pointed to response biases as underlying crossmodal congruency effects.

In contrast, Mossbridge et al. (2011) found a congruency effect on sensitivity but not criterion. Investigating a potential correspondence between the direction of frequency change and elevation in the visual field, Mossbridge et al. paired coloured stimuli that were either high or low in the visual field with tones with swept frequencies that either increased or decreased. They found that participants' sensitivity in a colour matching task was greater when the frequency sweep and the elevation were congruently paired (increasing frequency/high in the field or decreasing frequency/low in the field) compared with incongruent pairings (increasing frequency/low in the field or decreasing frequency/high in the field). Rather than reflecting a sensory process, Mossbridge et al. took their findings as reflecting a visual-spatial cueing effect.

To summarise, multisensory integration supports congruency effects being perceptual, the relative and categorical nature of congruency effects suggests some late cognitive influence, and the SDT findings support late decisional processes and sensory processes that may reflect attentional cueing. This issue of uncertainty regarding a mechanism is the focus of the present study, and we have three broad aims. First, using SDT in a visual discrimination paradigm, we aim to dissociate sensory from decisional factors underlying CMC congruency effects (Zeljko & Grove, 2017). Second, we aim to distinguish any sensory factor between purely bottom-up stimulus driven processes and top-down attentional cueing. Finally, we aim to discriminate low-level stimulus feature related sensory processes from higher level integrated object related perceptual processes. We next elaborate on each aim.

We first emphasise that a SDT discrimination task will dissociate sensory from decisional factors. In conditions of decision-making under uncertainty, SDT provides an analytical framework for independently categorising a decision makers' sensitivity to the information in an underlying signal with their criterion for making the decision (Green & Swets, 1966; Macmillan & Creelman, 2005). In a perceptual task, the sensitivity ( $d'$ ) and criterion ( $c$ ), are associated with sensory capacity and decisional strategy or bias, respectively (e.g. McDonald et al., 2012; McDonald, Teder-Sälejärvi, &

Hillyard, 2000; Odgaard, Arieh, & Marks, 2003). There are, however, caveats regarding these interpretations. Witt, Taylor, Sugovic, and Wixted (2015), for example, demonstrated that in a signal discrimination (vs detection) paradigm, perceptual effects (specifically, perceptual bias) can manifest as changes in criterion without any accompanying change in sensitivity. To preface our results, we find no criterion effects and so no longer consider this caveat.

A more important consideration is that, in certain paradigms, it is possible that differences in sensitivity could reflect differences in memory encoding. In fact, prior investigations have found that unisensory recall is improved for multisensory stimuli if the stimuli in each modality are semantically congruent compared to semantically incongruent (Thelen & Murray, 2013). We do not consider this literature relevant to the current study, however, since there has been no prior work that we are aware of that suggests a role of memory encoding in CMC congruency effects, and we are not sure how a memory account could explain the various congruency effects that have been observed by other investigators (e.g. RT effects, ventriloquist results). We agree with Marks et al. (2003) that "detection theory nevertheless offers a set of principles, as well as useful empirical procedures, for parcelling out effects on sensory processes, that is, effects on underlying sensory or perceptual representations, from effects on later decisional processes, that is, effects on criterion and judgment" (p. 142).

Next, we consider the issue of crossmodal attentional cueing. Stimuli in one modality have been shown repeatedly and reliably to direct attentional resources in another modality, both exogenously and endogenously. In exogenous cueing, a peripheral stimulus acts to direct spatial attention, providing performance benefits at the cued location (e.g. Posner, 1980; Posner & Cohen, 1984; McDonald et al., 2000; Störmer, McDonald, & Hillyard, 2009). Exogenous cueing is involuntary, stimulus driven, transient, and is thought to require at least 100 ms to be deployed (see Carrasco, 2011, for a review). In endogenous cueing, the cue is symbolic and can cue attention not only to a region of space, but also to a stimulus feature. For example, a centrally presented arrow can cue spatial attention to the direction that the arrow is pointing, thereby improving task performance at the cued location (e.g. Posner, 1980), or alternatively, a centrally presented cue can direct attention to a particular stimulus feature. In a visual-only paradigm, Andersen and Müller (2010) used a centrally presented colour cue to either one of two superimposed red or blue random dot kinematograms and found that reaction times were faster to the cued feature. Symbolic cues are voluntary and take about 200 ms to direct endogenous attention in a goal or conceptually driven fashion (e.g. van Ede, de Lange, & Maris, 2012).

It is possible that if CMC congruency effects are sensory and genuinely enhance perception, that they do so by a process of attentional cueing. For example, a high (or low) pitch tone may direct spatial attention endogenously high (or low) in the visual

field in a manner similar to a left-pointing (or right-pointing) arrow directing spatial attention to the left (or right). Similarly, a high (or low) pitch tone may direct attention to the visual feature of white (or black), or bright (or dim). Indeed, as already noted, Mossbridge et al. (2011) interpreted their SDT findings regarding changing pitch and visual elevation as reflecting visual-spatial cueing. Further, Chiou and Rich (2012) also suggested that, at least for the elevation/pitch CMC, it was plausible pitch was biasing spatial attention.

Critically, reorienting of attention can only improve discrimination of a stimulus if the stimulus is still present or easily retrieved from memory (van Ede et al., 2012). Therefore, to control for attentional cueing, we varied the duration of our stimuli, ranging from 50 ms to 250 ms to cover typical deployment times, and we backward masked the visual target to limit the accrual and persistence of stimulus information in memory.

Finally, if congruency effects are sensory in nature, then we are interested in whether they influence low-level sensory processing (e.g. basic feature processing in primary or early extrastriate visual cortex) or higher level perceptual processing (e.g. object-level processing in higher visual areas or association areas; e.g. Kandel, 2013). Essentially, if congruency effects provide a genuine perceptual enhancement, then is that enhancement limited to the specific feature that is part of the crossmodal correspondence (e.g. visual lightness or elevation), or does the enhancement extend beyond the specific feature to increase the salience of the broader stimulus more generally? Two prior CMC studies have prompted this question, each finding a congruency effect in relation to a stimulus feature that was not part of the CMC.

First, Evans and Treisman (2010), investigating bidirectional congruency effects between auditory pitch and visual size, elevation, and spatial frequency, found the typical effect of improved reaction time when participants were required to discriminate a stimulus feature relevant to the CMC (e.g. pitch or elevation). However, they also found a similar (although slightly smaller) congruency effect for the same CMCs, when instead the task was to discriminate a stimulus feature irrelevant to the CMC. For example, for stimuli in which pitch and elevation were manipulated, participants were faster to discriminate the orientation of a grating in the visual stimulus when elevation and pitch were congruently compared with incongruently paired. Similarly, Mossbridge et al. (2011), investigating a proposed CMC between frequency change and elevation in the visual field, found a congruency effect on sensitivity, but not on criterion in a colour matching task. That is, participants sensitivity to colour discrimination improved for targets whose elevation was congruently paired with frequency change (compared with incongruently paired), even though colour was unrelated to the CMC (frequency change/elevation).

We suggest that there are at least two possible sensory explanations compatible with these observed effects. If congruency effects are sensory in nature, they may reflect later perceptual effects, occurring *after* feature integration, such that enhancement of the CMC feature increases salience of the object more generally, allowing for improved performance. Alternatively, congruency effects may reflect low-level sensory effects, perceptually enhancing the specific feature only. In this case, improved performance in colour matching or grating discrimination would be due to a congruency benefit causing *later* attentional or response selection effects. The previous studies, each using stimuli of long duration, are potentially compatible with either of these alternatives.

In the present study, we test whether the mechanisms underlying two audiovisual CMCs (lightness/pitch and elevation/pitch)<sup>2</sup> are sensory or decisional in nature and, if they are sensory, we determine if they rely on attentional cueing effects and discriminate between low-level sensory processes and higher level processes. In an SDT paradigm, we presented participants with an audiovisual stimulus that randomly and independently varied along two visual dimensions (lightness and elevation) and one auditory dimension (pitch). Specifically, the visual component was either black or white (lightness variation) and presented either up or down relative to fixation (elevation variation), and the tone was either high or low (pitch variation). The combined audiovisual stimulus therefore could be either congruent or incongruent with respect to each CMC simultaneously. For example, black/up/high pitch would be simultaneously incongruent with respect to lightness/pitch (black/high) and congruent with respect to elevation/pitch (up/high). Critically, for each CMC, each congruency was counterbalanced with respect to the other CMC. So, for example, half of the lightness/pitch congruent trials were elevation/pitch congruent and half were elevation/pitch incongruent.

The visual component of the audiovisual stimulus was a circle presented in dynamic visual noise and the auditory component was a pure sine wave tone. Participants were tasked with discriminating only one of the visual features, either the lightness or the elevation about fixation, with approximately half of the participant pool allocated to each task condition. To control for attentional cueing effects, we masked the visual stimulus and varied the stimulus duration (from 50 ms to 250 ms in 50-ms increments) between subjects with approximately one-fifth of each task condition participant group randomly allocated to each duration condition. Our analysis considers the

<sup>2</sup> Our choice of CMCs was somewhat arbitrary since any pair of visual features that both exhibit a CMC with pitch could have reasonably been used in our paradigm. That said, we did aim to use visual features that shared a similar level of complexity (elevation and lightness are both low-level visual primitives, whereas size and angularity, for example, are higher level features), and we considered that low-level features would provide a good starting point for these investigations.

congruency effect of each CMC (lightness/pitch and elevation/pitch) on sensitivity and criterion measures of each discrimination as a function of stimulus duration.

Our predictions are as follows. First, we expect to find a congruency effect on sensitivity but not criterion, such that sensitivity in visual discrimination will be higher for visual stimuli that are paired with congruent (as compared with incongruent) but irrelevant auditory stimuli. Second, if attentional cueing makes a substantial contribution to this effect, then we expect a stimulus duration interaction such that the effect is lost for very short duration stimuli (50 ms). Third, if the congruency effect on sensitivity occurs at a low sensory level, then we expect it to be feature specific, so that only the CMC relevant to the discrimination task should cause a change in performance. For example, for lightness discrimination, we would expect lightness/pitch congruency, but not elevation/pitch congruency to influence sensitivity. To be clear, this contrasts with both Evans and Treisman (2010) and Mossbridge et al. (2011), who found that elevation/pitch congruency did change performance on the irrelevant tasks (grating and colour discrimination, respectively).

## Method

### Participants

Five hundred and forty-three psychology students at the University of Queensland (163 males, 376 females, and four of unspecified gender; age 20.6 years  $\pm$  3.9 years) participated in the experiment as part of an assignment exercise for an undergraduate psychology course. The experiment was conducted over two consecutive semesters, participation was during class time with classes of approximately 20 to 25 students participating concurrently under the supervision of a single tutor, and participation classes were spread over a week in each semester. Participation was voluntary, all participants were naïve as to the purpose of the experiment, and the experiment was cleared in accordance with the ethical review processes of the University of Queensland and within the guidelines of the National Statement on Ethical Conduct in Human Research.

Our sample size was not determined according to a priori considerations of effect size. Rather, it was simply determined by the number of students enrolled in the course for which the experiment was set as an assignment exercise. We do note, however, that we selected the number of duration conditions in an effort to ensure that, after an anticipated level of exclusion (see Results), the number of participants in each between-subjects condition would be in the order of what is typical for SDT experiments investigating these types of perceptual effects.

## Design

The experiment was a mixed-factorial design with participants conducting an unsped-up visual discrimination task on an audiovisual stimulus (visual component: a disc embedded in dynamic visual noise; auditory component: a pure tone). The position of the disc (elevation: above or below fixation), the colour of the disc (lightness: black or white), and the pitch of the tone (pitch: high or low) were varied randomly within subjects and within blocks. The duration of the audiovisual stimulus (duration: 50 ms, 100 ms, 150 ms, 200 ms, or 250 ms) was varied between subjects with participants randomly allocated to a duration condition. The visual property to be discriminated was varied between subjects, with each semester group performing a different task: The first semester cohort completed the lightness task (was the disc black or white;  $N = 304$ : 87 males, 215 females, and two of unspecified gender; age 20.3 years  $\pm$  3.2 years), and the second semester cohort completed the elevation task (was the disc up or down;  $N = 239$ : 76 males, 161 females, and two of unspecified gender; age 21.1 years  $\pm$  4.6 years).

### Stimuli and procedure

Stimuli were generated on a mix of PCs (Dell Optiplex 9030 AIO and 9010 AIO machines with 3.1 GHz Intel Core i7 CPUs, 8 GB RAM, and all running Microsoft Windows 7 Enterprise Version 6.1.7601 SP1 Build 7601) using MATLAB (R2015b, 2015) and the Psychophysics Toolbox extensions (Version 3.0.11; Brainard, 1997; Kleiner et al., 2007). Visual stimuli were viewed on Dell Optiplex displays (resolution 1,920  $\times$  1,080 pixels; 677  $\times$  381 mm), sounds were presented via participant-supplied earbuds or headphones (of various unrecorded makes and models), and responses were made using standard computer keyboards positioned directly in front of the participants. Participants were instructed to position themselves so that they were comfortable, their eyes were approximately 80 cm from the display, and their head position was unrestrained. They were further instructed to try and maintain this position for the duration of the experiment (approximately 15 minutes).

The visual stimulus was a disc (either black or white, 2 degrees in diameter) embedded in dynamic visual noise comprising a square (18 degrees  $\times$  18 degrees) of small elements (approximately 1.5 arcmin  $\times$  1.5 arcmin) that each randomly varied in monochrome intensity at 60 Hz). The dynamic visual noise was centred with the display centre, and the disc was horizontally centred and vertically offset 6 degrees either above or below a fixation cross (black, subtending 0.3 degrees) at the centre of the display. The screen background was grey. The auditory stimulus was a pure sinusoidal tone (either 600 Hz or 1800 Hz) sampled at 44.1 kHz and presented binaurally. Participants were instructed to set the volume so

that the tone was clearly audible, but comfortable. Because this was a visual discrimination task in the presence of an irrelevant tone, as long as the tone was obviously above threshold, we considered individual differences in sound level unimportant. In any case, if some participants set the volume too low for congruency effects to manifest, then it would only reduce any overall effect. To avoid ceiling effects, the disc was made partially transparent. The level of transparency set for each lightness (black or white) and each stimulus duration was based on pilot testing that aimed to enable average lightness discrimination with an accuracy of 70%.<sup>3</sup>

Participants were informed that they would be performing a computer-based visual discrimination task that would take approximately 15 minutes and that detailed instructions would be provided on-screen as the experiment progressed. After launching the experimental program, participants first entered basic demographic information (age, gender—female, male, or unspecified) and were then presented with four consecutive screens of detailed instructions.

All participants completed 256 trials split into four consecutive blocks with participant-controlled breaks between each block. Each trial consisted of the presentation of an audiovisual stimulus, with each of the following three parameters varied: disc lightness (lightness: black or white), disc elevation (elevation: up or down), and pitch of the tone (pitch: high or low). There were 32 trials of each audiovisual combination presented randomly. Stimulus duration (duration: 50 ms, 100 ms, 150 ms, 200 ms, or 250 ms) was randomly set by the experimental program upon commencement and remained fixed for each participant throughout the experiment. A practice block of 16 trials (two of each audiovisual combination, randomly presented), with the disc visibility well above threshold (i.e. transparency was very low making the task easy), was run before the four experimental blocks.

A single trial consisted of the following sequence of events. The fixation cross alone was visible for between 750 ms and 1,000 ms (the prestimulus interval: randomised across trials), after which the dynamic visual noise appeared (noise onset) and immediately began randomly updating at 60 Hz. After a randomly jittered interstimulus interval (733 ms, 817 ms, 900 ms, or 983 ms), the disc (either black or white) appeared either above or below fixation (disc onset) and the tone (either 600 Hz or 1800 Hz) commenced (tone onset). Disc onset and tone onset were simultaneous. The disc remained visible and the tone continued for the stimulus duration (either 50 ms, 100 ms, 150 ms, 200 ms, or 250 ms), after which they simultaneously offset. The dynamic visual noise remained visible

and continued to randomly update such that the total duration of the noise from onset to offset was 1,300 ms. After the offset of the dynamic visual noise, response instructions appeared and remained visible until a response was made. In the lightness task, participants were tasked with making a two-alternative forced-choice discrimination of disc lightness (i.e. determining if the disc was black or white). In the elevation task, participants were tasked with making a two-alternative forced-choice discrimination of disc elevation (i.e. was the disc up or down relative to fixation; see Fig. 1). Responses were untimed and made with the left and right arrows of a standard computer keyboard with the following task-specific mappings: lightness: left (black), right (white); elevation: left (up), right (down).

### Stimulus congruency

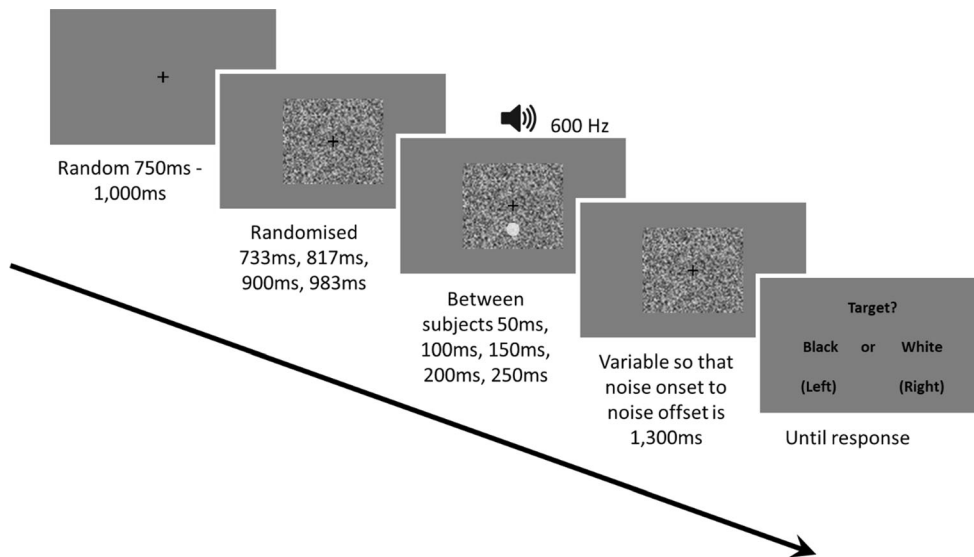
Each trial includes an audiovisual stimulus defined by the lightness (black or white) and elevation (up or down) of the disc, and the pitch of the tone (high or low). For the purposes of examining CMCs, we have defined each of the eight trial types in terms of *both* its lightness/pitch congruency *and* its elevation/pitch congruency (see Table 1).

To preempt our analysis, we will consider how the signal detection theory parameters sensitivity ( $d'$ ) and criterion ( $c$ ) vary with stimulus congruency. That is, we will compare  $d'$  and  $c$  for congruent versus incongruent trials for each CMC (lightness/pitch and elevation/pitch) separately. Consider the trial parameters (stimulus features and congruencies) as listed in Table 1. If we group trials by lightness/pitch congruency (Trial Types 2, 4, 5, and 7 are congruent whereas Trial Types 1, 3, 6, and 8 are incongruent), we see that within each congruency group all other trial parameters are internally counterbalanced. The lightness/pitch congruent grouping has equal proportions of black and white targets, up and down targets, high and low tones, and elevation/pitch congruent and incongruent pairs (and similarly for the lightness/pitch incongruent grouping). Additionally, all these trial parameters are equally represented in each congruency group. So, in considering, for example, lightness/pitch congruent versus incongruent trials, the analysis is independent of all other trial parameter variations, both in stimulus features and in the other congruency, and therefore provides results for this congruency condition in isolation.

### Signal detection theory

We conducted our analyses based on signal detection theory (SDT; Green & Swets, 1966; Macmillan & Creelman, 2005), arbitrarily defining white as the target for the lightness task and down as the target for the elevation task. We then classified individual trials depending on the task, the stimulus, and the response as either a HIT (for the lightness task if stimulus

<sup>3</sup> We omit the details of the pilot testing because absolute transparency of the stimuli is irrelevant to our analysis. In theory, the transparencies could be arbitrarily set without compromising our comparisons. As long as average sensitivities are reasonable (we aimed for  $d'$  values between 1 and 2) and are substantially similar across conditions (again,  $d'$  on average between 1 and 2), our analyses are valid.



**Fig. 1** Sample stimulus sequence for the lightness task (lightness: white; elevation: down; pitch: low). The sequence was similar for the elevation task, except that the final response screen displayed “up or down” instead of “black or white”

= white and response = white; for the elevation task if stimulus = down and response = down), a MISS (for the lightness task if stimulus = white and response = black; for the elevation task if stimulus = down and response = up), a FALSE ALARM (for the lightness task if stimulus = black and response = white; for the elevation task if stimulus = up and response = down), or a CORRECT REJECTION (for the lightness task if stimulus = black and response = black; for the elevation task if stimulus = up and response = up). We tabulated participant hit and false-alarm rates for five groupings of trials (overall, lightness/pitch congruent, lightness/pitch incongruent, elevation/pitch congruent, and elevation/pitch incongruent), and from these data we calculated individual sensitivity ( $d'$ ) and criterion ( $c$ ) measures for each trial grouping using Equations 1 and 2 below, where  $Z$  is the Gaussian inverse distribution function:

$$d' = Z(Hr) - Z(FAr) \tag{1}$$

**Table 1** Lightness/pitch and elevation/pitch congruencies for the eight trial types of varying lightness (black or white), elevation (up or down) and pitch (high or low)

Trial type	Lightness	Elevation	Tone	Lightness /pitch	Elevation /pitch
1	Black	Up	High	Incongruent	Congruent
2	Black	Up	Low	Congruent	Incongruent
3	Black	Down	High	Incongruent	Incongruent
4	Black	Down	Low	Congruent	Congruent
5	White	Up	High	Congruent	Congruent
6	White	Up	Low	Incongruent	Incongruent
7	White	Down	High	Congruent	Incongruent
8	White	Down	Low	Incongruent	Congruent

$$c = -0.5 \times (Z(Hr) + Z(FAr)). \tag{2}$$

## Results

### Participant screening

A  $d'$  measure of zero characterises chance performance while less than zero indicates that the probability of a ‘yes’ response is higher for trials in which the target stimulus is absent than for those where it is present.<sup>4</sup> Such results reflect poor task compliance and therefore justify removal from the data set for further analysis.

To allow for some variability, we have consequently defined chance target detection as being indicated by  $d' = 0 \pm 0.1$  and excluded participants from the analysis if the overall target detection was at chance or worse (overall  $d' < 0.1$ ) or detection in any of the four congruency conditions was worse than chance (any congruency  $d' < -0.1$ ). We have permitted chance detection levels in a congruency condition to account for the possibility that a participant may be at chance levels in one congruency condition (e.g. lightness/pitch incongruent) and above chance in the corresponding congruency alternative (e.g. lightness/pitch congruent).

Following participant screening, we excluded 82 participants from the lightness task (final  $N = 43, 45, 34, 34,$  and  $66$  for each stimulus duration condition, respectively) and 38 participants from the elevation task (final  $N = 62, 52, 40, 49,$  and  $39$  for each stimulus duration condition, respectively).

<sup>4</sup> Macmillan and Creelman (2005) note that  $d'$  has the mathematical properties of a distance measure, including positivity; although negative values can occur by chance, “ $d'$  should not be negative in the long run” (p. 14).

## SDT analysis

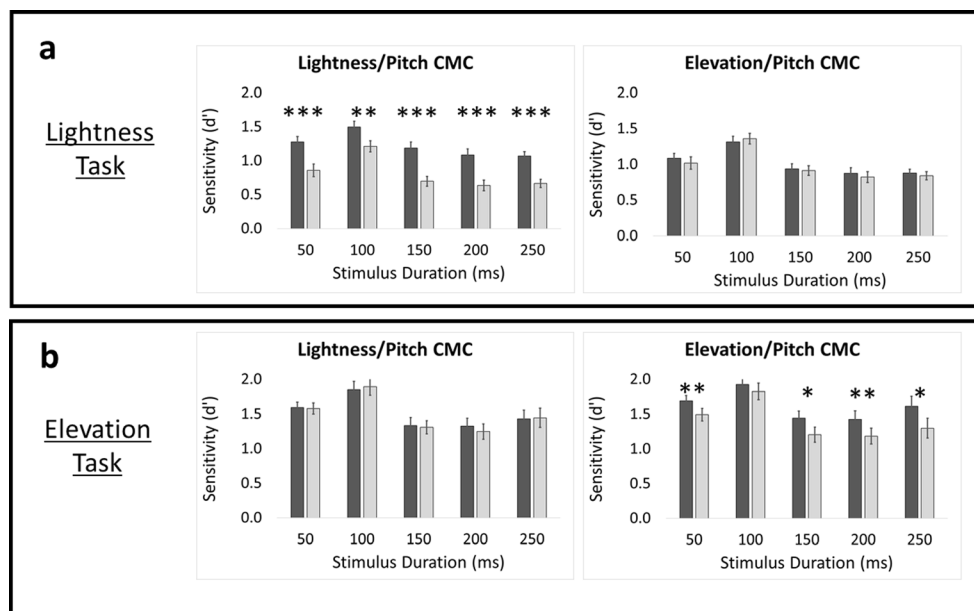
To examine the effect of crossmodal congruency for each duration and CMC (lightness/pitch and elevation/pitch) within each task (lightness and elevation), we conducted four separate mixed-factorial analyses of variance (ANOVAs): one for the sensitivity ( $d'$ ) and one for the criterion ( $c$ ) for each task group (lightness and elevation). We first determined the  $d'$  and  $c$  values for each participant for each congruency condition and each CMC (resulting in four  $d'$  values and four  $c$  values for each participant), and these were the units for statistical analyses. Each ANOVA is a  $2 \times 2 \times 5$  (congruency: congruent or incongruent; CMC: lightness/pitch or elevation/pitch; duration: 50 ms, 100 ms, 150 ms, 200 ms, or 250 ms), with congruency and CMC being within-subjects variables and duration a between-subjects variable.

For participant sensitivity ( $d'$ ) in the lightness task, we found significant main effects of congruency,  $F(1, 217) = 95.48, p < .001, \eta_p^2 = 0.306$ ; CMC,  $F(1, 217) = 49.90, p < .001, \eta_p^2 = 0.187$ ; and duration,  $F(4, 217) = 9.76, p < .001, \eta_p^2 = 0.152$ ; and a significant interaction between congruency and CMC,  $F(1, 217) = 71.21, p < .001, \eta_p^2 = 0.247$ . No other interactions reached significance: Congruency  $\times$  Duration,  $F(4, 217) = 1.33, p = .262, \eta_p^2 = 0.024$ ; CMC  $\times$  Duration,  $F(4, 217) = 0.68, p = .608, \eta_p^2 = 0.012$ ; Congruency  $\times$  CMC  $\times$  Duration,  $F(4, 217) = 0.247, p = .911, \eta_p^2 = 0.005$ . For participant sensitivity ( $d'$ ) in the elevation task, we found significant main effects of congruency,  $F(1, 237) = 23.64, p < .001, \eta_p^2 = 0.091$ , and duration,  $F(4, 237) = 4.98, p = .001, \eta_p^2 = 0.077$ , and a significant interaction between congruency and CMC,  $F(1, 237) = 14.27, p < .001, \eta_p^2 = 0.057$ . The main effect

of CMC failed to reach significance, CMC,  $F(1, 237) = 3.61, p = .059, \eta_p^2 = 0.015$ , and there were no other significant interactions: Congruency  $\times$  Duration,  $F(4, 237) = 0.69, p = .597, \eta_p^2 = 0.012$ ; CMC  $\times$  Duration,  $F(4, 237) = 0.44, p = .780, \eta_p^2 = 0.007$ ; Congruency  $\times$  CMC  $\times$  Duration,  $F(4, 237) = 0.85, p = .493, \eta_p^2 = 0.014$ . We note that the pattern of results for each task are similar, except that the main effect of CMC failed to reach significance in the elevation task and the effect sizes are smaller for the elevation task, which we discuss below. For group mean  $d'$  results for each task, see Fig. 2.

There were no significant main effects or interactions for participant criterion ( $c$ ) for either the lightness task: congruency,  $F(1, 217) = 0.84, p = .362, \eta_p^2 = 0.004$ ; CMC:  $F(1, 217) = 1.22, p = .270, \eta_p^2 = 0.006$ ; duration,  $F(4, 217) = 1.65, p = .164, \eta_p^2 = 0.029$ ; Congruency  $\times$  CMC,  $F(1, 217) = 0.58, p = .444, \eta_p^2 = 0.003$ ; Congruency  $\times$  Duration,  $F(4, 217) = 1.53, p = .195, \eta_p^2 = 0.027$ ; CMC  $\times$  Duration,  $F(4, 217) = 0.42, p = .792, \eta_p^2 = 0.008$ ; Congruency  $\times$  CMC  $\times$  Duration,  $F(4, 217) = 0.35, p = .843, \eta_p^2 = 0.006$ , or the elevation task: congruency,  $F(1, 237) = 0.01, p = .916, \eta_p^2 = 0.000$ ; CMC,  $F(1, 237) = 0.15, p = .704, \eta_p^2 = 0.001$ ; duration,  $F(4, 237) = 0.88, p = .478, \eta_p^2 = 0.015$ ; Congruency  $\times$  CMC,  $F(1, 237) = 0.43, p = .512, \eta_p^2 = 0.002$ ; Congruency  $\times$  Duration,  $F(4, 237) = 0.11, p = .978, \eta_p^2 = 0.002$ ; CMC  $\times$  Duration,  $F(4, 237) = 1.59, p = .177, \eta_p^2 = 0.026$ ; Congruency  $\times$  CMC  $\times$  Duration,  $F(4, 237) = 0.30, p = .877, \eta_p^2 = 0.005$ . For group mean criterion results for each task, see Fig. 3.

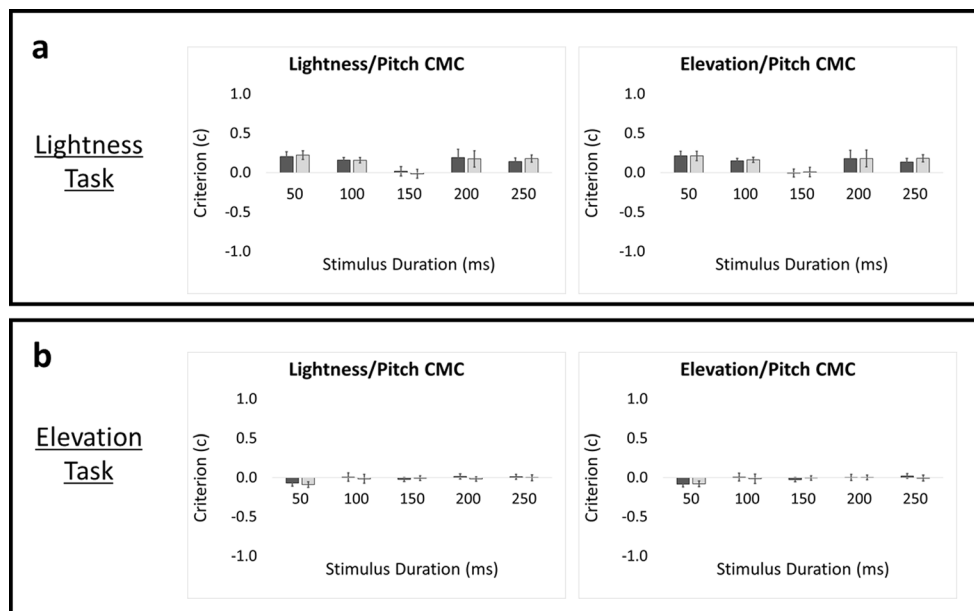
Our primary interest being the duration and task dependencies of any congruency effect, we followed up with planned



**Fig. 2** Group mean sensitivity ( $d'$ ) ( $\pm$ SEM) for congruent (dark bars) versus incongruent (light bars) stimulus pairings for each stimulus duration for each CMC (lightness/pitch and elevation/pitch) for **a** the

lightness task and **b** the elevation task. Asterisks indicate significance: \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$





**Fig. 3** Group mean criterion ( $c$ ) ( $\pm SEM$ ) for congruent (dark bars) versus incongruent (light bars) stimulus pairings for each stimulus duration for each CMC (lightness/pitch and elevation/pitch) for **a** the lightness task and **b** the elevation task

paired  $t$  tests, comparing the sensitivity ( $d'$ ) for congruent versus incongruent stimulus pairings in each duration condition within each CMC and for each task. For the lightness task, the sensitivity ( $d'$ ) was significantly higher for congruent stimuli compared with incongruent stimuli for all stimulus durations for the lightness/pitch CMC (see Table 2), but for none of the stimulus durations for the elevation/pitch CMC. We found the reverse for the elevation task, and sensitivity ( $d'$ ) was significantly higher for congruent stimuli compared with incongruent stimuli for all but one of the stimulus durations for the elevation/pitch CMC (see Table 3), but for none of the stimulus durations for the lightness/pitch CMC. We note that although effect sizes were small for some highly significant main effects and interactions in the ANOVA analysis (due to the very large sample sizes), the effect sizes of our main findings (i.e. the pairwise comparisons of  $d'$  in congruent vs incongruent trials) range from medium to very large, as evidenced by the Cohen's  $d$  measures.

To summarise, these results reveal three findings. First, we found a significant effect of congruency on

participants' sensitivity to discriminate both visual lightness and elevation in the visual field, with congruent stimulus pairings resulting in increased sensitivity, but there was no effect on criterion. Second, the congruency effect was entirely task dependent. When the instructed task was to discriminate lightness, a congruency effect was only seen for congruent versus incongruent lightness/pitch stimulus pairs and not for elevation/pitch congruency. Similarly, when the instructed task was to discriminate elevation, the congruency effect was only seen for elevation/pitch congruency and was abolished for lightness/pitch congruency. Third, the observed congruency effect was independent of stimulus duration in both tasks. We note that for the elevation/pitch CMC in the elevation task, the difference in sensitivity for congruent versus incongruent stimuli in the 100-ms stimulus duration condition failed to reach significance. We suggest that this may be due to the elevated overall sensitivity in this condition (around  $d' = 2.3$ ). While not at a level that could be described as ceiling, it is possibly high enough

**Table 2** Lightness task, lightness/pitch CMC sensitivity ( $d'$ ) paired-sample  $t$ -test comparisons for congruent versus incongruent stimulus pairings

Duration	Congruent	Incongruent	$t$	$df$	$p$	Cohen's $d$
50	1.23	0.86	4.37	42	<.001	0.66
100	1.50	1.21	3.31	44	.002	0.49
150	1.19	0.70	5.23	33	<.001	0.90
200	1.08	0.64	5.20	33	<.001	0.89
250	1.07	0.67	6.57	65	<.001	0.81

**Table 3** Elevation task, elevation/pitch CMC sensitivity ( $d'$ ) paired-sample  $t$ -test comparisons for congruent versus incongruent stimulus pairings

Duration	Congruent	Incongruent	$t$	$df$	$p$	Cohen's $d$
50	1.77	1.56	3.17	61	.002	0.40
100	2.36	2.29	0.91	51	.370	–
150	1.83	1.64	2.47	39	.018	0.39
200	1.67	1.46	2.74	48	.009	0.39
250	1.88	1.58	2.41	38	.021	0.39

that a subtle effect like CMC congruency fails to have an impact. In fact, possibly for the same reason the congruency effect in the elevation/pitch CMC (elevation task) is less than that in the lightness/pitch CMC (lightness task), across all duration conditions. This overall pattern is because we set the visual target transparency based only on pilot testing of the lightness task and held it constant for both tasks (for a given transparency, it will be easier to discriminate elevation than lightness).

Although the task dependence finding agrees with our hypothesis regarding congruency effects being feature specific and therefore reflective of low-level sensory processes, we note a potential confound. Rather than reflecting a feature-specific aspect to congruency effects, our task dependence finding may simply reflect the strong task demands induced in each discrimination task. For example, the fact that elevation/pitch congruency did not influence lightness discrimination may be because in the lightness task, participants are so focussed on the task-relevant visual feature (lightness) that elevation is effectively ignored and a potential elevation/pitch congruency has nothing to act on.

To address this issue, we next present a dual task follow-up experiment, wherein both visual discriminations (lightness and elevation) must be made on each trial, thereby making both congruencies task relevant. If the congruency effects are in fact feature specific, then the pattern of results seen in the single task discriminations should also be seen in the dual task discrimination, and each congruency should only influence the relevant discrimination. However, if task demands are driving the observed task dependency, then in a dual task paradigm each CMC should influence each discrimination, since both lightness and elevation are task relevant to the participant in each trial.

### Dual task follow-up experiment

The stimuli, design, and procedure for the dual task experiment were identical to those used for the two single task experiments (lightness discrimination and elevation discrimination), except as noted here.

One hundred and ninety-five psychology students at the University of Queensland (55 males, 137 females, and three of unspecified gender; age 20.9 years  $\pm$  4.5 years) participated in the experiment, and none had participated in either of the two single task experiments. As in the single task experiments, all participants viewed the same audiovisual stimulus; however, we reduced the between-subjects duration conditions from five to two (50 ms and 250 ms). Given the new task, stimulus duration remained a factor that we wished to examine; however, based on the lack of duration interactions in the single task experiments, we reasoned that restricting the examination to only the extreme durations would be adequate. One

hundred and four participants were allocated to the 50-ms duration condition and 91 to the 250-ms duration condition.

The key difference in the dual task experiment is that, in each trial, all participants were required to discriminate both the lightness and the elevation of the visual target. The two discrimination tasks were prompted sequentially following offset of the dynamic visual noise, and the order of the questions was counterbalanced between subjects (i.e. each subject answered every trial in the same order). Responses were made using pairs of keys (< and > with the right hand for one task, and z and x with the left hand for the other task). The mapping between key pair and task was counterbalanced between subjects, but the left key of each pair always corresponded to *black* (with the right key *white*) and *up* (with the right key *down*).

## Results

### Participant screening

We again screened participants and excluded those responding at chance levels or worse (an overall  $d' < 0.1$ ) or if detection on any of the four congruency conditions was worse than chance (any congruency  $d' < -0.1$ ). We excluded 12 participants in total, one from the 50-ms duration condition (final  $N = 103$ ) and 11 from the 250-ms duration condition (final  $N = 80$ ).

### SDT analysis

To examine the effect of crossmodal congruency for each discrimination (lightness and elevation), each duration (50 ms and 250 ms), and each CMC (lightness/pitch and elevation/pitch), we conducted two separate mixed-factorial analyses of variance (ANOVAs): one for the sensitivity ( $d'$ ) and one for the criterion ( $c$ ). Each ANOVA is a  $2 \times 2 \times 2 \times 2$  (task: lightness or elevation; congruency: congruent or incongruent; CMC: lightness/pitch or elevation/pitch; duration: 50 ms or 250 ms) with task, congruency and CMC being within-subjects variables and duration a between-subjects variable.

For participant sensitivity ( $d'$ ) we found significant main effects for task,  $F(1, 181) = 338.15, p < .001, \eta_p^2 = 0.651$ ; congruency,  $F(1, 181) = 30.23, p < .001, \eta_p^2 = 0.143$ ; and duration,  $F(1, 181) = 37.72, p < .001, \eta_p^2 = 0.089$ ; and significant interactions between task and duration,  $F(1, 181) = 18.46, p < .001, \eta_p^2 = 0.093$ , and task, CMC, and congruency,  $F(1, 181) = 46.50, p < .001, \eta_p^2 = 0.204$ . No further main effects or interactions reached significance. CMC,  $F(1, 181) = 0.10, p = .748, \eta_p^2 = 0.001$ ; CMC  $\times$  Duration,  $F(1, 181) = 0.31, p = .748, \eta_p^2 = 0.001$ ; Congruency  $\times$  Duration,  $F(1, 181) = 0.90, p = .345, \eta_p^2 = 0.005$ ; Task  $\times$  CMC,  $F(1, 181) = 0.67, p = .415, \eta_p^2 = 0.004$ ; Task  $\times$  CMC  $\times$  Duration,  $F(1, 181) = 0.18, p = .672, \eta_p^2 = 0.001$ ; Task  $\times$  Congruency,  $F(1, 181) = 0.01, p = .928, \eta_p^2 < 0.001$ ; Task  $\times$  Congruency  $\times$  Duration,

$F(1, 181) = 0.46, p = .499, \eta_p^2 = 0.003$ ; CMC  $\times$  Congruency,  $F(1, 181) = 2.63, p = .107, \eta_p^2 = 0.014$ ; CMC  $\times$  Congruency  $\times$  Duration,  $F(1, 181) = 1.18, p = .280, \eta_p^2 = 0.006$ ; Task  $\times$  CMC  $\times$  Congruency  $\times$  Duration,  $F(1, 181) = 0.33, p = .569, \eta_p^2 = 0.002$ . For group mean  $d'$  results for each task, see Fig. 4.

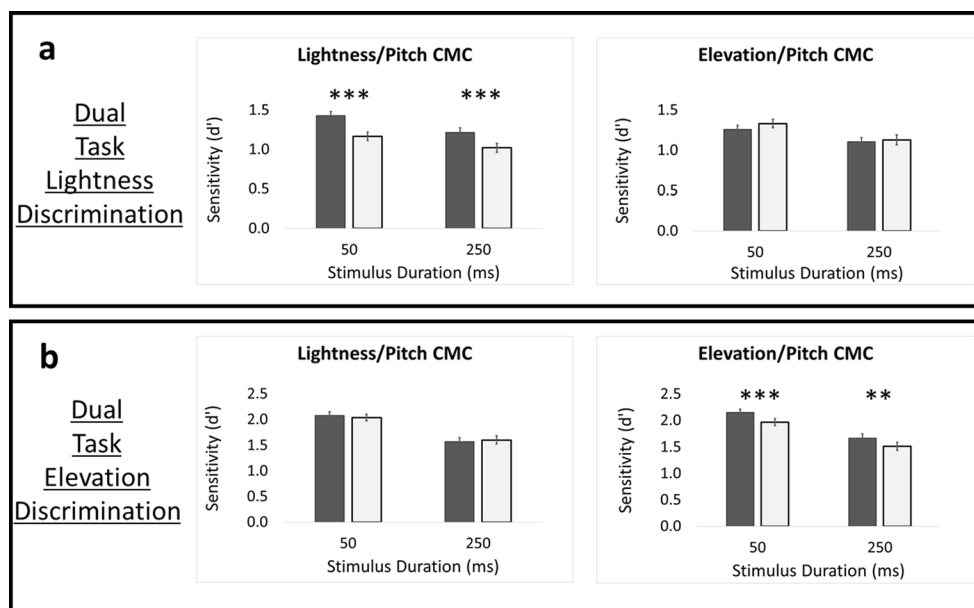
For participant criterion (c), we found a significant main effect of task,  $F(1, 181) = 101.58, p < .001, \eta_p^2 = 0.359$ , and a significant interaction between task and duration,  $F(1, 181) = 32.30, p < .001, \eta_p^2 = 0.151$ . No other main effects or interactions reached significance: CMC,  $F(1, 181) < 0.01, p = .974, \eta_p^2 < 0.001$ ; congruency,  $F(1, 181) = 0.50, p = .479, \eta_p^2 = 0.003$ ; duration,  $F(1, 181) = 0.46, p = .497, \eta_p^2 = 0.003$ ; CMC  $\times$  Duration,  $F(1, 181) = 0.68, p = .411, \eta_p^2 = 0.004$ ; Congruency  $\times$  Duration,  $F(1, 181) = 0.45, p = .501, \eta_p^2 = 0.003$ ; Task  $\times$  CMC,  $F(1, 181) = 0.07, p = .797, \eta_p^2 < 0.001$ ; Task  $\times$  CMC  $\times$  Duration,  $F(1, 181) = 0.04, p = .838, \eta_p^2 < 0.001$ ; Task  $\times$  Congruency,  $F(1, 181) = 0.11, p = .737, \eta_p^2 = 0.001$ ; Task  $\times$  Congruency  $\times$  Duration,  $F(1, 181) = 2.15, p = .144, \eta_p^2 = 0.012$ ; CMC  $\times$  Congruency,  $F(1, 181) = 3.06, p = .082, \eta_p^2 = 0.017$ ; CMC  $\times$  Congruency  $\times$  Duration,  $F(1, 181) = 0.01, p = .938, \eta_p^2 < 0.001$ ; Task  $\times$  CMC  $\times$  Congruency,  $F(1, 181) = 0.23, p = 0.634, \eta_p^2 = 0.001$ ; Task  $\times$  CMC  $\times$  Congruency  $\times$  Duration,  $F(1, 181) = 0.47, p = 0.494, \eta_p^2 = 0.003$ . For group mean criterion results for each task, see Fig. 5.

Once again, our primary interest being the duration and task dependencies of any congruency effect, we followed up with planned paired  $t$  tests, comparing the sensitivity ( $d'$ ) for congruent versus incongruent stimulus pairings in each duration condition within each CMC and for each task. For the lightness task, the sensitivity ( $d'$ ) was significantly higher for congruent stimuli compared with incongruent stimuli for both

stimulus durations for the lightness/pitch CMC (see Table 4), but for neither of the stimulus durations for the elevation/pitch CMC. We found the reverse for the elevation task, and sensitivity ( $d'$ ) was significantly higher for congruent stimuli compared with incongruent stimuli for both of the stimulus durations for the elevation/pitch CMC (see Table 5), but for neither of the stimulus durations for the lightness/pitch CMC. As in the single task experiments, we again note that although some of the highly significant main effects and interactions exhibit very small effect sizes (due to the very large sample size), the effect sizes of the main findings (i.e. the pairwise comparisons of  $d'$  for congruent versus incongruent trials) are in the medium range, as evidenced by the Cohen's  $d$  measures.

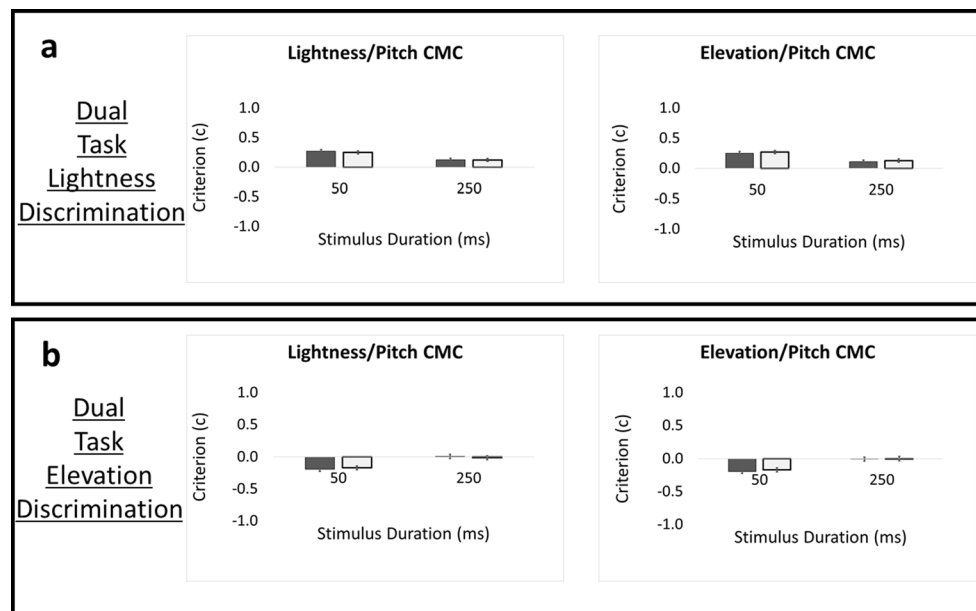
Regarding the significant Task  $\times$  Duration interactions, we note two points. First, in each duration condition, the average  $d'$  for elevation discrimination is higher than the average  $d'$  for lightness discrimination since, as we suggested earlier, it is easier to discriminate the location of a given stimulus than its lightness. Second, despite our attempt to set the visual target transparency for each duration to hold lightness discrimination (on average) constant for both 50-ms and 250-ms presentations, the 250-ms condition was more difficult as reflected by lower  $d'$  values (50 ms:  $M = 1.68, SE = 0.05$ ; 250 ms:  $M = 1.34, SE = 0.06$ ). We suggest that the finding that two different tasks (elevation and lightness discrimination) are differentially influenced by changes in stimulus visibility and presentation time is not relevant to the influence of CMC congruency and so will not be analysed further.

To summarise, the dual task experiment yielded the same pattern of results as the single task experiments combined. That is, significant congruency effects on sensitivity only, that



**Fig. 4** Group mean sensitivity ( $d'$ ) ( $\pm$ SEM) for congruent (dark bars) versus incongruent (light bars) stimulus pairings for each stimulus duration for each CMC (lightness/pitch and elevation/pitch) for **a** the

lightness task and **b** the elevation task. Asterisks indicate significance: \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$



**Fig. 5** Group mean criterion ( $c$ ) ( $\pm SEM$ ) for congruent (dark bars) versus incongruent (light bars) stimulus pairings for each stimulus duration for each CMC (lightness/pitch and elevation/pitch) for **a** the lightness task and **b** the elevation task

are of medium effect size, in both the short and long stimulus duration conditions, for both lightness discrimination and elevation discrimination. Critically, the effect of a particular CMC remained discrimination dependent despite both discriminations being made on each trial. To be clear, on each trial the task was to discriminate both lightness and elevation, making both features simultaneously task relevant, and both lightness/pitch and elevation/pitch congruency varied independently. While lightness/pitch congruency improved lightness discrimination, it had no influence on elevation discrimination. Similarly, while elevation/pitch congruency improved elevation discrimination, it had no influence on lightness discrimination. The congruency effects appear to be feature specific rather than reflecting task demands.

## Discussion

Our aim in this study was to investigate the sensory versus decisional origins of two audiovisual CMCs (lightness/pitch and elevation/pitch) in a signal detection discrimination paradigm while controlling for attentional cueing effects and

dissociating low-level sensory from higher level perceptual effects. We considered congruency effects on both sensitivity and criterion of each discrimination as a function of stimulus duration, and we proposed three hypotheses.

First, we expected to observe a congruency effect on discrimination sensitivity but not on response criterion. This hypothesis was confirmed, and visual discrimination sensitivity was significantly higher for visual targets paired congruently (compared with incongruently) with auditory tones while criterion was unaffected. Further, this congruency effect was observed for both audiovisual CMCs investigated (lightness/pitch and elevation/pitch), and the effect sizes were moderate to high.

Our second hypothesis was that if attentional cueing rather than sensory effects were causing the increased discrimination sensitivity, then the sensitivity benefit should be lost for very short stimulus durations. We reasoned that there would not be enough time for a simultaneous but brief auditory stimulus to be first identified as high or low pitch, interpreted in terms of the appropriate CMC and associated with the relevant visual feature (light or dark, or up or down), and then to cue attention to that visual feature before the masked visual stimulus had vanished. We found no significant duration interactions for

**Table 4** Dual task: Lightness discrimination, lightness/pitch CMC sensitivity ( $d'$ ) paired-sample  $t$ -test comparisons for congruent versus incongruent stimulus pairings

Duration	Congruent	Incongruent	$t$	$df$	$p$	Cohen's $d$
50	1.43	1.17	-5.52	102	<.001	0.54
250	1.22	1.02	-4.00	79	<.001	0.48

**Table 5** Dual task: Elevation discrimination, elevation/pitch CMC sensitivity ( $d'$ ) paired-sample  $t$ -test comparisons for congruent versus incongruent stimulus pairings

Duration	Congruent	Incongruent	$t$	$df$	$p$	Cohen's $d$
50	2.16	1.97	-4.08	102	<.001	0.40
250	1.67	1.52	-2.70	79	.009	0.30

either the lightness/pitch CMC or the elevation/pitch CMC. The sensitivity increase for congruent versus incongruent stimulus pairings was statistically equivalent across all stimulus durations from 50 ms to 250 ms.

Finally, we hypothesised that a purely low-level sensory effect should be feature specific, so, for example, lightness/pitch congruency should improve lightness discrimination but not elevation discrimination. This was confirmed, and we found that congruency effects for each CMC depended entirely on the feature being discriminated both in the single task experiments and in the follow-up dual task experiment (which ruled out an influence of task demand). Specifically, the audiovisual stimuli varied independently in both lightness/pitch and elevation/pitch congruency, regardless of the task. However, only lightness/pitch congruency, but not elevation/pitch congruency improved lightness discrimination. Similarly, only elevation/pitch congruency, but not lightness/pitch congruency, benefited elevation discrimination.

Our interpretation of these findings is that the CMC congruency effects investigated in this study (lightness/pitch and elevation/pitch) reflect a crossmodal influence on low-level sensory processing. First, congruency effects reflect sensory processes. In SDT, the sensitivity ( $d'$ ) and criterion ( $c$ ), are typically associated with sensory capacity and decisional strategy, respectively. Our findings of congruency effects on  $d'$  but not  $c$  reflect congruency-dependent changes in sensory capacity. However, we noted that any increased discriminability of low-level visual features could be due to changes in sensory processing, or they could simply be the result of attentional cueing to the relevant feature (Chiou & Rich, 2012; Mossbridge et al., 2011). We suggest that our duration results, in particular, significant congruency effects on sensitivity for 50-ms stimuli, argue against attentional cueing.

CMC congruency effects are not simply sensory in nature but are likely low-level sensory processing effects. That is, congruency effects appear to exert an influence during early processing of low-level stimulus features rather than later, higher level processing of integrated objects. Even in a dual task paradigm where participants had to discriminate both the lightness and the elevation of a visual target, only lightness/pitch congruency influenced lightness discrimination, whereas elevation/pitch congruency influenced elevation discrimination. That is, even when both visual features were task relevant, each congruency only enhanced the feature to which it was related. These findings appear to contrast with the literature. Whereas we found congruency effects only when the task involved a CMC feature, both Evans and Treisman (2010) and Mossbridge et al. (2011) found congruency effects for tasks that were unrelated to the CMCs (i.e. improved response times to discriminating grating direction for elevation/pitch congruent vs incongruent stimuli and improved sensitivity to colour discrimination for frequency change/elevation congruent vs incongruent stimuli, respectively). We suggest

that an alternative explanation for the findings of both Evans and Treisman (2010) and Mossbridge et al. (2011) is that the improved performance in grating discrimination or colour matching may be a ‘downstream’ effect. Given that each of these studies were response-time paradigms and the stimulus remained visible until a response was made, it is plausible that congruency benefits to a “non-CMC” feature like grating direction or colour may result from later attentional or response selection benefits that stem from the feature specific congruency benefits.

Our suggestion of a crossmodal influence on low-level processing is consistent with electrophysiological findings showing that cross-sensory interactions can occur very early in sensory processing. Giard and Peronnet (1999) found significant audiovisual event-related potential (ERP) effects in occipital electrodes 40–90 ms poststimulus and determined that this early pattern did not correspond to any ERP event in the unimodal response pattern. This was extended by Molholm et al. (2002) with materially increased EEG electrode density. They found an early (46 ms) audiovisual interaction over right parieto-occipital cortex and speculated that low-level auditory input modified the visual signal before the low-level visual sensory analysis was complete. Raji et al. (2010) considered the timing of unisensory versus low-level cross-sensory interactions and determined that in the case of audiovisual interactions, the primary sensory cortices were likely directly influencing primary sensory cortices crossmodally with a signal conduction delay of about 30 ms.

Early crossmodal interactions have been linked further with multisensory integration and associated with behavioural outcomes using the flash-beep illusion. In this illusion, a single short visual flash accompanied by two short beeps is often reported as two flashes (fission illusion), or, alternatively, a double flash accompanied by a single beep is reported as a single flash (flash fusion). In fMRI studies with the flash-beep illusion, Watkins, Shams, Tanaka, Haynes, and Rees (2006) showed that flash fission was associated with increased activity in early visual areas, and flash fusion was associated with decreased activity (Watkins, Shams, Josephs, & Rees, 2007), each result implying an effect of audition on low-level crossmodal processing that is correlated with multisensory perceptual outcomes.

Our novel finding is that early crossmodal interactions involve more than simply the presence of a crossmodal stimulus activating low-level sensory areas. These findings suggest that, in this case at least, the early cross-sensory interactions display a level of adaptability and different crossmodal features (say, a high vs a low pitch tone) can differentially activate low-level vision, biasing it according to specific CMC relationships. Further, multiple features can be biased independently in parallel. This interpretation fits with earlier CMC congruency effects including response-time benefits (e.g. Chiou & Rich, 2012; Evans & Treisman, 2010; Gallace

& Spence, 2006; Marks, 1987), degraded bimodal spatial (e.g. Bien et al., 2012) and temporal (e.g. Parise & Spence, 2008, 2009) offset judgements, increased sensitivity in SDT paradigms (Andersen & Mamassian, 2008), and others.

Three hypotheses have been proposed to account for congruent pairing of specific crossmodal features (Spence, 2011). First, congruency effects may reflect neural firing rates (e.g. Hidaka et al., 2013; Parise & Spence, 2013). Consider an intensity-mediated CMC like brightness/loudness. Since intensity (both visual brightness and auditory loudness) is encoded in the neural firing rate, it might simply be that crossmodal signals that correspond in firing rate are preferentially associated. Second, it has been suggested that congruency effects may reflect the natural statistics of the environment (e.g. Parise, 2016; Parise, Knorre, & Ernst, 2014; Parise & Spence, 2013). Smaller objects have higher resonant frequencies than larger objects and so, consistent with the observed congruency effects of the size/pitch CMC, smaller objects are likely to create sounds of higher pitch than are larger objects. Finally, congruency effects may be semantically mediated (e.g. Parise & Spence, 2013). For example, in the visual elevation/pitch CMC, congruent stimulus features have the same relative magnitude, and we use the same words to distinguish relative magnitude when describing both elevation and pitch (i.e. ‘high’ and ‘low’).

Our findings support the ‘natural statistics of the environment’ account for the two CMCs investigated here. Although lightness and pitch might plausibly have correlated neural firing rates, the same cannot be said for elevation and pitch. Conversely, while there may be a semantic connection between elevation and pitch (although we intentionally used up/down for elevation to distinguish it from the natural high/low of pitch), the same cannot be said for lightness and pitch.

Finally, we speculate that these congruency effects result from a top-down context-dependent biasing of the sensory processes. In support of this, we note prior CMC studies that found congruency effects depend both on the intermixed presentation of two stimulus values in each modality (e.g. Gallace & Spence, 2006) and the relative nature of the irrelevant stimulus (Chiou & Rich, 2012). Each of these requirements suggest that the irrelevant stimulus needs to be interpreted within the context of the recent environment to have a specific influence. An irrelevant tone acts crossmodally as a ‘high’ pitch only in comparison with some other tone that has been recently experienced that is lower in pitch. Comparison to recently heard sounds, interpreting the sound and linking it to a specific visual feature, and doing this with adaptability, all suggest high-level cognitive involvement. Furthermore, we note that in the case of lightness, associating black with low pitch and white with high pitch is not obvious and so requires interpretation, and similarly our use of up and down for elevation rather than high and low (which would match the levels of pitch) and left and right arrows also ensured that associations were not obvious.

In conclusion, we suggest that lightness/pitch and elevation/pitch congruency effects involving low-level stimulus features rely upon direct connections between low-level sensory cortices and result from a biasing of low-level sensory processes to enhance processing of particular feature values in parallel. For example, low-level auditory cortex distinguishes between high and low pitch tones, and each biases low-level visual processing towards either white (high pitch) or black (low pitch) visual features, and in parallel to either the high or low elevation in the visual field. Further, this low-level mapping may be under top-down control to operationalise the categorical and relative effects in line with the current perceptual experience. We also suggest that these congruency effects are likely learned through perceptual experience and reflect the natural statistics of the environment.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. doi: <https://doi.org/10.1016/j.cub.2004.01.029>
- Andersen, S. K., & Müller, M. M. (2010). Behavioural performance follows the time course of neural facilitation and suppression during cued shifts of feature-selective attention. *Proceedings of the National Academy of Sciences*, *107*(31), 13878–13882.
- Andersen, T. S., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision Research*, *48*(25), 2537–2544.
- Bernstein, I. H., & Edelman, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, *87*(2), 241–247.
- Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synaesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage*, *59*(1), 663–672.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525.
- Chiou, R., & Rich, A. N. (2012). Crossmodality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, *41*(3), 339–353.
- Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and crossmodal conflict in the human brain. *Journal of Neurophysiology*, *106*(6), 2896–2909.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169.
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6–6. doi: <https://doi.org/10.1167/10.1.6>
- Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*(7), 1191–1203.
- Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage*, *124*, 876–886.
- Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A

- behavioural and electrophysiological study. *Journal of Cognitive Neuroscience*, 11(5), 473–490.
- Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley.
- Hidaka, S., Teramoto, W., Keetels, M., & Vroomen, J. (2013). Effect of pitch–space correspondence on sound-induced visual motion perception. *Experimental Brain Research*, 231(1), 117–126.
- Kandel, E. (2013). Principles of neural science (5th). New York: McGraw-Hill Medical.
- Klapetek, A., Ngo, M. K., & Spence, C. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics*, 74(6), 1154–1167.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, 36(14), 1.
- Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLOS ONE*, 2(9), e943.
- Macmillan, N. A., & Creelman, C. D. (2005). Detection theory (2nd). London: Erlbaum.
- Marks, L. E. (1987). On crossmodal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3), 384.
- Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Crossmodal interactions in auditory and visual discrimination. *International Journal of Psychophysiology*, 50(1), 125–145.
- McDonald, J. J., Green, J. J., Störmer, V. S., & Hillyard, S. A. (2012). Cross-modal spatial cueing of attention influences visual perception. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes*. Boca Raton: CRC Press/Taylor & Francis.
- McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, 407(6806), 906–908.
- Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116(4), 323–336.
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons: I. Temporal factors. *Journal of Neuroscience*, 7(10), 3215–3229.
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221(4608), 389–391.
- Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, 365(2), 350–354.
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research*, 14(1), 115–128.
- Mossbridge, J. A., Grabowecy, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual–spatial attention. *Cognition*, 121(1), 133–139.
- Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. *Perception & Psychophysics*, 65(1), 123–132.
- Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters*, 442(3), 257–261.
- Parise, C., & Spence, C. (2013). Audiovisual crossmodal correspondences in the general population. In J. Simner & E. M. Hubbard (Eds.), *Oxford handbook of synesthesia*. Oxford: Oxford University Press.
- Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research*, 29(1/3), 7–28.
- Parise, C. V., Harrar, V., Ernst, M. O., & Spence, C. (2013). Cross-correlation between auditory and visual signals promotes multisensory integration. *Multisensory Research*, 26(3), 307–316.
- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, 111(16), 6104–6108.
- Parise, C. V., & 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: a study using the implicit association test. *Experimental Brain Research*, 220(3/4), 319–333.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3–25.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes*, 32, 531–556.
- Raij, T., Ahveninen, J., Lin, F. H., Witzel, T., Jääskeläinen, I. P., Letham, B., & Hämäläinen, M. (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *European Journal of Neuroscience*, 31(10), 1772–1782.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4), 971–995.
- Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*, 22(1), 245–260.
- Störmer, V. S., McDonald, J. J., & Hillyard, S. A. (2009). Crossmodal cueing of attention alters appearance and early cortical processing of visual stimuli. *Proceedings of the National Academy of Sciences*, 106(52), 22456–22461.
- Thelen, A., & Murray, M. M. (2013). The efficacy of single-trial multisensory memories. *Multisensory Research*, 26(5), 483–502.
- van Ede, F., de Lange, F. P., & Maris, E. (2012). Attentional cues affect accuracy and reaction time via different cognitive and neural processes. *Journal of Neuroscience*, 32(30), 10408–10412.
- Van Wanrooij, M. M., Bremen, P., & John Van Opstal, A. (2010). Acquired prior knowledge modulates audiovisual integration. *European Journal of Neuroscience*, 31(10), 1763–1771.
- 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(1), 138.
- Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *NeuroImage*, 37(2), 572–578.
- Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, 31(3), 1247–1256.
- Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception*, 44(3), 289–300.
- Zeljko, M., & Grove, P. M. (2017). Sensitivity and bias in the resolution of stream-bounce stimuli. *Perception*, 46(2), 178–204.