

# Summation of visual attributes in auditory–visual crossmodal correspondences

Clare Jonas<sup>1</sup> · Mary Jane Spiller<sup>1</sup> · Paul Hibbard<sup>2</sup>

Published online: 10 July 2017  
© Psychonomic Society, Inc. 2017

**Abstract** Crossmodal correspondences are a feature of human perception in which two or more sensory dimensions are linked together; for example, high-pitched noises may be more readily linked with small than with large objects. However, no study has yet systematically examined the interaction between different visual–auditory crossmodal correspondences. We investigated how the visual dimensions of luminance, saturation, size, and vertical position can influence decisions when matching particular visual stimuli with high-pitched or low-pitched auditory stimuli. For multidimensional stimuli, we found a general pattern of summation of the individual crossmodal correspondences, with some exceptions that may be explained by Garner interference. These findings have applications for the design of sensory substitution systems, which convert information from one sensory modality to another.

**Keywords** Multisensory perception · Crossmodal correspondences · Vision · Hearing

We live in a multisensory world, filled with sights, sounds, smells, textures, and tastes. We need to correctly integrate the information from different senses to create a unified understanding of the world—the *binding problem*. This article deals

with “property binding” (Treisman, 1996): linking together the different sensory properties of individual objects.

Shams and Kim (2010) suggested that, faced with multi-sensory input, brains attempt to minimize perceptual errors across all domains, using at least some top-down processes. Some combinations of information are therefore more likely to be bound together than others. This can happen through crossmodal correspondences (CMCs): pairs of cross-sensory stimuli that “go together,” apparently automatically (e.g., Evans & Treisman, 2010; but see Spence & Deroy, 2013). One example is the *kiki–bouba* effect: Participants typically pair spiky shapes with names containing high-pitched vowels (e.g., *kiki*), and round shapes with names containing low-pitched vowels (e.g., *bouba*; e.g., Bremner et al., 2013). CMCs occur in many sensory pairings: high luminance pairs with tactile softness (Ludwig & Simner, 2013), and blackberry odor pairs with “piano” (Crisinel & Spence, 2011). CMCs may occur for a variety of reasons, including (adult remnants of) neonatal inability to differentiate sensory inputs, statistical coupling of sensory dimensions in the environment, and semantic “matching” of stimuli (e.g., Mondloch & Maurer, 2004; Spence, 2011; Walker, Walker, & Francis, 2012).

Early studies on CMCs generally explored complex stimuli (e.g., Karwowski, Odbert, & Osgood, 1942, had participants draw visual responses to music); more-recent studies have focused on single CMCs. However, we lack information about how CMCs interact. This topic has been systematically approached only by Eitan and Rothschild (2011), who studied imagined tactile qualities of musical notes, and by Woods, Spence, Butcher, and Deroy (2013), in an online study of interactions between sounds, shapes, and emotions.

Interactions between CMCs are important, since real-world objects do not have only two sensory dimensions. For example, drums have visual, tactile, and auditory properties. A drum may have a dark color but a light weight (i.e., opposing

---

✉ Clare Jonas  
c.n.jonas@uel.ac.uk

<sup>1</sup> School of Psychology, University of East London, London, UK

<sup>2</sup> Department of Psychology, University of Essex, Colchester, UK

ends of the dark–light and heavy–light dimensions; Ward, Banissy, & Jonas, 2008). Do we predict that the drum makes a high sound because of its weight (Walker et al., 2012), or a low sound because of its color (Hubbard, 1996)?

In this study, we investigated the existence of interactions between auditory–visual CMCs (Spence & Deroy, 2013). We displayed visual stimulus pairs varying in luminance (lightness), saturation (color intensity), size, and/or vertical position, with auditory stimulus pairs varying in pitch. Participants decided which auditory stimuli “went with” which visual stimuli. Our goal was to determine the principles used to combine multiple CMCs.

We tested three models of CMC interaction. First we tested the *summation model*, based on sensory cue integration models (Trommershäuser, Körding, & Landy, 2011), in which the strengths of the individual CMCs add. When CMCs are consistent, crossmodal associations are strengthened. When CMCs conflict, they cancel out completely or partially, depending on their relative strengths. Second, we examined the *hierarchy model*, in which there was a hierarchy of CMCs, with some dominating others. The third was the *majority model*, where most (but not all) characteristics were paired with a specific pitch (e.g., a small, low luminance, low position stimulus would pair with high pitch, in terms of size, but with low pitch, in terms of luminance/position). In this model, participants’ pitch choices were predicted by the majority of the feature correspondences (in this case, low pitch).

## Method

### Participants

Because this was a novel line of research, and relied on proportions of responses across participants as the dependent measure, we wanted to sample as many participants as possible in the time available. We collected data online (<https://uelpsihology.org/soundvision>), recruiting 113 participants (76 female, 30 male, two other, and five who declined to respond; 18–67 years of age, mean = 30.82, *SD* = 11.39) from personal contacts and online communities of volunteers. Seventy-nine of these informants were monolingual English speakers, ten were bilingual native speakers of English, and the remaining 24 were nonnative speakers of English.

All participants gave informed consent. The experiment was approved by the Research Ethics Committee of the University of East London.

### Materials, design and procedure

The visual stimuli were two circles on a midgray background (Table 1). The circles varied in luminance, saturation, size, and

position. We chose four hues: red (hue in HSL system: 0), yellow (58), green (120), and blue (240). Within participants, hue was held constant, and the other characteristics varied. Each characteristic had three levels: low/large, medium, and high/small. For luminance and saturation, “low” was a value of 16%, “medium” 50%, and “high” 85% in the HSL system. All circles were presented with their centers aligned. We report positions and sizes as they appeared on a 56-cm widescreen monitor where the image occupied a rectangle with width 106 mm and height 79 mm (monitor sizes will have varied, since this was an online experiment). “Low” circles had centers 56 mm from the top of the image background, “medium” 40 mm, and “high” 23 mm. “Large” circles had a diameter of 25 mm, “medium” of 16 mm, and “small” of 8 mm.

Each pair of circles was either the same (i.e., both medium) or opposite (e.g., one large, one small) on all four within-participants characteristics. This gave us four “levels” of stimuli. At Level 1, the circles varied in one characteristic (e.g., one high and the other low luminance, but for all other characteristics both were medium). At Level 2, the circles varied on two characteristics; at Level 3, on three characteristics; and at Level 4, on all four. We describe pairs according to the characteristics of the left circle; the right circle’s characteristics are implied in that description. Participants saw every possible combination of circles twice; the second time, the circles’ left–right positions were reversed (a total of 80 stimuli).<sup>1</sup>

We used the responses to Level 1 stimuli to predict the responses at Levels 2–4. Therefore, it was unimportant that the perceptual distances between values were not identical across the stimulus dimensions; participants needed only distinguish between the values on each dimension. The experiment was programmed using Javascript.

The auditory stimuli were created using Audacity (<http://audacityteam.org/>). These were two pure-tone sine waves, each of 1,000-ms duration. One was at a pitch of 261.63 Hz, the other at 523.25 Hz. In each trial, participants heard both beeps; their order was counterbalanced across trials. The order of beeps was counterbalanced across participants, who were randomly assigned across the eight conditions (4 hues × 2 auditory orders). Twenty-eight participants were assigned to the red condition, 28 to green, 30 to yellow, and 27 to blue.

At the start of each trial, the visual stimuli appeared on the screen (see Fig. 1). Participants clicked to play the first auditory stimulus, with the second following automatically after 2,000 ms silence. Participants could replay the stimuli as

<sup>1</sup> Due to a programming error, participants did not see the Level 1 stimulus with medium luminance/size/position and low saturation, but instead were shown another Level 1 stimulus (medium luminance/saturation/position, small size) twice.

**Table 1** Example visual stimuli for each of the four levels

Characteristic	Level 1	Level 2	Level 3	Level 4
Luminance	+/-	+/-	+/-	+/-
Saturation	=	-/+	-/+	-/+
Size	=	=	-/+	-/+
Position	=	=	=	+/-
N trials	8	24	32	16

At Level 1, the two circles vary in only one characteristic; at Level 2, in two characteristics; at Level 3, in three characteristics; and at Level 4, in four characteristics. +/- indicates that the characteristic is high/small on the left and low/large on the right; -/+ indicates that the characteristic is low/large on the left and high/small on the right; = indicates that the characteristic is the same in both circles.

needed before deciding which beep went with which visual stimulus.

Prior to the 80 experimental trials, the participants completed four practice trials with stimuli not used in the main study.

**Initial analysis and statistics**

For the reported analyses, we used data from all participants. Similar results were found for monolingual native English speakers alone.

Initial analysis of the Level 1 stimuli established the association strengths of each individual CMC. Figure 2a shows the proportions of participants who chose high beeps for each stimulus at Level 1. In all cases, we found reliable and significant correspondences between sensory

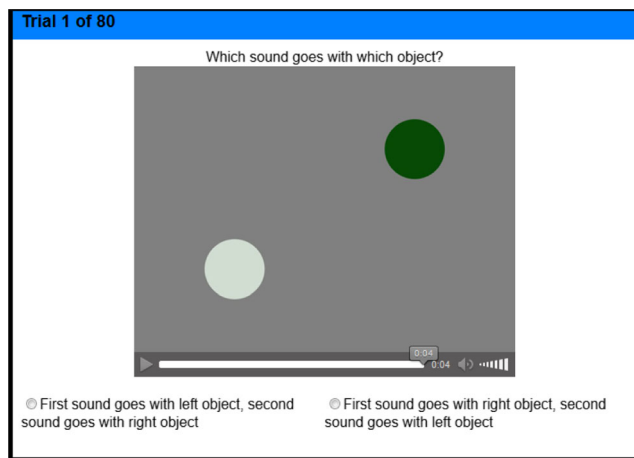
dimensions. The high beep was associated with stimuli with higher luminance, saturation, or position, or stimuli that were smaller. Because the auditory stimuli were matched for physical amplitude, the high beep was probably perceived as louder (International Organisation for Standardization (ISO), 2003). It is thus likely that the strength and direction of association was determined by pitch and loudness. This does not affect our interpretation of the results, which concern how different visual dimensions combined in determining the CMCs.

Association strengths were modeled by assuming that each value on each visual dimension has a particular strength of association with the high beep, relative to the low beep. We also assumed some variation in association strength across the population, modeled using a normal distribution. Using probits, we transformed the proportions of participants choosing each beep for each visual stimulus, to quantify the association strengths in units of the standard deviation of the variability (Thurstone, 1927):

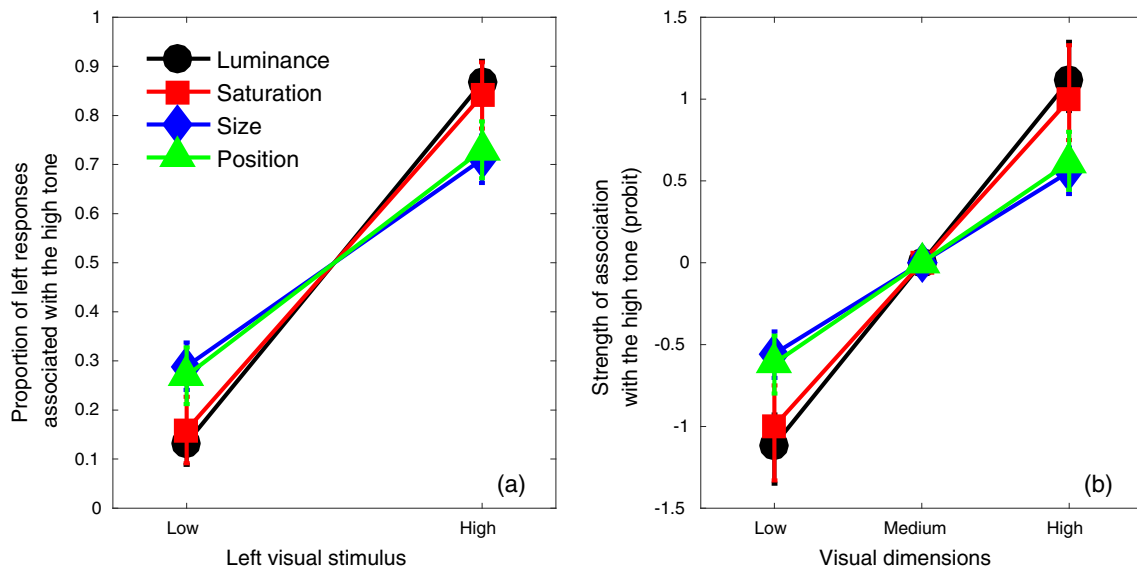
$$p(R = LEFT|S) = \phi(S), \tag{1}$$

where  $R = LEFT$  represents a participant choosing the left stimulus,  $S$  is the association strength, and  $\phi$  is the cumulative distribution function of the standard normal distribution. Association strength is quantified in terms of the variability in responses across observers.

Probit values for the Level 1 stimuli are plotted in Fig. 2b. These values were fixed at 0 for neutral stimuli: When both circles have the same value on a dimension, there can be no preference associated with that dimension. These associations were used to predict the outcomes for stimuli containing variations in multiple dimensions. We predicted these results using each model as follows:



**Fig. 1** Example trial (high luminance, medium saturation, medium size, low position), as viewed by the participant after the video has been played. The participant could not see the radio-button decisions beneath the video until it had been played once



**Fig. 2** Results for the Level 1 conditions, in which the stimuli varied on only a single dimension. (a) Proportions of participants who chose a high-pitched beep as the one that went with each stimulus. Error bars show binomial 95% confidence intervals. (b) Strengths of the associations

**Summation**

The simplest assumption is that association strengths will add:

$$S_{TOTAL} = S_{LUM} + S_{SAT} + S_{SIZE} + S_{POS}. \tag{2}$$

This model assumes that all dimensions are equally important in determining association strengths.

**Hierarchy**

In this model, there is a hierarchy of CMCs. For any stimulus, the CMC is predicted by the dominant association, which is not necessarily the dimension with the strongest association when presented alone. Rather, this model assumes a specific order in which the dimensions are considered, with the association determined by the first dimension, within this order, on which stimuli differ. Since we tested four CMCs, there were 24 (4 × 3 × 2 × 1) possible hierarchies. We calculated correlations between the predicted and actual responses for each stimulus, for all hierarchies, and chose the hierarchy that best fit the data. This method provided considerable freedom to achieve the best fit; the other models contained no free parameters.

**Majority**

In this model, where there was conflict between the directions of CMCs, the response was determined by majority vote, regardless of the strengths of the individual CMCs. If all stimulus dimensions, and experimental manipulations, had the same strength, then the predictions of the summation and

between the frequency of the auditory stimulus and each dimension of the visual stimuli, calculated using probit analysis (see the text for details). “Low” and “high” map to “large” and “small,” respectively, for the size dimension

majority models would agree. However, if, for example, one dimension was particularly dominant, this might outweigh the combined effects of other dimensions that predicted the opposite response.

**Results**

For each model, we calculated correlations between the predicted and actual responses for Level 2, 3, and 4 stimuli (Table 2). The summation model predicted the data well, with all correlations being significant. The correlations for the majority model were also significant, but lower than those for the summation model. The correlations for the hierarchy model, which did not take account of all CMCs, were in all cases

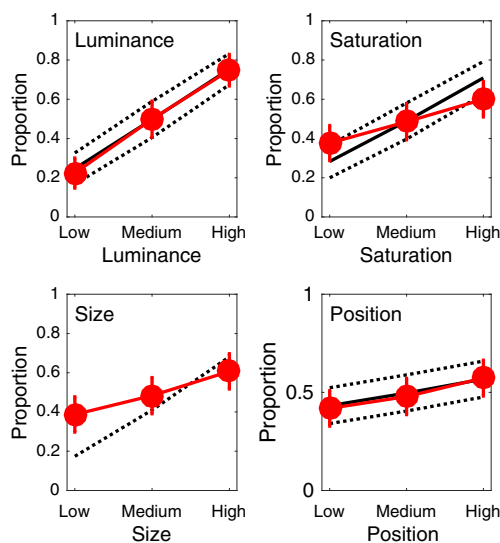
**Table 2** Correlation coefficients and significance levels for the fits of the probit summation, hierarchy, and majority models

Model	Stimulus Level	Correlation Coefficient	Significance Level
Summation	Level 2	.90	$p < .000001$
Summation	Level 3	.83	$p < .000001$
Summation	Level 4	.77	$p = .000575$
Hierarchy	Level 2	.83	$p = .000001$
Hierarchy	Level 3	.43	$p = .0146$
Hierarchy	Level 4	.06	$p = .8207$
Majority	Level 2	.85	$p < .000001$
Majority	Level 3	.70	$p = .000007$
Majority	Level 4	.74	$p = .001006$

lower, and nonsignificant for Level 4 stimuli. Therefore, for stimuli containing multiple CMCs, all visual dimensions contribute to participants' decisions.

To further test the summation model, we created a generalized linear model with a binomial distribution and a probit linking function. A full factorial model was used, with color saturation and luminance, the width of the stimulus, and its distance from the center of the screen as covariates. Each of these covariates was significant (luminance, Wald  $\chi^2 = 1,734.8, p < .001$ ; saturation, Wald  $\chi^2 = 424.1, p < .001$ ; size, Wald  $\chi^2 = 348.0, p < .001$ ; position, Wald  $\chi^2 = 203.1, p < .001$ ). None of the two-way interactions were significant, but there were significant three-way interactions between luminance, size, and position (Wald  $\chi^2 = 4.10, p = .043$ ); luminance, saturation, and size (Wald  $\chi^2 = 9.00, p = .003$ ); and saturation, size, and position (Wald  $\chi^2 = 6.69, p = .01$ ).

We also predicted the main effect and two-way interaction results using probits for the Level 1 stimuli (Fig. 2b), using a linear regression after centering the data for each dimension. The results were combined according to Eq. 2, and the predicted proportion of “left” responses was calculated from the resulting probit value. These results are plotted in Figs. 3 (main effects) and 4 (two-way interactions). These report good predictions for luminance and size. However, the effect of saturation, in particular, was less than expected. A simple linear model therefore does not appear to fully account for associations made when stimuli vary across multiple visual dimensions. This apparent difference was tested using a generalized linear model with saturation as a covariate, fit separately to the data from different levels of luminance. The effect of



**Fig. 3** Proportions of “left” responses associated with the higher tone, as a function of each visual dimension, for stimuli pooled over all other visual dimensions. The color symbols indicate the participants' responses, and the solid black lines the predictions of the probit model. Error bars and dotted black lines represent 95% binomial confidence limits of the data and the model predictions, respectively

saturation was significantly greater for neutral-luminance stimuli ( $b = .026$  [95% confidence limits: .024–.028]; Wald  $\chi^2 = 739.2; p < .001$ ) than for those with low ( $b = .001$  [–.0001 to .0003]; Wald  $\chi^2 = 14.934; p = .24$ ) or high ( $b = .003$  [.002–.005]; Wald  $\chi^2 = 15.8; p < .001$ ) luminance. Participants' responses were only strongly influenced by saturation when luminance was neutral.

To interpret the significant three-way interactions, we performed separate analyses for each stimulus size, with luminance and position, luminance and saturation, or saturation and position as predictors (Table 3). We found significant main effects of luminance, saturation, size, and position in all conditions. For medium-sized objects, there was a significant interaction between luminance and saturation, consistent with the reduced effect of saturation at low and high levels of luminance.

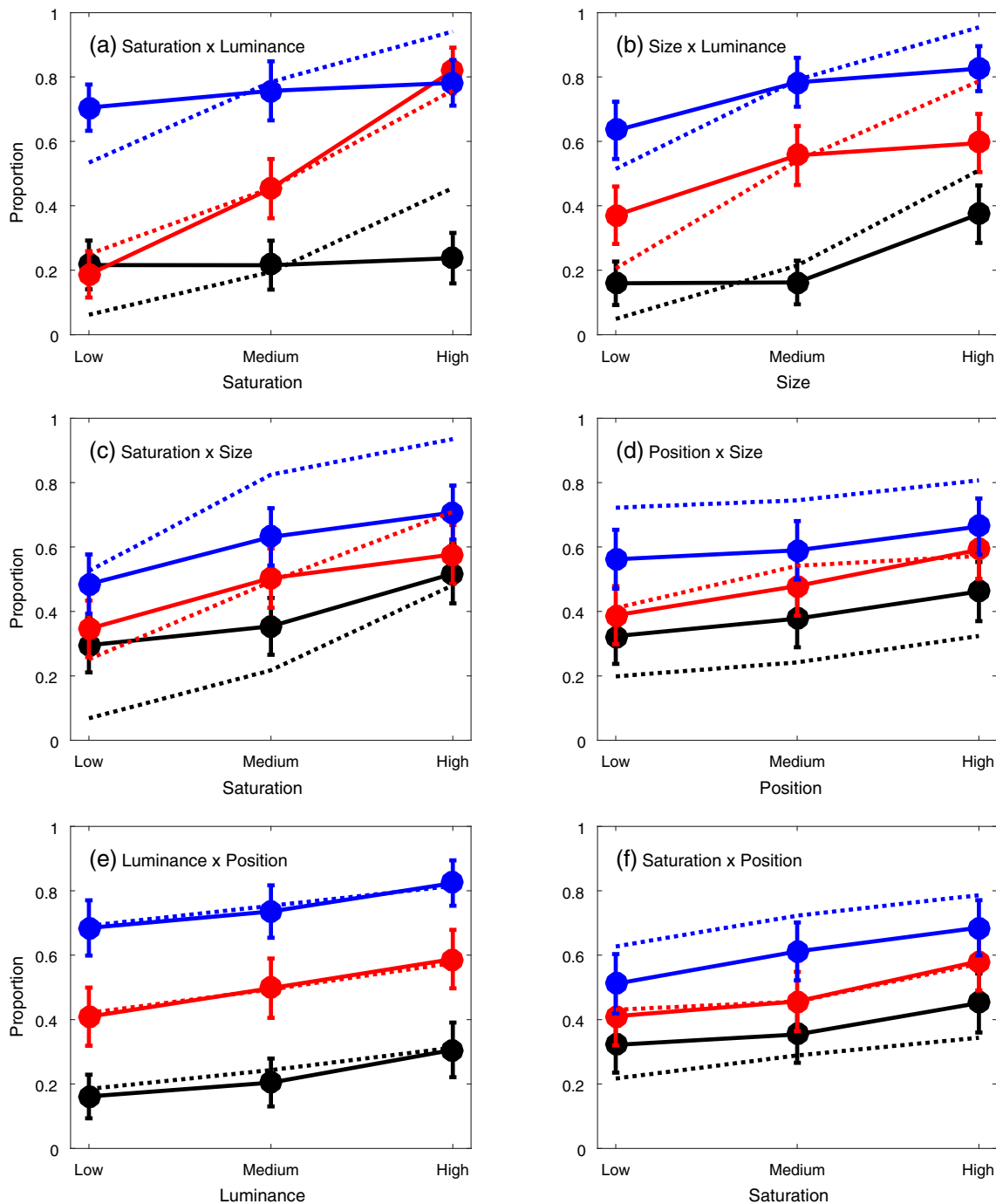
All calculations were performed using the HSL system. It is possible that different results could have been obtained if the stimuli were analyzed in a different color space. For example, the CIE Luminance  $\times$  Chroma Hue (LCh) space might be considered more appropriate, since distances in this space relate to just-noticeable differences in color. We recalculated our probit predictions in the LCh color space, but found little difference in the overall fits of the model, regardless of whether the HSL ( $R^2 = .534$  across all stimuli) or the LCh ( $R^2 = .525$ ) space was used.

## Discussion

We examined how visual characteristics interact to determine which auditory pitch “goes with” a given visual stimulus. We found the predicted associations of high pitch with high luminance, high saturation, small size, and high position when one visual characteristic was varied (following, e.g., Evans & Treisman, 2010; Hamilton-Fletcher, 2015; Klapetek et al., 2012). Our study extends previous research by using visual stimuli that differed on two or more characteristics. A linear summation model predicted participants' choices more accurately than a majority or a hierarchy model, although some results did not fit this model.

The summation model's overall success in predicting participants' responses suggests a general strategy of weighting the available visual cues to determine the best auditory match, perhaps via neural intensity matching (Spence, 2011) or a generalized system for dealing with magnitude (Walsh, 2003). However, we need to account for the few results that violate the model (the lower effects of saturation at low and high luminances, and the three-way interactions of luminance, position, size; luminance, saturation, and size; and saturation, position, and size). The decreased effects of saturation at low and high





**Fig. 4** Proportions of “left” responses associated with the higher tone, as a function of each pair of visual dimensions, for stimuli pooled over all other visual dimensions. In all cases, one dimension is plotted on the horizontal axis, and the black, red, and blue symbols (color only in the

online figure) represent the “low,” “medium,” and high values on the other dimension, respectively. The dashed lines of each color show the predictions of the probit model. Error bars and dotted lines indicate 95% binomial confidence limits of the data and the model fits, respectively

luminances appear to be the result of Garner interference. In Garner’s (1976) paradigm, participants are presented with stimuli that vary along two perceptual dimensions, and then make decisions about one dimension. Information from the irrelevant dimension can interfere with decision making about the relevant information. When this happens, the dimensions are *integral* and

viewed as one super-dimension. In our results, luminance and saturation integrate to form one super-dimension (see, e.g., Burns & Shepp, 1988), except when luminance is medium and does not differ between the two stimuli. However, other violations of the model are not clear-cut instances of Garner interference. One possibility is that the dimensions are incompletely integrated, so that

**Table 3** Results of the generalized linear models, performed separately for small, medium, and large stimuli, with luminance and position or luminance and saturation as predictors

Predictor	Wald Chi-Square	Degrees of Freedom	Significance Level
Luminance × Position (small stimuli)			
Luminance	477.1	1	$p < .001$
Position	49.5	1	$p < .001$
Luminance × Position	0.036	1	$p = .85$
Luminance × Position (medium stimuli)			
Luminance	788.1	1	$p < .001$
Position	84.38	1	$p < .001$
Luminance × Position	0.442	1	$p < .506$
Luminance × Position (large stimuli)			
Luminance	416.0	1	$p < .001$
Position	41.54	1	$p < .001$
Luminance × Position	1.99	1	$p = .159$
Luminance × Saturation (small stimuli)			
Luminance	122.3	1	$p < .001$
Saturation	484.6	1	$p < .001$
Position × Saturation	3.089	1	$p = .079$
Luminance × Saturation (medium stimuli)			
Luminance	840.9	1	$p < .001$
Saturation	176.9	1	$p < .001$
Position × Saturation	14.01	1	$p < .001$
Luminance × Saturation (large stimuli)			
Luminance	412.8	1	$p < .001$
Saturation	116.8	1	$p < .001$
Position × Saturation	1.48	1	$p = .224$
Saturation × Position (small stimuli)			
Saturation	107.5	1	$p < .001$
Position	43.27	1	$p < .001$
Saturation × Position	0.712	1	$p = .397$
Saturation × Position (medium stimuli)			
Saturation	119.6	1	$p < .001$
Position	101.1	1	$p < .001$
Saturation × Position	3.26	1	$p = .071$
Saturation × Position (large stimuli)			
Saturation	106.6	1	$p < .001$
Position	23.3	1	$p < .001$
Saturation × Position	1.25	1	$p = .263$

participants' decisions are influenced by the dimensions at unequal relative strengths, but also by interactions between the different dimensions.

### Explaining summation in the context of theories about CMCs

How CMCs arise is a matter of ongoing investigation (e.g., Lindborg & Friborg, 2015). Eventually it should be possible to make a broad taxonomy of the fundamental mechanisms of CMCs. Some probably occur earlier in processing than others (e.g., a CMC based on statistical features of the environment probably occurs earlier than a language-based one), so early-occurring CMCs are likely to impact on later ones.

It is also possible that some CMCs begin at an early stage of processing and spread to other CMCs (e.g., a structural CMC that becomes encoded in language). These hypothetical

CMCs would likely have more effect on perceptions and decisions than those that occur at only one level. That is, if a multilevel CMC conflicts with a single-level one, the multilevel CMC is likely to “win.”

### Limitations and future directions

Online testing has advantages, including ease and speed of participant recruitment, but also disadvantages (Woods, Velasco, Levitan, Wan, & Spence, 2015). Repeated participation is one concern. However, since this study was unpaid and was informally reported by some participants to be tedious, the participants would not have repeatedly participated for money or fun.

The variety of participant hardware and system settings used will have affected the exact presentation of the stimuli. However, because we asked participants to judge the comparative visual features of stimuli presented at the same time, this cross-participant variance should not matter. This does, however, mean that our experiment cannot speak to whether CMCs and the interactions between them are relative or absolute. Consequently, an important next step will be to replicate this experiment in laboratory conditions. We do not expect very different results: When millisecond accuracy in presentation or response collection is not required, participants largely behave similarly in the lab and online (Woods et al., 2015).

To keep the experiment short, we tested only one auditory dimension. Therefore, we cannot know whether our findings are specific to the relationships of visual dimensions with pitch, or whether the same interactions would occur if we tested, say, duration instead. This question may also be applied to other sensory pairings—for example, differing tactile stimuli being matched with visual stimuli.

A consideration for future research is whether the relationships between CMCs could have appeared if we had presented visual characteristics varying in a single dimension alongside auditory or tactile stimuli varying in multiple dimensions. Is summation a general feature of CMCs, or is it unique to vision? Evidence showing that timbre, pitch, and loudness interact to varying extents in speeded classification paradigms (Melara & Marks, 1990) suggests that similar results would be seen at least with aurally multidimensional CMCs.

Finally, it is not clear whether the CMC interactions we have reported are implicit or explicit. This could be tested using speeded classification tasks (see Marks, 2004). For example, temporal-order judgments (e.g., Parise & Spence, 2009) would allow an exploration of whether interactions occur at perceptual or decisional levels. An analysis of response times would also allow an exploration of the impacts of multiple conflicting or converging cues on decision making.

## Applications

CMCs are used in packaging design (e.g., Becker, van Rompay, Schifferstein, & Galetzka, 2011), though not always successfully (Crisinel & Spence, 2012). Since real-world objects have multiple sensory dimensions, the existence of nonsummative effects of different dimensions indicates that it is important to consider which features of packaging or advertising are most strongly associated with the dimension that needs to be emphasized.

Our findings will also be helpful for designers of sensory substitution devices, such as the vOICe (Meijer, 1992), which allow the “translation” of information from one sense to another (for a review, see Hamilton-Fletcher & Ward, 2013). Having explicit knowledge about the relationships between different CMCs will allow for better design of default settings that are intuitively correct to most, reducing the time needed to learn to use such devices (Auvray, Hanneton, & O’Regan, 2007). The findings could also help when comparing devices that pair a single quality (e.g., pitch) with others, such as saturation (Bologna, Deville, Pun, & Vinckenbosch, 2007) and luminance (Doel, 2003).

**Author note** We thank Tony Leadbetter for his technical assistance, and Danielle van Versendaal and Sarah Hamburg for commenting on an earlier draft of the manuscript.

## References

- Auvray, M., Hanneton, S., & O’Regan, J. K. (2007). Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with “The vOICe.”. *Perception*, *36*, 416–443. doi:10.1068/p5631
- Becker, L., van Rompay, T. J., Schifferstein, H. N., & Galetzka, M. (2011). Tough package, strong taste: The influence of packaging design on taste impressions and product evaluations. *Food Quality and Preference*, *22*, 17–23. doi:10.1016/j.foodqual.2010.06.007
- Bologna, G., Deville, B., Pun, T., & Vinckenbosch, A. (2007). Transforming 3D coloured pixels into musical instrument notes for vision substitution applications. *EURASIP Journal on Image and Video Processing*, *2007*, 076204. doi:10.1155/2007/76204
- Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). “Bouba” and “Kiki” in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition*, *126*, 165–172. doi:10.1016/j.cognition.2012.09.007
- Burns, B., & Shepp, B. E. (1988). Dimensional interactions and the structure of psychological space: The representation of hue, saturation, and brightness. *Perception & Psychophysics*, *43*, 494–507. doi:10.3758/BF03207885
- Crisinel, A.-S., & Spence, C. (2011). A fruity note: Crossmodal associations between odors and musical notes. *Chemical Senses*, *37*, 151–158. doi:10.1093/chemse/bjr085
- Crisinel, A.-S., & Spence, C. (2012). Assessing the appropriateness of “synaesthetic” messaging on crisps packaging. *Food Quality and Preference*, *26*, 45–51. doi:10.1016/j.foodqual.2012.03.009
- Doel, K. (2003, July). *SoundView: Sensing color images by kinesthetic audio*. Paper presented at the 2003 International Conference on Auditory Display, Boston, MA.
- Eitan, Z., & Rothschild, I. (2011). How music touches: Musical parameters and listeners’ audiotactile metaphorical mappings. *Psychology of Music*, *39*, 449–467. doi:10.1177/0305735610377592
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6. doi:10.1167/10.1.6
- Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *8*, 98–123. doi:10.1016/0010-0285(76)90006-2
- Hamilton-Fletcher, G. (2015). How touch and hearing influence visual processing in sensory substitution, synaesthesia and cross-modal correspondences (Unpublished doctoral dissertation). University of Sussex, Brighton, UK.
- Hamilton-Fletcher, G., & Ward, J. (2013). Representing colour through hearing and touch in sensory substitution devices. *Multisensory Research*, *26*, 503–532. doi:10.1163/22134808-00002434
- Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology*, *109*, 219–238.
- International Organization for Standardization. (2003). Acoustics—Normal equal-loudness-level contours (ISO Standard 226: 200). Retrieved from <https://www.iso.org/standard/34222.html>
- Karwoski, T. F., Odbert, H. S., & Osgood, C. E. (1942). Studies in synesthetic thinking: II. The role of form in visual responses to music. *Journal of General Psychology*, *26*, 199–222. doi:10.1080/00221309.1942.10545166
- 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*, 1154–1167. doi:10.3758/s13414-012-0317-9
- Lindborg, P., & Friborg, A. K. (2015). Colour association with music is mediated by emotion: Evidence from an experiment using a CIE Lab interface and interviews. *PLoS ONE*, *10*, e0144013. doi:10.1371/journal.pone.0144013
- Ludwig, V. U., & Simner, J. (2013). What colour does that feel? Tactile–visual mapping and the development of cross-modality. *Cortex*, *49*, 1089–1099. doi:10.1016/j.cortex.2012.04.004
- 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.
- Meijer, P. B. L. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, *39*, 112–121. doi:10.1109/10.121642
- Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch and loudness. *Perception & Psychophysics*, *48*, 169–178. doi:10.3758/BF03207084
- Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch–object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 133–136. doi:10.3758/CABN.4.2.133
- Parise, C., & Spence, C. (2009). “When birds of a feather flock together”: Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE*, *4*, e5664. doi:10.1371/journal.pone.0005664
- Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, *7*, 269–284. doi:10.1016/j.plrev.2010.04.006
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*, 971–995. doi:10.3758/s13414-010-0073-7
- Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*, *22*, 245–226. doi:10.1016/j.concog.2012.12.006
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286. doi:10.1037/h0070288



- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6, 171–178. doi:[10.1016/S0959-4388\(96\)80070-5](https://doi.org/10.1016/S0959-4388(96)80070-5)
- Trommershäuser, J., Körding, K., & Landy, M. S. (2011). *Sensory cue integration*. Oxford, UK: Oxford University Press.
- Walker, L., Walker, P., & Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception*, 41, 1186–1192. doi:[10.1068/p7149](https://doi.org/10.1068/p7149)
- Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7, 483–488. doi:[10.1016/j.tics.2003.09.002](https://doi.org/10.1016/j.tics.2003.09.002)
- Ward, J., Banissy, M. J., & Jonas, C. N. (2008). Haptic perception and synaesthesia. In M. Grunwald (Ed.), *Human haptic perception: Basics and applications* (pp. 259–265). Basel, Switzerland: Birkhäuser.
- Woods, A. T., Spence, C., Butcher, N., & Deroy, O. (2013). Fast lemons and sour boulders: Testing crossmodal correspondences using an internet-based testing methodology. *i-Perception*, 4, 365–379. doi:[10.1068/i0586](https://doi.org/10.1068/i0586)
- Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the Internet: A tutorial review. *PeerJ*, 3, e1138. doi:[10.7717/peerj.1058](https://doi.org/10.7717/peerj.1058)