CrossMark

Are crossmodal correspondences relative or absolute? Sequential effects on speeded classification

Riccardo Brunetti[1] · Allegra Indraccolo[1] · Claudia Del Gatto[1] · Charles Spence[2] ·
Valerio Santangelo[3,4]

Abstract Crossmodal correspondences have often been demonstrated using congruency effects between pairs of stimuli in different sensory modalities that vary along separate dimensions. To date, however, it is still unclear the extent to which these correspondences are relative versus absolute in nature: that is, whether they result from pre-defined values that rigidly link the two dimensions or rather result from flexible values related to the previous occurrence of the crossmodal stimuli. Here, we investigated this issue in a speeded classification task featuring the correspondence between auditory pitch and visual size (e.g., congruent correspondence between high pitch/small disc and low pitch/large disc). Participants classified the size of the visual stimuli (large vs. small) while hearing concurrent high- or low-pitched task-irrelevant sounds. On some trials, visual stimuli were paired instead with "intermediate" pitch, that could be interpreted differently according to the auditory stimulus on the preceding trial (i.e., as "lower" following the presentation of a high pitch tone, but as "higher" following the presentation of a low pitch tone). Performance on sequence-congruent trials (e.g., when a small disc paired with the intermediate-pitched tone was preceded by a low pitch tone) was compared to sequence-incongruent trials (e.g., when a small disc paired with the intermediate-

pitch tone was by a high-pitched tone). The results revealed faster classification responses on sequence-congruent than on sequence-incongruent trials. This demonstrates that the effect of the pitch/size correspondence is relative in nature, and subjected to trial-by-trial interpretation of the stimulus pair.

In recent years, there has been a rapid growth of research interest in the study of crossmodal correspondences, defined as the brain's tendency to associate different features of stimuli, or dimensions, across the senses (see Spence, 2011; Spence & Deroy, 2013, for reviews). To date, correspondences have been documented between several different dimensions, such as between auditory pitch and visual size: in this case, participants typically associate the smaller of two objects with the higher-pitched of two sounds, and the relatively larger object with the lower-pitched sound (Bien, ten Oever, Goebel, & Sack, 2012; Evans & Treisman, 2010; Gallace & Spence, 2006; Parise & Spence, 2012), as reflected in faster reaction times (RTs; and possibly also lower error rates).

While the different kinds of correspondences and their effects on perception/performance have been widely documented, the question of whether these effects are *absolute* or *relative* is a complex one. In other words, whether a specific correspondence (e.g., auditory pitch/visual size) produces effects only with a pre-determined audiovisual range of stimuli (e.g., a specific high pitch and a specific small size) or whether instead these effects are context-sensitive (e.g., a pitch higher than the preceding one or a visual stimulus smaller than the one presented on the preceding trial) is the question that is

✉ Riccardo Brunetti
   riccardo.brunetti@gmail.com

[1] Cognitive and Clinical Psychology Laboratory, Department of Human Sciences, Università Europea di Roma, Rome, Italy

[2] Department of Experimental Psychology, Oxford University, Oxford, UK

[3] Department of Philosophy, Social Sciences & Education, University of Perugia, Perugia, Italy

[4] Neuroimaging Laboratory, Santa Lucia Foundation, Rome, Italy

⧌ Springer

being investigated. When the correspondence is relative, it is, then, not clear whether this context-sensitive effect is attributable to a general context (e.g., the experimental block in which a stimulus happens to be presented) or rather to the specific sequence defined by the transition from the stimuli presented on the immediately preceding trial(s). Of course, at the outset, it should be stated that there is evidence that correspondences are of different types in this regard: while many correspondences might be context-sensitive, certain others might, perhaps, be absolute (Marks, Szczesiul, & Ohlott, 1986; see also Spence & Deroy, 2013). However, it seems more likely, at least to us, that a specific mixture of absolute and context-sensitive is present in every correspondence. Namely, every correspondence might have effects within certain absolute ranges of stimulus values (e.g., between certain values, intensities, degrees), while nevertheless still being sensitive to relative variations inside the boundaries of those ranges. In other words, all crossmodal correspondences might be relative, when falling within certain absolute boundaries.

Some studies reported that certain correspondences operate under absolute constraints. Guzman-Martinez, Ortega, Grabowecky, Mossbridge, and Suzuki (2012) used Gabor displays having one of three different spatial frequencies. The participants adjusted the auditory amplitude modulation (AM) rate and the pitch of the sound until it perceptually matched the visual spatial frequency of the Gabor display. The observers consistently matched a specific auditory AM rate to each of the visual spatial frequencies. The authors demonstrated that these matches also persisted when the observer had to determine an auditory AM rate to only one visual spatial frequency (the same result was obtained between tactile AM rate and visual spatial frequency). Guzman-Martinez and his colleagues concluded that a consistent and absolute crossmodal mapping between visual spatial frequency and auditory AM rate does indeed exist (see Guzman-Martinez et al., 2012; for additional absolute effects, see Pedley & Harper, 1959; Smith, Grabowecky, & Suzuki, 2007).

Importantly, however, these results have been only partially confirmed by subsequent research. For instance, Orchard-Mills, Van der Burg, and Alais (2013), tried to replicate the relationship between AM auditory stimulus and visual spatial frequency that had been obtained by Guzman-Martinez et al. (2012). Even though these authors confirmed the existence of a linear relationship between AM auditory noise and visual spatial frequency (so pointing toward an absolute mapping), the subsequent evidence that this crossmodal interaction is flexible and based on relatively unspecific frequencies (e.g., the evidence that correspondence effects also appear with variable pairs of visual and auditory frequencies) leaves room for their being both absolute constraints as well as some degree of relative context-dependent flexibility (e.g., within those constraints).

In general, most researchers have argued that correspondences are relative rather than absolute (certainly as far as

the crossmodal correspondence between auditory pitch and visual elevation is concerned; see Marks, 1974, 1989; Spence, 2011). For instance, Chiou and Rich (2012) demonstrated that the presentation of a 900 Hz pure tone led to an upward shift of participants' visual spatial attention when the other tone in a block of trials was 100 Hz. However, when the same tone was presented in another block of trials that contained 1700 Hz tones, it led to a downward shift of participants' spatial attention instead. While these results demonstrate that many correspondences are relative in nature, they do not clarify the level of flexibility of these context-sensitive effects. That is, one might ask whether is a consistent repetition of the crossmodal pair of stimuli across a large number of trials required for the correspondence to be established (cf. Chiou & Rich, 2012; Ernst, 2007) and/or are correspondences flexible enough to be set or updated on a trial-by-trial basis (cf. Bernstein & Edelstein, 1971)? Probably, there is some kind of interaction between long-term established correspondences (e.g., structural or statistical, see Spence, 2011) and recent trial history.

For many years now, different authors have been exploring the hypothesis that crossmodal correspondences are relative by analyzing sequential effects (Martino & Marks, 2001; Walker, 2012). For instance, Walker and Walker (2016) recently explored the flexibility of crossmodal correspondences by using intermediate stimuli (in this specific case, the size of hidden response keys) that could be interpreted as smaller or bigger according to the preceding stimuli. In their investigation, they confirmed the relative nature of the haptic size/ visual brightness crossmodal mapping, obtaining opposite congruency effects with the same stimulus (e.g., the response key of a certain size was congruent with darker or lighter stimuli according to which stimulus preceded it). These results were interpreted by the authors as corroborating the idea that the crossmodal correspondence between haptic size and visual brightness is conceptual in nature (see also Walker, 2016), i.e., this mapping would reflect interactions that are free from specific, absolute values, being dependent on relationships that can be mapped onto different stimuli.

However, whether these relative effects are under the influence of rapidly changing sequence features on a speeded classification task, on a trial-by-trial basis, is still unexplored (see Spence & Deroy, 2013, footnote 7; see also Spence, Nicholls, & Driver, 2001, for a similar approach concerning the contextual effect on the current trial performance as a function of what happened on the preceding trial). Here, we aimed at investigating this issue by presenting participants with sequences of audiovisual stimuli; namely, visual discs paired with auditory tones. The participants had to classify whether each disc was large or small while ignoring task-irrelevant auditory stimuli. In this speeded classification task, the small and large size discs could be paired with a high, low, or medium pitch tone. The trials featuring a high or low

pitched tone, can be congruent or incongruent relative to the visual stimuli (stand-alone congruency). Crucially, the interpretation of the medium pitch tone, and the probability for a crossmodal correspondence effect occurring, was contextually related to the previous trial (Spence et al., 2001): a medium pitch tone could be considered as relatively high or low depending on whether it was preceded by a low or high pitch tone, respectively (sequential congruency).

If the crossmodal association under study is flexible enough to be established on a trial-by-trial basis, and if the sequence is presented with a pace that can elicit sequential effects, we would expect to observe consistent correspondences modulating the speeded classification task. That is, a large disc paired with a medium pitch tone preceded by a higher pitch tone could be defined as "sequentially congruent". By contrast, a small-sized disc paired with the same medium pitch tone preceded by a higher pitch tone would be defined as "sequentially incongruent". If the crossmodal mapping is relative and context-dependent, faster classification performance would be expected for sequentially-congruent than for sequentially-incongruent trials, indicating the rapid establishment of the audiovisual correspondence. Finally, we investigated if this sequential effect may also affect the stand-alone congruency (e.g., the congruency that links size and pitch in a single trial with a high or low sound).

Methods

Participants

Thirty-one undergraduates took part in the experiment for course credit. One participant was excluded from further analyses due to poor performance (accuracy < 2 SD from the group average) leaving a total of 30 participants (7 males; mean age = 21.3 years; SD 2.16; range = 19–26 years).

All of the participants reported normal or corrected-to-normal hearing and vision. The participants were naïve as to the purpose of the study, which has been approved by the Ethics Review Board of Università Europea.

Apparatus and materials

Stimulus presentation, conditions, pseudo-randomization, and the recording of responses were all controlled by a custom-made script in the PsychoPy (v 1.80) programming environment, running on a 15" 2.4 GHz MacBook Pro laptop computer. The sounds were presented through Edirol MA-15D speakers at a comfortable listening level. The auditory component consisted of one of three sinusoidal tones: high (4500 Hz), medium (1200 Hz), or low (300 Hz), matched for

perceived loudness[1] (Scharf, 1978). Each tone was presented for 300 ms and included two 5 ms fading-ramps. The visual stimuli consisted of two black discs of small (2.1° visual angle) and large (5.2° visual angle) size, presented at the centre of a grey (50%) display for 300 ms. In order to ensure the co-localization of the auditory and visual stimuli, the display was set exactly in the middle of the two loudspeakers.

Design and procedure

The participants were seated in a testing room, facing a monitor placed approximately 60 cm in front of their head. They were instructed to stay still in front of the screen. After having received the instructions, and after having completed a few practice trials, the experiment began. This consisted of a visual speeded classification task in which the participants had to discriminate the size of the disc (large vs. small), while ignoring the simultaneously presented task-irrelevant auditory tone in an uninterrupted and fast sequence. Each trial started with the presentation of a fixation cross for 1200 ms followed by the presentation of a synchronized audiovisual stimulus pair for 300 ms, with an inter-trial interval of 1200 ms during which time the participants had to respond (see Fig. 1). The inter-trial interval was set at 1200 ms, to elicit sequential effects (see Spence et al., 2001).

The task was divided into two blocks of trials. In the first block (Baseline block), we assessed whether our stimuli and experimental settings would replicate previous findings demonstrating the Pitch/Size correspondence (e.g., using two pitches and two sizes, we expected to observe facilitation for crossmodally congruent stimuli, e.g., high pitch tone/small disc and low pitch tone/large disc). In the second block (Experimental block, see Fig. 1), a sound with an intermediate pitch (i.e., a sound lower in pitch than the high-pitched tone and higher in pitch than the low-pitched tone) was included in the task-irrelevant auditory dimension. "Ambivalent" trials were defined as those that included this intermediate pitch sound. This means that this intermediate sound could be interpreted as low pitched if preceded by a high-pitched tone or as high pitched if preceded by a low-pitched tone. Accordingly, if paired with a small or large disc, the intermediate sound could generate a "sequentially incongruent" (e.g., large disc, with medium pitch interpreted as high or a small disc, with medium pitch interpreted as low) or a "sequentially congruent" trial (e.g., large disc, with medium pitch interpreted as low or a small disc, with medium pitch interpreted as high).

---

[1] By using an equal loudness contour, we have pinpointed the db level that matched the loudness (measured in phon) for our pitches. For each tone, we adjusted the db on audacity software and measured them with a sound meter, to verify the adjustment.

Sequence of events and types of trials

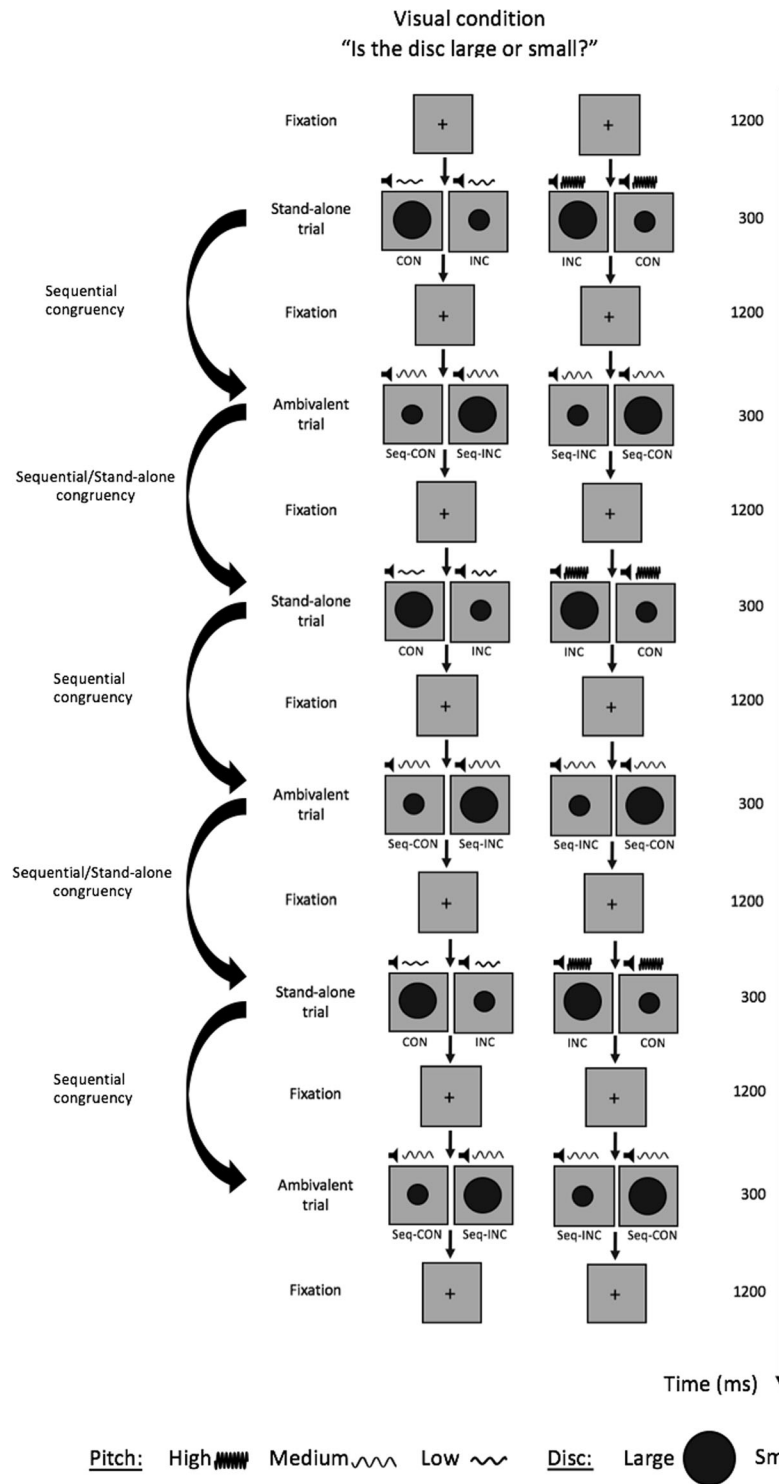## Visual condition
"Is the disc large or small?"

Fig. 1 Speeded classification task in the Experimental block. The participants indicated whether the *disc* was large or small while ignoring the task-irrelevant sounds. In the "ambivalent" trials, the task-irrelevant auditory stimulus consisted of the presentation of a sound having an intermediate pitch. The interpretation of this ambivalent auditory stimulus was based on the preceding stand-alone trial, thus giving rise to sequentially congruent vs. incongruent ambivalent trials. For instance, a sequentially congruent trial might consist in a large disc paired with a "relatively" low pitch (i.e., an intermediate pitch tone preceded by a high pitch tone in the previous trial). The sequential congruency is also applicable to stand-alone trials, thus reinforcing the stand-alone congruency as a consequence of the preceding trial

"Ambivalent" trials were alternated with "stand-alone" trials in the Experimental block (see Fig. 1). "Stand-alone" trials were defined as those featuring only high or low sounds. These trials were included in order to establish the subsequent congruency or incongruency of the "ambivalent" trials (e.g., the ones featuring a sound with an intermediate tone). These stand-alone trials featured a "stand-alone congruency", e.g., the pitch/size couple that is presented in each stand-alone trial could be either congruent or incongruent in itself. Moreover, these stand-alone trials also featured a sequential congruency: a congruency effect strengthened by the change implied by the preceding ambivalent trial. Namely, "sequentially/stand-alone congruent" were defined as those two consecutive congruent stand-alone trials where both pitch and size vary in a congruent direction together. Conversely, we called "sequentially/stand-alone incongruent" those two consecutive incongruent stand-alone trials where the pitch and size do not vary together congruently.

Importantly, the presentation of the stimuli was pseudo-randomized such that each audiovisual combination at the "ambivalent" trial level was equiprobably preceded by each audiovisual combination on the "stand-alone" trials. Consequently, half of the time, the stand-alone / ambivalent pair was sequentially congruent, while on the other half of trials it was sequentially incongruent, thus allowing us to assess the impact of the different types of sequences on speeded classification performance.

In both blocks (baseline and experimental), the participants had to respond as rapidly and accurately as possible on every trial. Participants responded by pressing, with the right hand, the "M" or "N" keys on the keyboard as a function of whether the disc was small or large. In order to verify whether the participants were focused on the visual task but still able to discriminate the task-irrelevant sound, we randomly presented a display requiring the participants to judge the pitch of the sound presented on the last trial, by pressing "1" (low), "2" (intermediate), or "3" (high), with their left hand. These "catch" trials appeared every 10–30 trials.[2] Overall, in the baseline block, the participants were presented with 144 trials (72 congruent and 72 incongruent); in the experimental block, they were presented with 320 trials, including 160 "ambivalent" trials featuring intermediate task-irrelevant stimuli, i.e., 40 repetitions for each of the four main conditions (large disc and medium pitch, congruent or incongruent; small disc and medium pitch, congruent or incongruent; congruency was defined by the preceding sound); "stand-alone" trials, 160 in total, were divided in two categories, congruent or incongruent. The experiment had a total duration of approximately 20 min.

Results

RT and accuracy for Baseline block (stand-alone congruency only)

Those trials in which the participants failed to provide a response (overall 22.4% of trials on average)[3] were removed from the analysis of RT and accuracy. A two-tailed paired t-test between stand-alone crossmodally congruent and stand-alone crossmodally incongruent trials was performed on the RT data. This analysis revealed a significant main effect of Congruency [$T(29) = -4.726; P < .001$], indicating faster classification responses when the task-irrelevant dimension was congruent (mean 459 ms; SE ± 11) than when it was incongruent (mean 474 ms; SE ± 11).

A similar t-test was performed on the accuracy data from the baseline block. This analysis corroborated our main finding, yielding a significant effect of stand-alone Congruency [$T(29) = 3.784; P = .001$]. In fact, the participants responded more accurately on the congruent trials (mean .920; SE ± .01) than on the incongruent trials (mean .880; SE ±.01). These results therefore replicate previous findings related to the facilitation of speeded classification performance for crossmodally matched congruent pitch/size stimuli (Evans & Treisman, 2010; Gallace & Spence, 2006).

RT and accuracy for the Experimental block

Next, we moved on to assessing the main aim of the present study, related to the assessment of whether pitch/size correspondences could be established on a trial-by-trial basis. As before, we removed miss responses from the analysis, thus resulting in the removal of 16.6% trials overall.

Sequential effects on ambivalent trials

A two-tailed paired sample t-test was used to compare participants' performance following sequentially-congruent or sequentially-incongruent target trials. The analysis revealed a significant effect of Target trial type [$T(29) = -3.507; P = .001$], indicating faster categorization responses for sequentially congruent trials (mean 457 ms; SE ± 11) as compared to those trials that were sequentially incongruent (mean 471 ms; SE ± 11).

A similar t-test performed on the accuracy data from the experimental block revealed no effect of Target trial type [$T(29) = .225; p = .824$].

---

[2] Overall, participants scored 71% correct on this task, well above chance level (33%), thus indicating that they were able to properly discriminate sounds.

[3] Since the next bimodal stimulus appeared relatively quickly there was a fast cut-off point for accepting a response (i.e., the speed of the sequence was such that a tight upper limit was set for correct RTs that would be incorporated in the analysis).

Stand-alone trials

Given that the participants had to respond on every trial, and given that stand-alone trials can be categorized as crossmodally congruent or incongruent by themselves, here we report performance on the inter-target trials without considering the subsequent target trials. A two-tailed paired sample *t*-test was performed on the RT data. However, we did not obtain a significant difference between crossmodally congruent and incongruent trials in this case [T(29) = .613; *P* = .544].

A similar *t*-test on the accuracy data failed to show significant differences between the crossmodally congruent and crossmodally incongruent trials [T(29) = .021; *P* = .983].

Sequential effect on stand-alone trials

A two-tailed paired sample *t*-test on the RT data revealed a significant main effect of sequential/stand-alone congruency [T(29) = −2.238; *P* = .033]: participants were faster with sequentially congruent trials (mean 455 ms; SE ± 11) than with sequentially incongruent trials (mean 471 ms; SE ± 11).

Concerning the accuracy data, a similar *t*-test revealed a significant result for sequential/stand-alone congruency [T(29) = −2.541; *P* = .017]. That is, the participants were more accurate with sequentially congruent trials (mean .933; SE ± .013) than with sequentially incongruent trials (mean .899; SE ± .014).

Discussion

The aim of the present study was to investigate the sequential effects (e.g., effects related to the immediately-preceding context) of the pitch/size crossmodal correspondence. Our hope was to be able to shed some light on the dependence of this correspondence on relative vs. absolute coding. Specifically, we explored the flexibility of this crossmodal correspondence by analyzing the results derived from speeded classification task on a trial-by-trial basis. Our findings confirm the crossmodal congruency effect of audiovisual corresponding trials on the latency of participants' speeded classification responses. Crucially, this congruency effect emerged even when a stimulus (here the intermediate pitch tone) could be considered 'higher' or 'lower' based on the "high" or "low" pitch heard in the preceding trial. Indeed, the "valence" of the intermediate stimulus changed according to the preceding trial, producing faster classification responses on ambivalent congruent trials than on ambivalent incongruent trials. This result shows that the same "inducer" (i.e., the intermediate stimulus) matched equally well with the different values of the visual modality, according to the sequential magnitude established on the preceding trial.

Moreover, the lack of any effect of stand-alone congruency and the effect of sequential congruency on stand-alone trials reveals that even established associations (stand-alone congruency, clearly detected in the baseline) are influenced by the context. This provides strong support for our hypothesis concerning the relative flexibility of this kind of crossmodal association. The mere presence of a third auditory stimulus gives rise to sequential effects that may affect stand-alone congruency as well.

The results of the secondary task, namely performance that was well above chance (>33%), revealed that participants were able to discriminate the task-irrelevant stimuli. It is possible to speculate that asking participants to focus their attention on the irrelevant modality could have affected their performance (Chiou & Rich, 2015; Spence & Deroy, 2013). However, our results would appear to rule out this latter possibility, given the high percentage of hit responses in the primary task, demonstrating that participants found it easy to discriminate the task-relevant stimuli without any substantial interference from the secondary task. Nevertheless, future research is undoubtedly needed to further assess this point.

The current findings therefore highlight that the pitch/size correspondence is context-sensitive, and it is able to adapt rapidly, producing different interpretations of one and the same stimulus, depending on the preceding context. This result is in accord with explanations of this correspondence as being conceptually based (Walker, 2016; or even strategically based, see Del Gato, Brunetti, & Delogu, 2016, for the case of visual elevation/auditory pitch). Here, we show that this correspondence is context-sensitive, and flexible enough to be updated on a trial-by-trial basis. Indeed, the effect of the intermediate stimuli was determined by the stimulus presented on the preceding trial, while the context of having three different sounds in the block affected the inter-trial stand-alone congruency effects. Interestingly, this trial-by-trial effect might explain the absence of the congruency effect when the task-irrelevant dimension does not change within a block (Bernstein & Edelstein, 1971; Gallace & Spence, 2006; Orchard-Mills, Van der Burg, & Alais, 2016). Following Melara and O'Brien (1987), our results support the claim that crossmodal correspondences are crucially affected when there is stimulus variation along the irrelevant dichotomized dimension, namely, only when the relevant and irrelevant dimensions are both varied on an unpredictable trial-by-trial basis. Moreover, the extent of this variation is also significant, as shown by the analysis of performance on the inter-target trials. Therefore, these congruency effects depend on inter-trial variation, since this provides a context in which to define relative stimulus values and thereby to highlight crossmodal relations between dimensions (Eitan, Schupak, Gotler, & Marks, 2014; Martino & Marks, 2000). From these results, we might assert that flexibility is a core feature of these correspondences.

Overall, these results are in line with previous studies suggesting that the effect of correspondences on behaviour is determined in more of a relative than an absolute manner (Ben-Artzi and Marks, 1995; Chiou & Rich, 2012; Gallace & Spence, 2006; Melara & O'Brien, 1987), highlighting the relevance of the transition from the preceding trial on the strength of pitch/size correspondence. The correspondences where relative effects have been directly measured now include visual brightness/haptic size (Walker & Walker, 2016), visual spatial attention/auditory pitch (Chiou & Rich, 2012), and visual size/auditory pitch (the present study). Taking into account the classification proposed by Spence (2011), the importance of the context in which we perceive the stimuli is in line with the concept of "statistical correspondence", a kind of correspondence originating from learning and reflecting an adaptive response of our brain to the regularities that are present in the world. The pitch/size correspondence has been explained exactly in terms of the stable correlation in the natural context between these two features (e.g., smaller objects tend to produce higher pitched sounds). Thus it makes sense that our brain takes advantage of the regularities (both those operating over the long and short term) that exist around us when deciding which of the many possible stimuli to integrate (von Kriegstein & Giraud, 2006; Spence, 2011). These regularities can systematically affect our responses at both perceptual (D'Ausilio, Brunetti, Delogu, Santonico, & Olivetti Belardinelli, 2010) and post-perceptual levels (i.e., short-term memory; Brunetti, Indraccolo, Mastroberardino, Spence, & Santangelo, 2017).

However, while the relative nature of this correspondence has been shown to be flexible enough to adapt quickly from one trial to the next, it may still be dependent on the specific stimulus magnitudes chosen. Namely, the values we used (specific pitches and sizes) may be well-suited to inducing adjustable congruity effects: these values may be facilitating some kind of crossmodal "tuning" that makes the congruency effect emerge. In other words, instead of being purely relative (e.g., entirely based on conceptual knowledge), the correspondence could have "ranges" of values that allow for congruency to appear (see also Parise, Knorre, & Ernst, 2014). Unfortunately, few authors focused on investigating these "ranges" (that may well be somehow "absolute"; see Guzman-Martinez et al., 2012; Lunghi & Alais, 2013; Lunghi, Binda, & Morrone, 2010), and thus this possibility needs further research.

Further research is also needed in order to investigate whether the source of these interactions is in sensory information processing or in later (more decisional) processing stages. Most studies in this field published to date have used tasks that minimize any decision/response selection (though see Parise & Spence, 2012, for an exception). As a result, they suggest that crossmodal correspondences might be integrated at a perceptual level. However, some other studies have demonstrated that post-perceptual processes might also contribute to these synesthetic congruencies. For example, Bien, Ten Oever, Goebel, and Sack (2012) have shown that if the multisensory integration is disrupted by transcranial magnetic stimulation (TMS), top-down signals no longer overrule the bottom-up information. This has the consequence of wiping out the pitch-size correspondence, and leaving the auditory locations unchanged (and easier to localize). The fact that top-down processes play a crucial role in the pitch/size correspondence provides additional evidence in favour of interpreting it as a conceptual correspondence (see Walker, 2016). Moreover, Stekelenburg and Keetels (2016) have recently demonstrated that the Colavita effect (the tendency to respond more to the visual aspect of simultaneously presented auditory and visual targets), induced by factors that contribute to the structural binding of audiovisual stimuli (Koppen, Alsius, & Spence, 2008), is not affected by crossmodal correspondences. This supports the view that audiovisual synesthetic associations are probably processed at a stage subsequent to the stage in which the Colavita effect occurs (for other evidence supporting the top-down nature of crossmodal correspondences see Orchard-Mills, Alais, & Van der Burg, 2013; Orchard-Mills et al., 2016). Finally, researchers have demonstrated that correspondences also influence other processes, such as spatial attention or working memory (Brunetti et al., 2017; Chiou & Rich, 2012). Specifically, the effects on exogenous spatial attention (Chiou and Rich, 2012) seem to show that these correspondences can also have an effect at earlier stages of processing. In light of such considerations, it might be interesting to investigate whether the relative and sequence-dependent effect that we have documented is able to affect participants' performance in other cognitive tasks (e.g., tasks focusing on mnemonic or attentional processes), and to investigate if and how different correspondences can be effective at different processing stages.

In conclusion, the results of the present study add novel insights concerning the influence of the crossmodal correspondence between pitch and size on human performance. The sequence effect could be a sign of the adaptive function of correspondences in terms of better adaptation to the statistical regularities of the environment. The dynamic interaction with the environment, reacting in a prompt and continuous manner, make us adjust to recent change.

References

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

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

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: crossmodal binding in pitch-size synesthesia: A combined TMS, EEG, and psychophysics study. *NeuroImage, 59*, 663–672.

Brunetti, R., Indraccolo, A., Mastroberardino, S., Spence, C., & Santangelo, V. (2017). The impact of cross-modal correspondences on working memory performance. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 819–831.

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

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

D'Ausilio, A., Brunetti, R., Delogu, F., Santonico, C., & Olivetti Belardinelli, M., 2010). How and when auditory action effects impair motor performance. *Experimental Brain Research, 201*, 323–330.

Del Gatto, C., Brunetti, R., & Delogu, F. (2016). Cross-modal and intra-modal binding between identity and location in spatial WM: the identity of objects does not help recalling their locations. *Memory, 24*, 603–615.

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

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

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

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

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

Koppen, C., Alsius, A., & Spence, C. (2008). Semantic congruency and the Colavita visual dominance effect. *Experimental Brain Research, 184*, 533–546.

Lunghi, C., & Alais, D. (2013). Touch interacts with vision during binocular rivalry with a tight orientation tuning. *PLoS ONE, 8*, e58754.

Lunghi, C., Binda, P., & Morrone, M. C. (2010). Touch disambiguates rivalrous perception at early stages of visual analysis. *Current Biology, 20*, R143–R144.

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

Marks, L. E. (1989). On crossmodal similarity: The perceptual structure of pitch, loudness and brightness. *Journal of Experimental Psychology, 15*, 586–602.

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

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

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

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

Orchard-Mills, E, Alais, D., & Van der Burg, E. (2013). Cross-modal associations between vision, touch and audition influence visual search through top-down attention, not bottom-up capture. *Attention, Perception, & Psychophysics, 75*, 1892–1905.

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

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

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

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

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

Scharf, B. (1978). Loudness. In E. C. Carterette & M. P. Friedmann (Eds.), Handbook of perception - Hearing (pp. 187–242). New York: Academic.

Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology, 17*, 1680–1685.

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

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

Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics, 63*, 330–336.

Stekelenburg, J. J., & Keetels, M. (2016). The effect of synesthetic associations between the visual and auditory modalities on the Colavita effect. *Experimental Brain Research, 234*, 1–11.

von Kriegstein, K., & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4*, e326.

Walker, L., & Walker, P. (2016). Cross-sensory mapping of feature values in the size–brightness correspondence can be more relative than absolute. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 138–150.

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

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