SUMMATION OF VISUAL ATTRIBUTES IN AUDITORY-VISUAL CROSSMODAL CORRESPONDENCES

Clare Jonas[1]*, Mary Jane Spiller[1], & Paul Hibbard[2]

[1]School of Psychology, University of East London, Water Lane, London E15 4LZ, U.K.

[2]Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, U.K.

Brief Report to be submitted to *Psychonomic Bulletin and Review*

Word count: 3998 (excluding abstract and keywords)

*Corresponding author:

Clare N. Jonas

School of Psychology

University of East London

Stratford Campus

Water Lane

London E15 4LZ

Tel: +44 (0) 20 8223 4659

Fax: +44 (0) 20 8223 4937

Email: c.n.jonas@uel.ac.uk

1. 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 objects than large objects. However, no study yet has 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 multi-dimensional stimuli, we found a general pattern of summation of 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

2. Introduction

We live in a multisensory world filled with sights, sounds, smells, textures, and tastes. We need to correctly integrate information from different senses to create a unified understanding of the world – the binding problem. This paper deals with 'property binding' (Treisman, 1996): linking together different sensory properties of individual objects.

Shams and Kim (2010) suggested that, faced with multisensory input, brains attempt to minimise 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), while blackberry odour 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. Karwoski, Odbert & Osgood, 1942, had participants draw visual responses to music); more recent studies have focused on single CMCs. We lack, though, 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 Woods, Spence, Butcher and Deroy (2013), in an online study of interactions between sounds, shapes and emotions.

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

We tested three models for CMC interaction. First, the *summation model*, based on sensory cue integration models (Trommershäuser, Kording, & Landy, 2011). Here, strengths of individual CMCs add. When CMCs are consistent, cross-modal associations are strengthened. When CMCS conflict, they cancel out completely or partially, depending on their relative strengths. Second, the *hierarchy model*, in which there is a hierarchy of CMCs, with some dominating others. Third, the *majority model*, where most (but not all) characteristics are paired with a specific pitch (e.g. a small, low luminance, low position stimulus pairs with high pitch in terms of size but with low pitch in terms of luminance/position). In this model, participants' pitch choices are predicted by the majority (in this case, low pitch).

3. Methods

3.1. Participants

As this is a novel line of research, and relies 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://uelpsychology.org/soundvision), recruiting 113 participants (76 female, 30 male, 2 other, 5 declined to respond; aged 18-67 years, mean = 30.82, *SD* = 11.39) from personal contacts and online communities of volunteers. Seventy-nine were monolingual English speakers; 10 were bilingual native speakers of English, the remaining 24 were non-native speakers of English.

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

3.2. Materials, design and procedure

Visual stimuli were two circles on a mid-grey background (Table 1). 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 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 sizes were presented with their centres aligned. We report positions and sizes as they appear on a 56cm widescreen monitor where the image occupied a rectangle with width 106mm and height 79mm (monitor sizes will have varied as this was an online experiment). ‘Low’ circles had centres 56mm from the top of the image

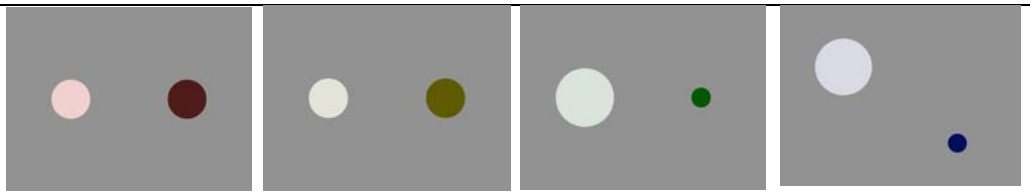
background, ‘medium’ 40mm, and ‘high’ 23mm. ‘Large’ circles had a diameter of 25mm, ‘medium’ 16mm, and ‘small’ 8mm.

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, circles varied on one characteristic (e.g. one high and the other low luminance, but for all other characteristics both medium). At Level 2, circles varied on two characteristics, at Level 3, on three characteristics, and at Level 4, on all four. We describe pairs using 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, their left-right positions were reversed (a total of 80 stimuli)[1].

We used responses to Level 1 stimuli to predict responses at Levels 2-4. Therefore, it is unimportant that perceptual distance between values is not identical across stimulus dimensions; participants need only distinguish between values on each dimension.

[1] Due to a programming error, participants did not see the Level 1 stimulus with medium luminance/size/position and low saturation and 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. At Level 1, the two circles vary only in one characteristic; at Level 2, in two characteristics; at Level 3, in three characteristics; and at Level 4, four characteristics. +/- indicates that the characteristic is high/small on the left and low/large on the right; -/+ indicates the characteristic is low/large on the left and high/small on the right; = indicates that the characteristic is the same in both circles.

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

The experiment was programmed using Javascript.

Auditory stimuli were created using Audacity (http://audacityteam.org/). These were two pure-tone sine waves, each of 1000ms duration. One was at a pitch of 261.63Hz, the other at 523.25Hz. In each trial, participants heard both beeps; their order was counterbalanced across trials. Order of beeps was counterbalanced across participants. Participants were randomly assigned across the eight conditions (four hues x two 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, visual stimuli appeared on the screen. Participants clicked to play the first auditory stimulus, with the second following automatically after 2000ms silence. Participants could replay stimuli as needed before deciding which beep went with which visual stimulus.

Prior to the 80 experimental trials, participants completed 4 practice trials with stimuli not used in the main study.

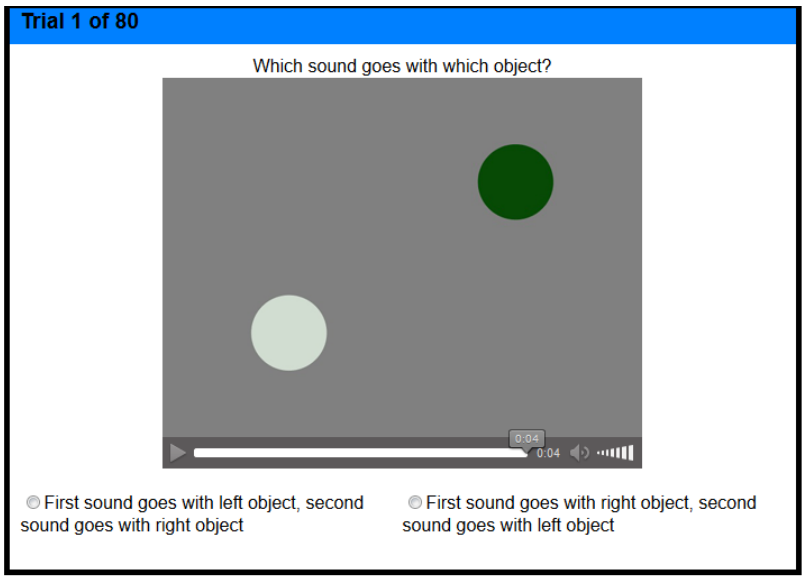


Figure 1: Example trial (high luminance, medium saturation, medium size, low position), as viewed by the participant after the video has been played. The participant could not see the radio-button decisions beneath the video until it had been played once.

4. Initial analysis and statistics

Reported analyses used data from all participants. Similar results were found for monolingual native English speakers alone.

Initial analysis of Level 1 stimuli established association strengths of each individual CMC. Figure 2(a) shows proportions of participants who chose high beeps for each stimulus at Level 1. In all cases, there were reliable and significant correspondences between sensory dimensions. The high beep was associated with stimuli with higher luminance, saturation, or position, or that were smaller. Because auditory stimuli were matched for physical amplitude, the high beep was probably perceived as louder (ISO, 2014). It is thus likely that the strength and direction of association is determined by pitch *and* loudness. This does not

affect our interpretation of the results, which concern how different *visual* dimensions combine in determining CMCs.

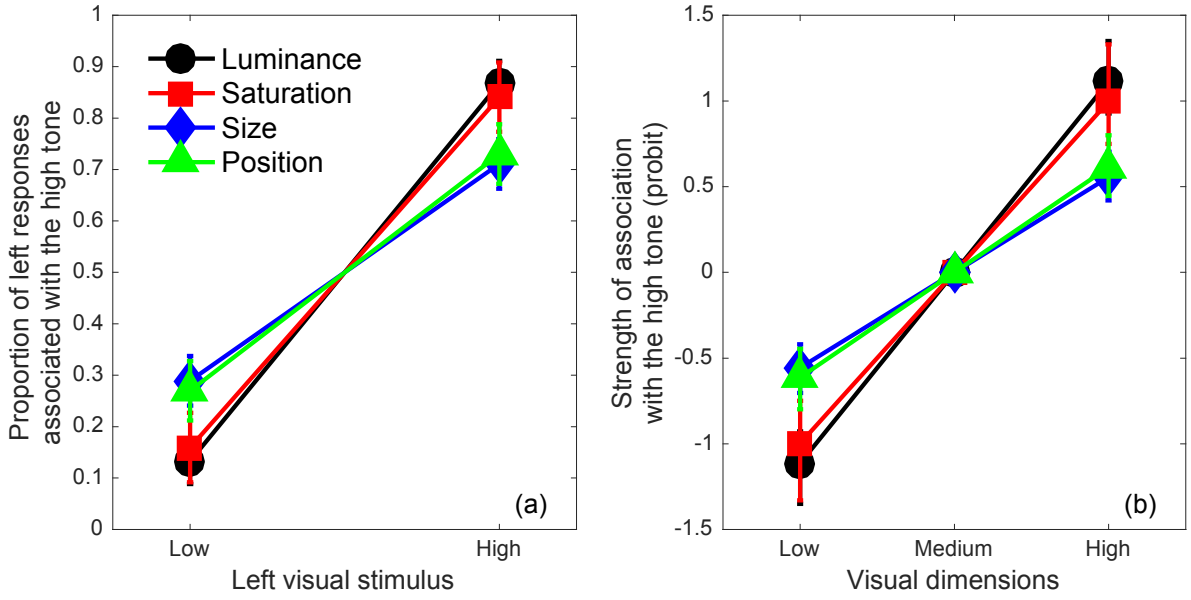


Figure 2: Results for the Level 1 conditions, in which the stimuli varied on only a single dimension. (a) The proportion of participants who chose a high pitched beep as the one that went with each stimulus. Error bars show binomial 95% confidence intervals. (b) The strength of the association between the frequency of the auditory stimulus, and each dimension of the visual stimuli, calculated using probit analysis (see text for details). 'Low' and 'high' map to 'large' and 'small' for the size dimension.

Association strengths were modelled 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 assume some variation in association strength across the population, modelled using a normal distribution. Using probits, we transformed proportions of participants choosing each beep for each visual stimulus, to quantify association strengths in units of the standard deviation of the variability (Thurstone, 1927):

$$p(R = LEFT \mid \mathrm{S}) = \phi(S) \qquad (1)$$

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

Probit values for Level 1 stimuli are plotted in Figure 2(b). 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 outcomes for stimuli containing variations in multiple dimensions. We predicted these results using each model as follows:

4.1. Summation

The simplest assumption is that association strengths will add:

$$S_{TOTAL} = S_{LUM} + S_{SAT} + S_{SIZE} + S_{POS} \quad (2)$$

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

4.2. Hierarchy

In this model, there is a hierarchy of CMCs. For any stimulus, the CMC is predicted by the dominant association. This is not necessarily the dimension with the strongest association when presented alone. Rather, it assumes a specific order in which dimensions are considered, with the association determined by the first dimension, within this order, on which stimuli differ. Since we tested four CMCs, there are 24 (4x3x2x1) possible hierarchies. We calculated correlations between predicted and actual responses for each stimulus, for all

hierarchies, and chose the hierarchy that best fit the data. This method provides considerable freedom to achieve the best fit; the other models contain no free parameters.

4.3. Majority

In this model, where there is a conflict between the directions of CMCs, the response is determined by majority vote, regardless of strengths of individual CMCs. If all stimulus dimensions, and experimental manipulations, had the same strength, then the predictions of the summation and majority models would agree. However, if for example one dimension was particularly dominant, then this might outweigh the combined effects of other dimensions that predicted the opposite response.

5. Results

For each model, we calculated correlations between predicted and actual responses for Level 2, 3 and 4 stimuli (Table 2). The summation model predicts the data well, with all correlations significant. Correlations for the majority model were significant, but lower than for the summation model. Correlations for the hierarchy model, which does not take account of all CMCs, were in all cases lower, and non-significant for Level 4 stimuli. Therefore, for stimuli containing multiple CMCs, *all* visual dimensions contribute to participant decisions.

Table 2: Correlation coefficients and significance levels for the fit 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

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 colour saturation and luminance, the width of the stimulus, and its distance from the centre of the screen, as covariates. Each was significant (Luminance: Wald χ^2=1734.8, p < .001; Saturation: Wald χ^2=424.1, p < .001; Size: Wald χ^2=348.0, p < .001; Position: Wald χ^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 χ^2 =4.10; p=0.043); luminance, saturation and size (Wald χ^2 ==9.00; p=0.003); and saturation, size and position (Wald χ^2 =6.69; p=0.01).

We also predicted main effect and two-way interaction results using probits for Level 1 stimuli (Figure 2b), using a linear regression after centring the data for each dimension. The results were combined according to Equation 2, and the predicted proportion of 'left' responses calculated from the resulting probit value. These results are plotted in Figures 3 (main effects) and 4 (two-way interactions). This gives a good prediction for luminance and size. However, the effect of saturation, in particular, is 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 different was tested using a generalised linear model with saturation as a covariate, fit separately to data from different levels of luminance. The effect of saturation was significantly greater for neutral luminance stimuli (b=.026 (95% confidence limits: .024-.028); Wald χ^2=739.2; p < .001) than for those with low (b=.001 (-.0001-.0003); Wald χ^2=14.934; p = .24) or high (b=.003; 0.002-0.005); Wald χ^2=15.8; p < .001) luminance. Participants' responses were only strongly influenced by saturation when luminance was neutral.

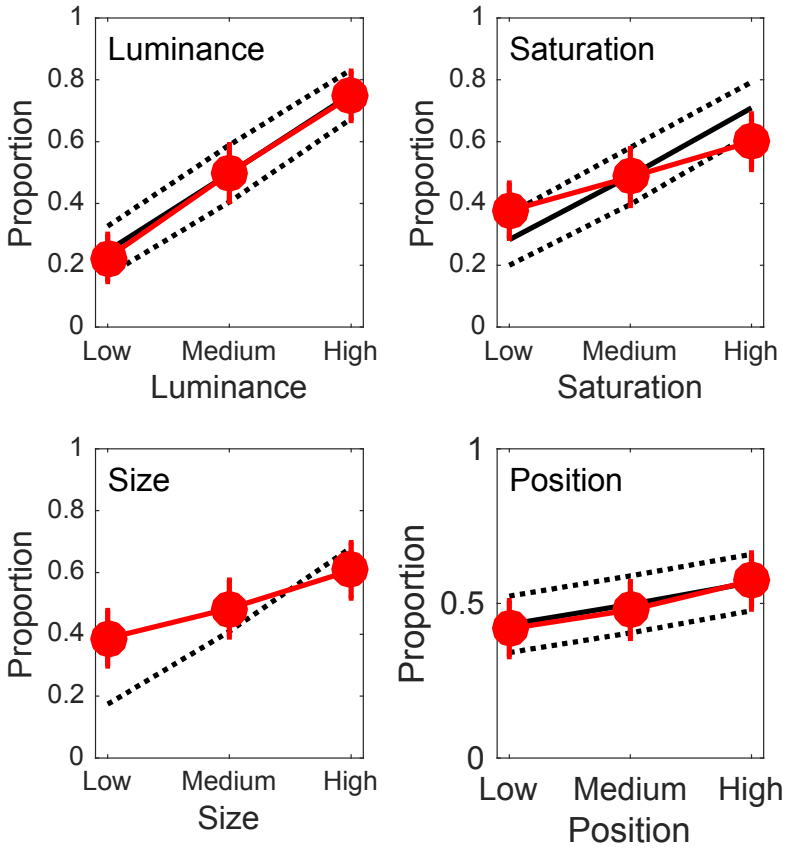


Figure 3: Proportion of left responses associated with the higher tone, as a function of each visual dimension, for stimuli pooled over all other visual dimensions. The red symbols indicate the participants' responses, the solid black line the predictions of the probit model. Error bars, and dotted black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

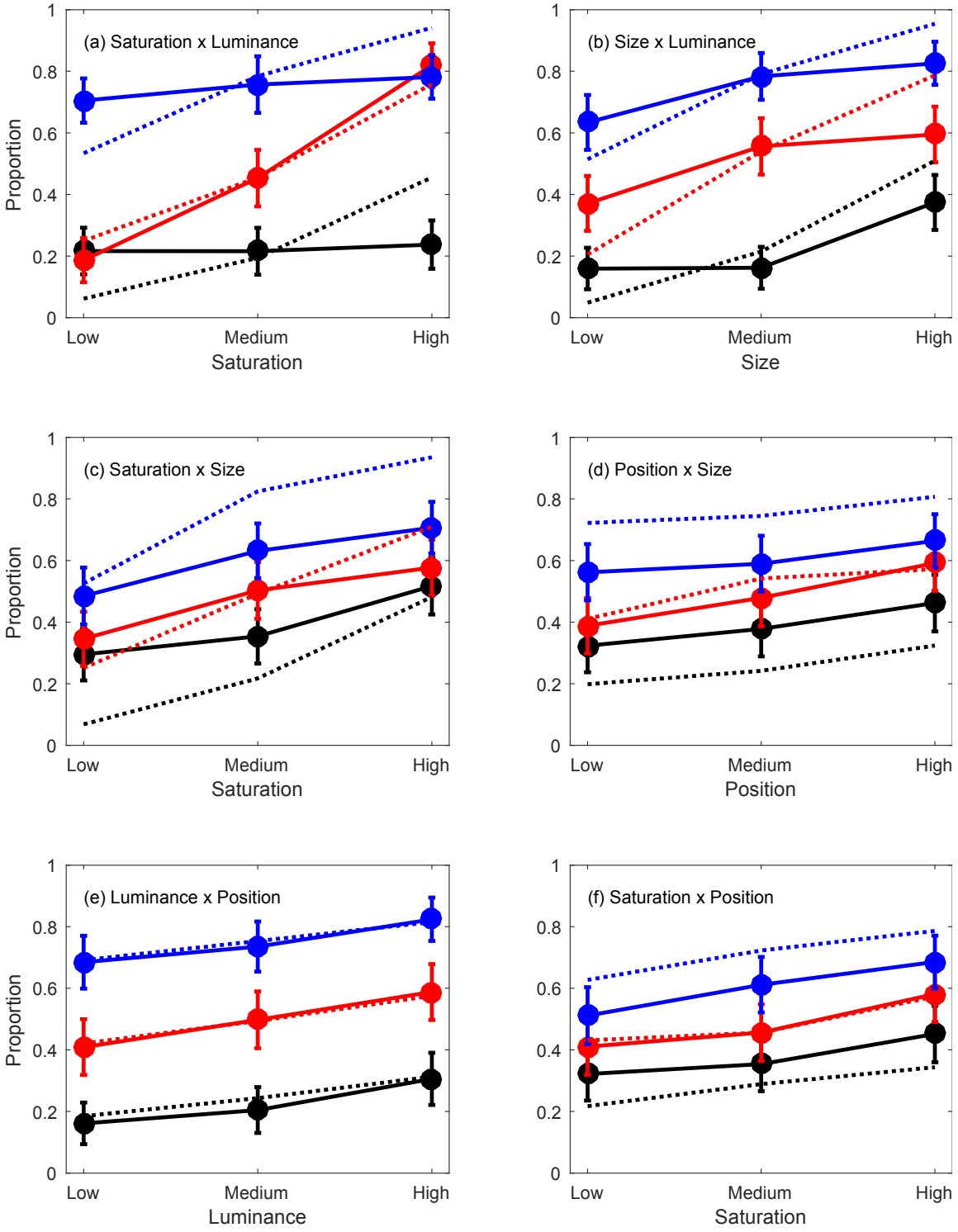


Figure 4 Proportion of left responses associated with the higher tone, as a function of each pair of visual dimensions, for stimuli pooled over all other visual dimensions. In all cases, one dimension is plotted on the horizontal axis, the black, red and blue symbols represented the 'low', 'medium' and high values on the other dimension, respectively. The dashed lines show the predictions of the probit model. Error bars, and dotted lines, indicate 95% binomial confidence limits of the data and model fits, respectively.

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 be obtained if stimuli are analysed in a different colour space. For example the CIE Luminance, Chroma Hue (LCh) might be considered more appropriate, since distances in this space relate to just-noticeable-differences in colour. We recalculated our probit predictions in LCh colour space, but found that there was little difference in the overall fit of the model regardless of whether the HSL (r^2=0.534 across all stimuli) or LCh space (r^2=0.525) was used.

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 factors.

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

6. 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 differing 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 participant responses suggests a general strategy of weighting available visual cues to determine the best auditory match, perhaps via neural intensity matching (Spence, 2011) or a generalised system for dealing with magnitude (Walsh, 2003). However, we need to account for the few results which violate the model (the lower effect of saturation at low and high luminances, and the three-way interactions of luminance/position/size, luminance/saturation/size and saturation/position/size). The decreased effect of saturation at low and high luminances appears to be the result of Garner interference. In Garner's (1976) paradigm, participants are presented with stimuli varying along two perceptual dimensions and make a decision 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 dimensions are incompletely integrated, so participant decisions are influenced by each dimension at unequal relative strengths, but also by interactions between different dimensions.

6.1. 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 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 others (e.g. a structural CMC that becomes encoded in language). These hypothetical CMCs would likely have more effect on perceptions and decisions than those which occur at only one level. That is, if a multiple-level CMC conflicts with a single-level CMC, the multiple-level CMC is likely to 'win'.

6.2. 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 informally reported by some participants to be tedious, participants will 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 is 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 only tested one auditory dimension. Therefore, we cannot know whether our findings are specific to the relationship of visual dimensions with pitch, or whether the same interactions will occur if we test, 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 relationships between CMCs could appear if we 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 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.

Last, 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 judgements (e.g. Parise & Spence, 2009) would allow exploration of whether interactions occur at perceptual or decisional levels. An analysis of response

times would also allow exploration of the impact of multiple conflicting or converging cues on decision making.

6.3. 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 non-summative 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 emphasised.

Our findings are also helpful for designers of sensory substitution devices such as the vOICe (Meijer, 1992), which allow 'translation' of information from one sense to another (for a review see Hamilton-Fletcher & Ward, 2013). Having explicit knowledge about relationships between different CMCs will allow 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).

7. Acknowledgements

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

8. 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*(3), 416-43. doi:1.1068/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*(1), 17-23. doi: 1.1016/j.foodqual.201.06.007

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*. 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*(2), 165-172. doi:1.1016/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*(5), 494-507. doi: 1.3758/BF03207885

Crisinel, A.-S., & Spence, C. (2011). A fruity note: Crossmodal associations between odors and musical notes. *Chemical Senses*, *37*, 151-158. doi: 1.1093/chemse/bjr085

Crisinel, A.-S., & Spence, C. (2012). Assessing the appropriateness of 'synaesthetic' messaging on crisps packaging. *Food Quality and Preference*, *26*(1), 45-51. doi: 1.1016/j.foodqual.2012.03.009

Doel, K. (2003). SoundView: Sensing color images by kinesthetic audio. Proceedings of 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*(4), 449-467. doi: 1.1177/0305735610377592

Evans, K.K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*, 6. doi:1.1167/1.1.6

Garner, W.R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *8*(1), 98-123. doi: 1.1016/0010-0285(76)90006-2

Hamilton-Fletcher, G. (2015). How touch and hearing influence visual processing in sensory substitution, synaesthesia and cross-modal correspondences (Doctoral dissertation, University of Sussex).

Hamilton-Fletcher, G., & Ward, J. (2013). Representing colour through hearing and touch in sensory substitution devices. *Multisensory Research*, *26*(6), 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*(2), 219-238.

ISO (2004) ISO 226:2003 Acoustics – normal equal-loudness-level contours.

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*(2), 199-222. doi: 1.1080/00221309.1942.10545166

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: 1.3758/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*(12), e0144013. doi:1.1371/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*(4), 1089-1099. doi: 1.1016/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: MIT Press.

Meijer, P.B.L. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, *39*, 112-121. doi: 1.1109/1.121642

Melara, R.D., & Marks, L.E. (1990). Interaction among auditory dimensions: timbre, pitch and loudness. *Perception and Psychophysics*, *48*(2), 169-178. doi: 1.3758/BF03207084

Mondloch, C.J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioural Neuroscience*, *4*(2), 133-136. doi: 1.3758/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:1.1371/journal.pone.0005664

Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, *7*, 269-284. doi: 1.1016/j.plrev.201.04.006

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971-995. doi: 1.3758/s13414-010-0073-7

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*, *22*(1), 245-26. doi:1.1016/j.concog.2012.12.006

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171-178. doi: 1.1016/S0959-4388(96)80070-5

Trommershäuser, J. Körding, K. and Landy, M.S. (2011). *Sensory cue integration*. Oxford: 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: 1.1068/p7149

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483-488. doi: 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). Birkhäuser Basel.

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*(6), 365-379. doi: 1.1068/i0586

Woods, A.T., Velasco, C., Levitan, C.A., Wan, X., & Spence, C. (2015). Conducting perception research over the Internet: a tutorial review. *PeerJ PrePrints*, *3*, e1138. doi: 1.7287/peerj.preprints.921v1