Psicológica (2019), 40, 85-104
doi: 10.2478/psicolj-2019-0006

Vertical Mapping of Auditory Loudness:

Loud is High, but Quiet is not Always Low

Laura Puigcerver[1,2,3,4*], Sara Rodríguez-Cuadrado[5], Víctor Gómez-Tapia[1,2,3] & Jordi Navarra[1,2,3,4]

[1] *Department of Cognition, Development and Educational Psychology, University of Barcelona, (Barcelona, Spain)*

[2] *Fundació Sant Joan de Déu, Sant Joan de Déu Hospital, (Barcelona, Spain)*

[3] *Children and Adolescent Mental Health Research Group. Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat (Barcelona, Spain).*

[4] *Child and Adolescent Psychiatry and Psychology Department, Sant Joan de Déu Hospital, Esplugues de Llobregat (Barcelona, Spain).*

[5] *Department of Developmental and Educational Psychology, Autonomous University of Madrid (Madrid, Spain).*

Although the perceptual association between verticality and pitch has been widely studied, the link between loudness and verticality is not fully understood yet. While loud and quiet sounds are assumed to be equally associated crossmodally with spatial elevation, there are perceptual differences between the two types of sounds that may suggest the contrary. For example, loud sounds tend to generate greater activity, both behaviourally and neurally, than quiet sounds. Here we investigated whether this difference percolates into the crossmodal correspondence between loudness and verticality. In an initial phase, participants learned one-to-one arbitrary associations between two tones differing in loudness (82dB vs. 56dB) and two coloured rectangles (blue vs. yellow). During the experimental phase, they were presented with the two-coloured stimuli (each one located above or below a central "departure" point) together with one of the two tones.

[*] **Corresponding author:** Laura Puigcerver, Fundació Sant Joan de Déu, Hospital Sant Joan de Déu – c/ Santa Rosa 39-57, Edifici Docent, 4th floor. 08950 Esplugues de Llobregat (Barcelona), Spain. E-mail: l.puigcerver@gmail.com. **Acknowledgments:** We would like to thank Irune Fernández-Prieto for her comments on an earlier version of the article. The present study was supported by the PSI2015-71143-P grant from Ministerio de Economía y Competitividad (MINECO), Spain to J.N.; and a FPI scholarship (MINECO) to L.P.

Participants had to indicate which of the two-coloured rectangles corresponded to the previously-associated tone by moving a mouse cursor from the departure point towards the target. The results revealed that participants were significantly faster responding to the loud tone when the visual target was located above (congruent condition) than when the target was below the departure point (incongruent condition). For quiet tones, no differences were found between the congruent (quiet-down) and the incongruent (quiet-up) conditions. Overall, this pattern of results suggests that possible differences in the neural activity generated by loud and quiet sounds influence the extent to which loudness and spatial elevation share representational content.

We live in a multisensory environment. The perceptual relationships between perceptual features gathered through different sensory modalities (e.g., pitch and spatial elevation or the size of objects) have been intensively studied for the last decades (see Parise, 2016; Spence, 2011). In fact, many *crossmodal correspondences* have been described to date. As a paradigmatic example, the representation of pitch is somehow linked to the representation of other perceptual features such as verticality (e.g., see Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006), visual size (see Bien, ten Oever, Goebel, & Sack, 2012) and brightness (Marks, 1987). High and low frequency sounds tend, for example, to be associated with high and low spatial positions, respectively (see Deroy, Fernandez-Prieto, Navarra, & Spence, 2018; Spence, 2011; Spence & Deroy, 2013 for reviews). Some of these correspondences have also been observed in prelinguistic infants during the first year of life (see Fernández-Prieto, Navarra, & Pons, 2015, for the association between pitch and size; Dolscheid, Hunnius, Casasanto, & Majid, 2014; Walker et al., 2010, 2014; for the association between pitch and spatial elevation; although see Lewkowicz & Minar, 2014). The association between pitch and verticality is also present in a vast majority of languages, where pitch is described as either *high* or *low* (Spence, 2011). Interestingly, this association is present even in cultures that lack this explicit linguistic label (Parkinson, Kohler, Sievers, & Wheatley, 2012). This may perhaps explain the extensive literature addressing this perceptual association, as well as the reason for pitch being the preferred dimension for sonification (Dubus & Bresin, 2013). In addition, a similar crossmodal phenomenon can also be observed in the representation of loudness (understood as the perceived amplitude of sounds): we raise the volume or turn it down. In fact, acoustic features other than pitch (e.g., loudness) can also be expressed in terms of verticality (see Walsh, 2003).

Before getting into more detail about the relationship between loudness and verticality, it is mandatory to review “generalist" theories of magnitudes such as the one proposed by Walsh (i.e., the so-called "a theory of magnitude" or "ATOM"). ATOM assumes the involvement of the same cognitive (and perhaps neural) multimodal resources for the representation of quantifiable stimulus properties (Bueti & Walsh, 2009; Walsh, 2003). According to this perspective, features of objects stimulating different sensory modalities would have an amodal representation based on magnitude. ATOM offers a common basis for space, time and quantity (Walsh, 2003), which would be linked by a common metric for action and the share of several resources to process information and to achieve certain goals (see Walsh, 2003). There is mounting behavioural and neuropsychological evidence supporting ATOM (Bien, ten Oever, Goebel, & Sack, 2012b; Evans & Treisman, 2011; Foster, Halpern, & Zatorre, 2013; Foster & Zatorre, 2010a, 2010b; Parise & Spence, 2009; Rusconi et al., 2006). A strong example in the specific case of space and number comes from the so-called "distance effect", for which it is easier to compare numbers that are far away. The proposed reason behind this effect is the presence of a mental number line for which we compare the locations of numbers, hence showing this relationship between number and space. Similarly, strong evidence is provided too in favour of the relationship between time and space, from a neuropsychological perspective (see Critchley, 1953, cited in Walsh, 2003).

Coming back to roots of Walsh's theory (2003; Bueti & Walsh, 2009), to understand the commonality of objects sharing "quantifiable stimulus properties", means to speak about "prothetic dimensions" or, in other words, those dimensions that can be understood as being "more" or "less" than. Roughly following his rationale, it is licit to understand loudness as a prothetic dimension. In this context, loudness would be polarized as "loud" or "quiet", just like spatial elevation would be classified in terms of "upper" or "lower". The two relative poles of spatial elevation, or henceforth, loudness, would tend to be associated and therefore represented together in the same end of the polarization. If we understand loudness as a magnitude, just as numbers are, a practical example of this common magnitude code would be present in the relationship between musical and spatial representation. The shared representation of different acoustic features and verticality has an impact on musicians' spatial abilities. In fact, it has been reported that professional musicians show above-average visuospatial reasoning abilities (Sluming, Brooks, Howard, Downes, & Roberts, 2007), mostly on the vertical dimension (Brochard, Dufour, & Després, 2004). Not without controversy, recent evidence suggesting the presence of difficulties in the spatial representation of external objects in individuals with amusia would also be consistent with this idea (see Douglas & Bilkey, 2007; though see Tillmann et al., 2010).

Returning to the specific case of loudness, it seems to have a link with other perceptual attributes. Although being less researched than pitch, previous studies have observed crossmodal associations between loudness and brightness (Bond & Stevens, 1969; Marks, 1974, 1987; Root & Ross, 1965; Stevens & Marks, 1965; Wicker, 1968), and between loudness and visual size (Smith & Sera, 1992). An association was reported between big/small visual stimuli and loud/quiet sounds, respectively (Smith & Sera, 1992). A correspondence between loudness and brightness has also been shown in prelinguistic infants (Lewkowicz & Turkewitz, 1980). Loudness has also been related to changes in distance, possibly due to the fact that objects approaching us tend to produce louder sounds than objects that move away from us. As an example, people seem to use loudness to determine the length of visually hidden objects (see Carello, Anderson, & Kunkler-Peck, 1998). Likewise, objects located nearby tend to be visually larger and acoustically louder than distant objects (Burro & Grassi, 2001; Cabrera & Tilley, 2003a, 2003b; Lipscomb & Kim, 2004; Walker, 1987). The intensity of sounds seems to have an adaptive, alertness-related meaning: we can infer the distance with respect to us and the size of an approaching object by the intensity of its sounds (Zahorik, Brungart, & Bronkhorst, 2005). The perception of an approaching sound (that is, increasing in loudness) can also enlarge the distance with respect to our bodies in which we can feel comfortable and safe, thus provoking a behavioural response (Ferri, Tajadura-Jiménez, Väljamäe, Vastano, & Costantini, 2015). We even have a finer discrimination for louder tones than for quieter tones (Schröger, 1996), as well as a perceptual bias for changes in rising loudness, which are interpreted by the system as larger than physically-equal changes in falling loudness (Neuhoff, 1998). These perceptual biases have implications for the alertness system, especially if a behavioural reaction is needed: if an object is approaching us, we may need to be very accurate perceiving any possible changes in the intensity of the sounds it produces to avoid it, if needed. These mechanisms of finer perception and reaction can be crucial if the object approaches us at a great speed. In contrast, if the object is far from us or getting further away, our judgment of the intensity of the sound would be less critical, as a rapid flight response may not be needed. In relation to this, higher emotional responses have been observed for sounds increasing in intensity, simulating an approaching effect, than for sounds decreasing in intensity. These effects are present especially when sounds have a negative emotional valence (Tajadura-Jiménez, Väljamäe, Asutay, & Västfjäll, 2010). An increase in loudness can therefore be experienced as a warning cue, interpreted as a potential danger approaching (Neuhoff, 1998). This "bias" for louder sounds is also reflected at a neural level. Several electroencephalography (EEG) studies have found larger amplitudes in event-related potentials (ERPs) such as the mismatch negativity (MMN), the N1, the N2b and the P3 for louder tones than for quieter tones (Näätänen, Paavilainen, Alho, Reinikainen, & Sams, 1989; Näätänen, Paavilainen, Rinne, & Alho, 2007; Rinne, Särkkä, Degerman, Schröger, & Alho, 2006).

Even though, as we have just seen, the possible crossmodal correspondence between loudness and verticality has been researched, it has not been fully understood yet. Besides, all the existing evidence on loudness and verticality has been garnered using paradigms in which the response code contained linguistic spatial connotations. When judging loudness, participants may implicitly use the spatial labels "up" or "down". In the present study, in order to reduce any potential confound linked to the use of linguistic labels, we examined whether this correspondence still occurs when the response code is shifted towards an orthogonal feature that does not contain any spatial connotation (such as colour). In the present study, the possible interaction between loudness and verticality was addressed by means of a testing method including an orthogonal response code. This approach allowed us to investigate whether the crossmodal correspondence between loudness and verticality can occur in a scenario where the response codes contain non-spatial linguistic labels.

Continuing now with those studies that have explored the relationship between loudness and verticality, Eitan and Granot (2006) found that sounds with descending loudness tended to be associated with downwards imagined movements. However, sounds with ascending loudness seemed to be more linked to faster imagined movements than to upward imagined movements. Therefore, loudness can be associated, in people's imagery at least, with either spatial elevation or movement speed depending on its intensity levels. In another study by Eitan and collaborators (2008), the Garner paradigm, which investigates how two perceptual dimensions can interact and interfere with each other, was used to investigate whether the processing of dynamic changes in loudness and visual movement interact or not (see Garner & Sutliff, 1974). Participants were presented with sounds that could either increase or decrease in loudness and visual stimuli that could move either upwards or downwards. The task consisted of judging changes in either loudness or in vertical visual motion. The results of this study revealed that the participants' speeded responses at judging changes in loudness were modulated by the vertical movement of the visual stimuli. However, participants' responses to changes in vertical movements of the visual stimuli were not affected by the differences in loudness.

In another study by Fernandez-Prieto and collaborators (Fernandez-Prieto, Spence, Pons, & Navarra, 2017), a crossmodal effect between loudness and verticality was found both in bilingual Spanish/Catalan speakers and monolingual English speakers. Participants had to judge whether a sound was louder or quieter than a probe sound. They responded by means of a vertically-placed keyboard, with a key located either above or below a central starting point. Participants performed faster when the response button coincided spatially with the sound (loud-up, quiet-down) than when it did not (loud-down, quiet-up). The same pattern of results was found in a study conducted by Bruzzi and collaborators (2017), in which participants were faster responding to loud and quiet sounds with crossmodally congruent vertical buttons. However, this previous research did not take into consideration the perceptual differences that have consistently been observed between loud and quiet sounds. As loud sounds have a larger impact than quiet sounds both on behaviour (Neuhoff, 1998; Schröger, 1996) and on neural activity (Näätänen et al., 1989, 2007; Rinne et al., 2006), it is plausible that a difference between louder and quieter sounds will be observed in terms of their ability to elicit vertical representations. Along these lines, in a previous study conducted by Fernández-Prieto and Navarra (2017), participants were faster at detecting a visual target in the upper visual field when this stimulus was preceded by a sound with rising pitch than when it was preceded by a sound with decreasing pitch. Interestingly, this congruency effect was not observed for sounds with falling pitch, suggesting the presence of a possible link between arousal (or alertness) and the vertical representation associated to pitch: the higher the pitch, the larger the brain response to the sound and its capability of generating the impression of "verticality".

All in all, more research is needed to understand the possible crossmodal correspondence between loudness and verticality, in which potential confounds derived from the use of linguistic labels are avoided. In the present study, we used a classification task in which participants learnt an association between a coloured visual stimulus and an either loud or quiet sound. Once the association was learnt, participants were presented with the two-coloured stimuli (one above and one below the fixation point), and only one of the two sounds, having to identify which of the two-coloured stimuli was associated to the sound based on what they just learnt. Participants responded moving a computer mouse to reach the visual stimulus with the cursor. Based on the existing literature (Bruzzi, Talamini, Priftis, & Grassi, 2017; Eitan et al., 2008; Fernández-Prieto & Navarra, 2017; Spence, 2011), we would expect participants to be faster in the crossmodally congruent pairs of stimuli (target appearing upwards together with a loud sound, and target appearing downwards together with a quiet sound).

METHOD

Participants. Twenty-nine healthy participants from a volunteer database at Hospital Sant Joan de Déu took part in the experiment (19 female, mean age 21 ± 3.98 years old). They had normal or corrected-to-normal vision, were right handed, had normal hearing and reported to be non-musicians. They received 5 euros for participating in the study. Written informed consent was obtained from all participants before taking part in the experiment. The study was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and had ethical approval from Hospital Sant Joan de Déu and Parc Sanitari Sant Joan de Déu Ethics Committee.

Apparatus. An Intel Core® computer with a 15-inch CRT monitor (Hyundai Q770, South Korea; refresh rate = 75 Hz, screen resolution = 800*600 pixels) and two loudspeakers (Altec Lansing V52420, China) were used for the experiment. Experiments were run using E-Prime 2.0 (Psychology Software Tools Inc., Pittsburg, PA) in a dimly lit and partially soundproof booth. Participants sat at a distance of approximately 60 cm from the screen.

Stimuli. The visual stimuli consisted of 2 rectangles (1.67*2.27 degrees of visual angle) that had an either blue (RGB 0,255,255) or yellow (RGB 255,255,0) frame, each one containing a smaller square filled with the same colour (0.57°*0.57°; see Figure 1). The brightness levels of the two coloured stimuli were considered (blue: 78.7% luminance relative to white; yellow: 92.8% of luminance with respect to white). These stimuli were presented on the computer screen on a black background (RGB 0,0,0,). The fixation point used in the test phase (see Procedure) consisted of a rectangle with a white frame (0.5°*0.76°) containing a smaller white square (0.25°*0.25°). The two-coloured rectangles were placed at a distance of 2.86° above (upper rectangle) or below (lower rectangle) the fixation point. The two tones that were associated with the coloured blue and yellow rectangles (see Procedure) were identical in frequency (2000 Hz) but differed in terms of loudness ('loud' tone: 82 dB; 'quiet' tone: 56 dB). The auditory stimuli lasted for 200 ms and were presented on the 2 loudspeakers placed at each side of the screen. Loudness levels were tested using a digital sound level meter (Sinometer JTS-1357, China).

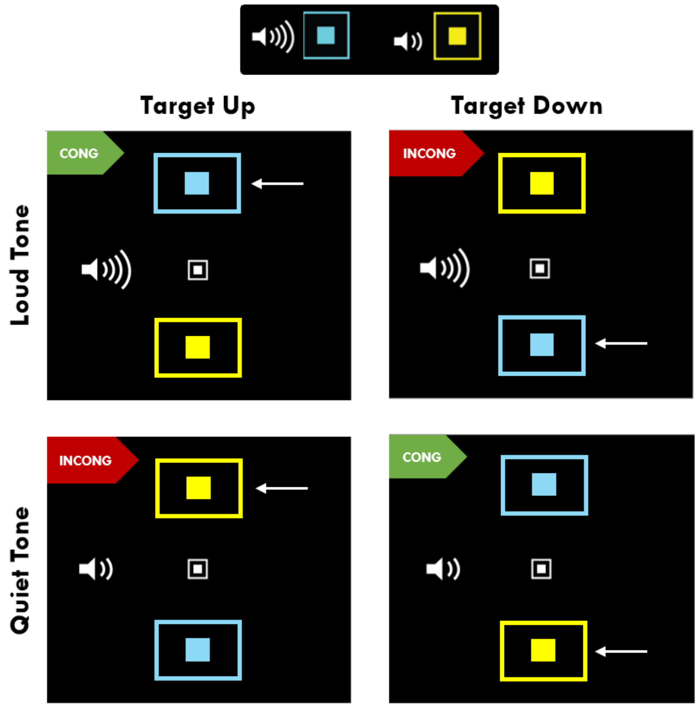

Figure 1. Stimuli and design. A coloured rectangle was associated with either a loud or a quiet sound (in the example, the blue rectangle is associated with the loud sound and the yellow rectangle with the quiet sound). The rectangles appeared above or below the fixation point, and the combination of the two positions and the two sounds generated congruent (loud sound associated with rectangle above and quiet sound associated with rectangle below) and incongruent conditions (loud sound associated with rectangle below and quiet sound associated with rectangle above).

Procedure. <u>Association learning phase</u>. Participants were instructed to associate each of the 2 colours of the rectangles (blue or yellow) with a specific tone (loud or quiet in terms of loudness). Half of the participants learnt to associate yellow with the loud tone and blue with the quiet tone, while the other half learnt the opposite association (see Figure 1). The two colour-loudness associations were learnt by participants in two steps. In the first step, participants were passively exposed to one of the two possible audiovisual associations mentioned, which they were asked to learn. That is, a yellow or a blue rectangle was displayed, in each trial, in the centre of the screen, together with the loud or quiet sound. Participants were presented with a total number of 20 trials, with a random order, that lasted 1200 ms each (1000 ms for gaze fixation plus 200 ms for the simultaneous presentation of the tone and the visual coloured target). This first step lasted 1 min in total. In the second step, participants were presented with either a blue or a yellow rectangle in the centre of the screen, which appeared concurrently with a loud or quiet sound, and were asked to decide whether there was an association between the two stimuli or not, according to the two associations previously learnt. The number of correct and incorrect audiovisual pairings was equal. The order of presentation of correct and incorrect trials was randomised. Feedback was provided after each response: the messages "correct" or "incorrect" were displayed in the screen after each response. The second part of the training lasted for 3 min and contained as many trials as the participant could answer in this interval.

Test phase**.** Right after participants completed the learning phase of the experiment, the test phase began. In this part, each test trial started with a fixation point that appeared at the centre of the screen. The participants were instructed to click, with the cursor, on the fixation point (being also the point of departure of the cursor). Then, after an interval that could vary randomly between 600 and 1000 ms, either the loud or the quiet tone was played. Simultaneously with the presentation of one of the two sounds, the two rectangles (blue and yellow) appeared on the computer screen, 5.53° above and below the fixation point. The position of the blue and yellow rectangles was randomised across trials. The participants' task consisted of producing a speeded movement with the computer mouse that would be translated into a vertical displacement of the cursor on the computer screen towards the rectangle that had previously been associated with the presented tone and clicking inside its area. An unlimited amount of time was given to participants for responding. Once the response was given, there was an inter-trial interval of 1000 ms (see Figure 2). Afterwards, the fixation point re-appeared in the centre of the screen and participants had to place the mouse back in the centre of the fixation point and click for the next trial to start. The experimental test phase included 240 trials. There were 60 trials for each loudness-position combination: high vertical position combined with the loud tone, low vertical position with a quiet tone, high vertical position with a quiet tone, and a low vertical position with loud tone. Note that the first two of these combinations were *congruent* in terms of the crossmodal associations previously described in the literature (see Spence, 2011), and the latter 2 combinations were *incongruent*. To familiarise with the task, participants underwent a 3min training including approximately 170 trials that were identical to those in the main test. The whole experimental session took 20 min approximately.

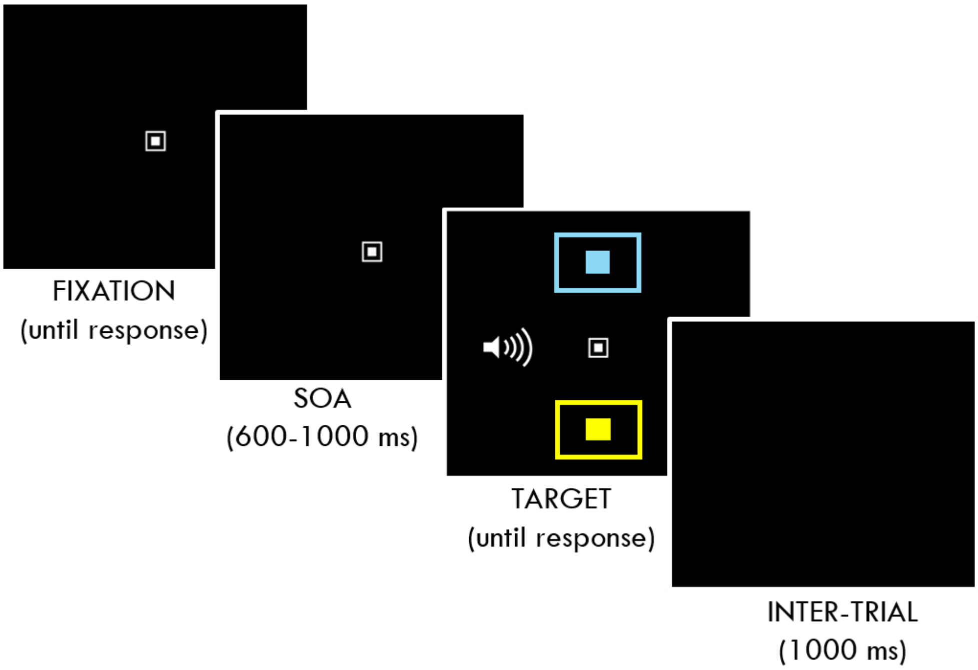

Figure 2. Experiment set-up. Participants were presented with the visual stimuli (a blue and a yellow rectangle) on the screen. There was a fixation and departure point between these two stimuli. Simultaneously with the coloured rectangles, one of 2 tones with different intensity was played. Participants moved the cursor, as fast as possible, towards the rectangle that corresponded to the tone, according to the previously-learnt association during the learning phase.

RESULTS

Association learning phase. Participants answered to an average of 70.93 trials in this phase (SD: 4.12), with a mean accuracy of 88.59% (SD: 5.85).

Test phase. Participants who did not fulfil a criterion of at least an accuracy of 60% were discarded. Data from 2 participants were removed from the analyses due to low performance rates (4% and 25% of correct responses). The average performance rate of the remaining participants (n=27) was 87.3% (standard deviation [SD]=11.4). Incorrect trials and those whose response times (RTs) were below 200 ms were discarded from the final analyses. The latter criterion was applied based on previous studies including similar tasks, using a mouse cursor to reach small targets (Phillips & Triggs, 2001). Likewise, times longer than 2000 ms were also considered as invalid as participants were not performing the speeded task correctly in that particular trial (i.e., due to distraction or other factors). An outlier correction was also applied, rejecting trials with RTs above or below 2.5 SD from each participant's individual mean. The number of outlier trials was very low across participants (mean=6, maximum number of trials discarded: 13 out of 240). Accuracy was calculated based on the area of the rectangle that participants had to click on and responses within these perimeters (see Stimuli) were considered valid. Additionally, responses were categorised as correct or incorrect based on the position of the visual target that was correct, for every trial, according to the previously-learned association.

To discard whether luminance levels influenced participants' responses (due to an association between loudness and brightness), we checked if there were differences between the two groups of participants that learnt the different associations (yellow-loud vs. blue-loud). Even though participants that learnt the loud-blue association performed overall slightly faster (mean: 752.59 ms; SD: 144.15) than participants that learnt the loud-yellow association (mean: 821.44 ms, SD: 144.95), we found no significant differences in reaction times between the two groups of participants ($t(25)=-1.229$, $p=.230$). In this sense, no effect of brightness and loudness was found in our experiment. Regarding the congruence analysis, a repeated-measures ANOVA was conducted for RTs, with *congruence* (congruent vs. incongruent) and *loudness* (loud vs. quiet) as within-subject variables. The results revealed a main effect of congruence ($F(1,26)=9.200$, $p=.005$, $\eta_p^2=.261$), reflecting faster RTs when responding to congruent than to incongruent visual targets (see Figure 3). There was also a main effect of loudness ($F(1,26)=12.187$, $p=.002$, $\eta_p^2=.319$): participants' responses were faster for the loud sounds than for the quiet sounds. Additionally, a significant interaction was found between congruence and loudness ($F(1,26)=11.845$, $p=.002$, $\eta_p^2=.313$). Subsequent Bonferroni post-hoc comparisons revealed that participants were significantly faster at responding to the loud tone when the visual target location was congruent (see Figure 3), that is, when the rectangle was presented on the upper position, than when it was incongruent (i.e., presented on the lower position; $t(26)=-4.996$, $p<.001$). Participants' responses for the loud congruent condition were also significantly faster than responses both in the quiet congruent ($t(26)=-5.043$, $p<.001$) and the quiet incongruent ($t(26=-4.697$, $p<.001$) conditions. No congruence effect was observed for the quiet tone, and no differences were observed between the quiet congruent and the quiet incongruent conditions ($t(26)=1.701$, $p=.101$).

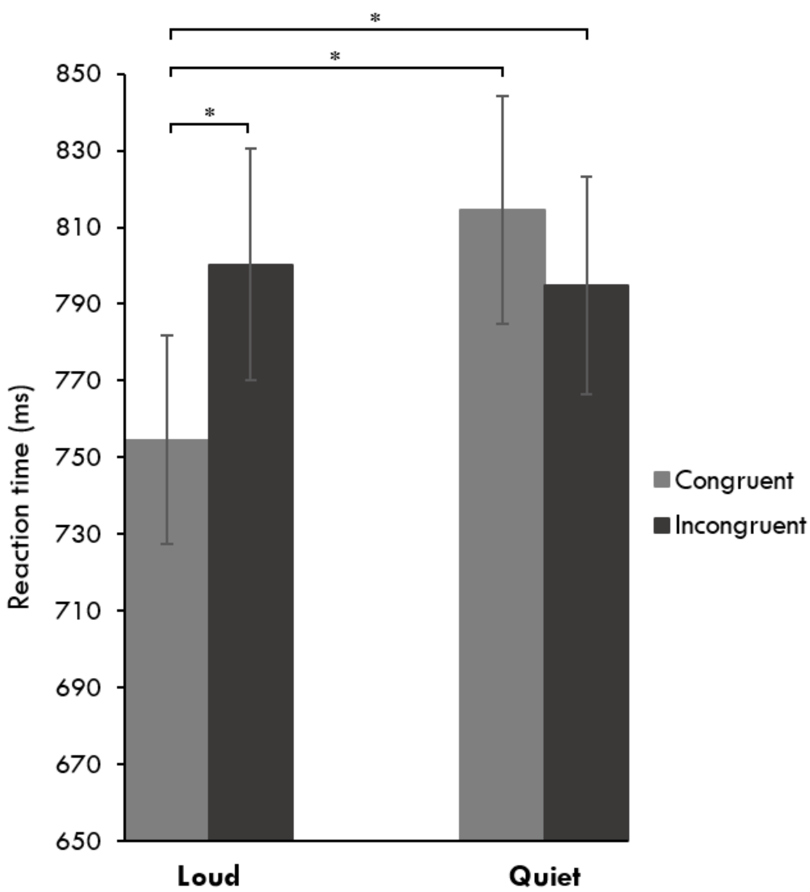

Figure 3. Results. Participants' responses were faster to the loud tone when this was crossmodally congruent with the position of the coloured rectangle than to the incongruent condition. Their responses were also faster in the loud congruent condition than in both quiet congruent and quiet incongruent conditions. However, the crossmodal effect was not observed for quiet sounds: RTs were not significantly different in the congruent or the incongruent conditions. Asterisks indicate a significance level <.001. The error bars plotted account for the standard error of the mean (SEM).

DISCUSSION

Literature on the field of crossmodal correspondences has tended to assume that these are similar for all of the possible perceptual combinations when congruent conditions are met (e.g., high and low pitch combined with upper and low positions in space, respectively; see Spence, 2011). However, and considering previously observed differences between different kinds of stimuli (e.g., behavioural and neural differences between loud and quiet sounds, see Näätänen et al., 1989, 2007; Neuhoff, 1998; Rinne et al., 2006; Schröger, 1996), we suspect that this may not always be the case. In the present study, we explored whether the possible crossmodal correspondence between loudness and verticality is influenced by basic differences in processing between loud and quite sounds. This would be reflected in different response times to loudness along the vertical plane. We also tested whether this crossmodal correspondence was present both for loud and quiet sounds separately, and if it had the same effects in both conditions. We used a task in which participants did not rely on any vertical linguistic label to execute the response (e.g., classifying sounds as being "high" or "low"), but instead responded according to an orthogonal, non-spatial feature (that is, colour). In this way, participants were instructed to respond based on the arbitrary relationship between loudness and colour, without the explicit implication of a linguistic label directly related to verticality. This allowed us to avoid a possible linguistic confound that could account for the association between loudness and verticality (see Fernandez-Prieto et al., 2017).

Our results are in line with previous findings that suggest the existence of a perceptual (or representational) link between loudness and spatial verticality (Bruzzi et al., 2017; Eitan & Granot, 2006; Eitan et al., 2008; Fernandez-Prieto et al., 2017). Specifically, the pattern of results found in the present study also implies that this link is stronger (or even limited to) for the association between loud sounds and high spatial positions than for the association between quiet sounds and low positions. Congruency effects were found, in our experiment, for loud (but not for quiet) sounds. Along these lines, in a previous study conducted by Fernández-Prieto and Navarra (2017), participants were faster at detecting a visual target in the upper visual field when this stimulus was preceded by a sound with rising pitch than when it was preceded by a sound with decreasing pitch. Interestingly, this congruency effect was not observed for sounds with falling pitch, suggesting the presence of a possible link between arousal (or alertness) and the vertical representation associated to pitch: the higher the pitch, the larger the brain response to the sound and its capability of generating the impression of "verticality". This may suggest that the larger behavioural (Schröger, 1996) and neural response (Näätänen et al., 1989) elicited by louder sounds, that are perhaps more salient, interact with the way loudness and spatial interact crossmodally.

Thanks to our methodological design where possible crossmodal effects were studied by means of an indirect task that was completely orthogonal with respect to both loudness and spatial elevation, we found that the crossmodal correspondence between loudness and verticality takes place even when participants are not using a response code implying loudness, spatial elevation, or spatial language labelling. We believe this manipulation is key to advance our knowledge on the crossmodal correspondence between loudness and verticality, as previous studies might have included spatial linguistic labels in the response codes (Bruzzi et al., 2017; Eitan et al., 2008; Fernandez-Prieto et al., 2017; Lidji et al., 2007; Rusconi et al., 2006). In the present study, participants were making a judgment on the basis of a previously-learnt association between colour and loudness, but not based on the association between loudness and verticality, or loudness per se. A similar approach has been used in previous studies that accounted for other types of crossmodal correspondences, such as between pitch and spatial elevation (Bernstein & Edelstein, 1971; Melara & O'Brien, 1987), as well as in studies that looked for the correspondence between pitch and size (Gallace & Spence, 2006; Marks, 1987). Considering that the latter two perceptual correspondences are thought to be very robust, it is plausible that the use of a testing method with orthogonal response codes when studying these phenomena can easily, and reliably, reveal the presence of these associations. However, in the case of loudness and verticality, this is of special relevance because this correspondence had not been studied at such a profound level (e.g., using methodological approaches that reduce the use of spatial linguistic labels to solve the task).

The results obtained in the present study resemble previous evidence on the relation between pitch and spatial elevation (see Deroy et al., 2018; Spence, 2011; Spence & Deroy, 2013) where faster responses have consistently been observed in congruent conditions (i.e., for upward or downward responses to a high- and low-pitched sounds, respectively). As opposed to some previous studies, in which congruency effects are studied as a whole; i.e., collapsing upward and downward conditions (Bruzzi et al., 2017; Eitan et al., 2008; Occelli, Spence, & Zampini, 2009; Rusconi et al., 2006), the present study addressed the crossmodal correspondence for loud and quiet tones separately. This allowed us to look for possible asymmetries in the perceptual correspondence between loud and quiet sounds and high and low positions in space. In this regard, a similar pattern of results was found in a previous study addressing the relationship between verticality and pitch: frequency sweeps with ascending pitch were associated with higher positions in space, but no effect was found for descending frequency sweeps (Fernández-Prieto & Navarra, 2017). These results were interpreted by the authors as suggesting the presence of a boost of crossmodal effects related to the higher neural and behavioural response produced by high pitch.

Following these lines and consistently with previous literature, a plausible explanation of the asymmetry of crossmodal correspondences between loud and quiet sounds is that such correspondences are mediated by more basic or survival-related brain mechanisms (e.g., alertness) that are triggered by auditory stimuli. Increases in loudness have been previously related to faster reaction times (Bach et al., 2008; Schröger, 1996), as well as to enhanced EEG-related signals (Jacobsen, Horenkamp, & Schröger, 2003; Näätänen et al., 1989, 2007; Rinne et al., 2006; Schröger, 1996). We could argue, therefore, that the higher arousing and neural response of loud sounds can be influencing how loudness is represented in the vertical plane. The fact that auditory stimuli with increasing loudness are capable of generating larger behavioural responses, in non-human primates, than stimuli with decreasing loudness (Ghazanfar, Neuhoff, & Logothetis, 2002) leads to the idea that this bias can have an adaptive function (Hall & Moore, 2003; Neuhoff, 1998). In fact, increases (but not decreases) in loudness have been reported to activate subcortical, alertness-related brain areas such as the amygdala (Bach et al., 2008). A combined interpretation of our results and the aforementioned previous literature may perhaps suggest that the greater behavioural and neural response to loud sounds may help the crossmodal effect between loudness and verticality to emerge.

In summary, we found evidence that goal-directed responses specifically towards an upper spatial position can be modulated by acoustic loudness. This association seems to be present even when the participants have to respond to non-spatial linguistic features, therefore using a colour-based response code. We hypothesise that this crossmodal association may be based on the presence of a common representation code for verticality and loudness. At the same time, our data suggests that the difference in the neural response between loud and quiet sounds (see Näätänen et al., 1989; Neuhoff, 1998) can have an impact on the vertical representation of loudness. Further studies addressing the interaction between loudness, verticality and the influence of arousal and alertness using different testing methods will contribute to better understand this phenomenon.

RESUMEN

Aunque la asociación perceptiva entre la verticalidad y la frecuencia auditiva ha sido ampliamente estudiada, la relación entre la intensidad y la verticalidad sigue sin entenderse completamente. Mientras que se asume que los sonidos más y menos intensos están asociados de forma igual con la elevación espacial, existen diferencias perceptivas entre los dos tipos de sonidos que sugieren lo contrario. Por ejemplo, los sonidos más intensos tienden a generar más actividad, tanto en el aspecto conductual como neuronal, que los sonidos más flojos. En este estudio, investigamos si esta diferencia influye en la correspondencia transmodal entre la intensidad y la

verticalidad. En una fase inicial, los participantes aprendieron asociaciones arbitrarias entre uno de dos tonos que diferían en intensidad (82dB vs. 56 dB) y uno de dos rectángulos coloreados (azul vs. amarillo). Durante la fase experimental, se les presentaron los dos estímulos coloreados (cada uno de ellos localizado por encima o debajo de un punto central de partida), junto con uno de los dos tonos. Los participantes tenían que indicar cuál de los dos rectángulos coloreados correspondía al tono previamente asociado moviendo el cursor del ratón desde el punto de partida hasta el objetivo. Los resultados mostraron que los participantes fueron significativamente más rápidos cuando respondían al tono más intenso cuando el objetivo visual se situaba arriba (condición congruente) que cuando se situaba abajo (condición incongruente). Para los sonidos menos intensos no se observaron diferencias entre las condiciones congruente (flojo-abajo) e incongruente (flojo-arriba). En general, este patrón de resultados sugiere que las posibles diferencias en la actividad neuronal generadas por sonidos de mayor y menor intensidad influyen el grado en el que la intensidad y la elevación espacial comparten contenido representacional.

REFERENCES

Bach, D. R., Schahinger, H., Neuhoff, J. G., Esposito, F., Salle, F. Di, Lehmann, C., … Seifritz, E. (2008). Rising Sound Intensity: An Intrinsic Warning Cue Activating the Amygdala. *Cerebral Cortex*, *18*(1), 145–150. https://doi.org/10.1093/cercor/bhm040

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, *87*(2), 241–247. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5542226

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012a). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage*, *59*(1), 663–672. https://doi.org/10.1016/j.neuroimage.2011.06.095

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012b). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage*, *59*(1), 663–672. https://doi.org/10.1016/J.NEUROIMAGE.2011.06.095

Bond, B., & Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception & Psychophysics*, *6*(6), 337–339. https://doi.org/10.3758/BF03212787

Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition*, *54*(2), 103–109. https://doi.org/10.1016/S0278-2626(03)00264-1

Bruzzi, E., Talamini, F., Priftis, K., & Grassi, M. (2017). A SMARC Effect for Loudness. *I-Perception*, *8*(6), 204166951774217. https://doi.org/10.1177/2041669517742175

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1525), 1831–1840. https://doi.org/10.1098/rstb.2009.0028

Burro, R., & Grassi, M. (2001). Experiments on size and height of falling objects. *Phenomenology of Sound Events, IST Project No. IST-2000-25287, Report 1*, 31–39.

Cabrera, D., & Tilley, S. (2003a). Parameters for auditory display of height and size. In *Proceedings of the 9th International Conference on Auditory Display (ICAD)*. Boston, MA: Georgia Institute of Technology.

Cabrera, D., & Tilley, S. (2003b). Vertical Localization and Image Size Effects in Loudspeaker Reproduction. In *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*. Banff, Canada: Audio Engineering Society.

Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of Object Length by Sound. *Psychological Science*, *9*(3), 211–214. https://doi.org/10.1111/1467-9280.00040

Critchley, M. (1953). *The parietal lobes*. Oxford, England: Williams and Wilkins.

Deroy, O., Fernandez-Prieto, I., Navarra, J., & Spence, C. (2018). Unraveling the Paradox of Spatial Pitch. In *Spatial Biases in Perception and Cognition* (pp. 77–93). Cambridge University Press. https://doi.org/10.1017/9781316651247.006

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic Infants Are Sensitive to Space-Pitch Associations Found Across Cultures. *Psychological Science*, *25*(6), 1256–1261. https://doi.org/10.1177/0956797614528521

Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nature Neuroscience*, *10*(7), 915–921. https://doi.org/10.1038/nn1925

Dubus, G., & Bresin, R. (2013). A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PLoS ONE*, *8*(12), e82491. https://doi.org/10.1371/journal.pone.0082491

Eitan, Z., & Granot, R. Y. (2006). How Music Moves. *Music Perception: An Interdisciplinary Journal*, *23*(3). https://doi.org/10.1525/mp.2006.23.3.221

Eitan, Z., Schupak, A., & Marks, L. E. (2008). Louder is Higher: Cross-Modal Interaction of Loudness Change and Vertical Motion in Speeded Classification. In *Proceedings of the 10th international conference on music perception and cognition (ICMP10)*.

Evans, K. K., & Treisman, A. (2011). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6–6. https://doi.org/10.1167/10.1.6

Fernández-Prieto, I., & Navarra, J. (2017). The higher the pitch the larger its crossmodal influence on visuospatial processing. *Psychology of Music*, *45*(5), 713–724. https://doi.org/10.1177/0305735616684205

Fernández-Prieto, I., Navarra, J., & Pons, F. (2015). How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behavior and Development*, *38*, 77–81. https://doi.org/10.1016/j.infbeh.2014.12.008

Fernandez-Prieto, I., Spence, C., Pons, F., & Navarra, J. (2017). Does Language Influence the Vertical Representation of Auditory Pitch and Loudness? *I-Perception*, *8*(3). https://doi.org/10.1177/2041669517716183

Ferri, F., Tajadura-Jiménez, A., Väljamäe, A., Vastano, R., & Costantini, M. (2015). Emotion-inducing approaching sounds shape the boundaries of multisensory peripersonal space. *Neuropsychologia*, *70*, 468–475. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2015.03.001

Foster, N. E. V., Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *NeuroImage*, *75*, 27–35. https://doi.org/10.1016/j.neuroimage.2013.02.044

Foster, N. E. V., & Zatorre, R. J. (2010a). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(6), 1350–1359. https://doi.org/10.1093/cercor/bhp199

Foster, N. E. V., & Zatorre, R. J. (2010b). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage*, *53*(1), 26–36. https://doi.org/10.1016/j.neuroimage.2010.06.042

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*(7), 1191–1203. https://doi.org/10.3758/bf03193720

Garner, W. R., & Sutliff, D. (1974). The effect of goodness onencoding time in visual pattern discrimination. *Perception & Psychophysics*, *16*(3), 426–430. https://doi.org/10.3758/BF03198567

Ghazanfar, A. A., Neuhoff, J. G., & Logothetis, N. K. (2002). Auditory looming perception in rhesus monkeys. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15755–15757. https://doi.org/10.1073/pnas.242469699

Hall, D. A., & Moore, D. R. (2003). Auditory Neuroscience: The Salience of Looming Sounds. *Current Biology*, *13*(3), R91–R93. https://doi.org/10.1016/S0960-9822(03)00034-4

Jacobsen, T., Horenkamp, T., & Schröger, E. (2003). Preattentive memory-based comparison of sound intensity. *Audiology & Neuro-Otology*, *8*(6), 338–346. https://doi.org/73518

Lewkowicz, D. J., & Minar, N. J. (2014). Infants Are Not Sensitive to Synesthetic Cross-Modality Correspondences: A Comment on Walker et al. (2010). *Psychological Science*, *25*(3), 832–834. https://doi.org/10.1177/0956797613516011

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Developmental Psychology*, *16*(6), 597–607. https://doi.org/10.1037/0012-1649.16.6.597

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1189–1207. https://doi.org/10.1037/0096-1523.33.5.1189

Lipscomb, S. D., & Kim, E. M. (2004). Perceived match between visual parameters and auditory correlates: an experimental multimedia investigation. In S. Lipscomb, R. Ashley, R. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception & Cognition (ICMPC8)*. Evanston, IL, USA: Adelaide, Australia: Casual Productions.

Marks, L. E. (1974). On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness. *The American Journal of Psychology*, *87*(1/2), 173. https://doi.org/10.2307/1422011

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384–394. https://doi.org/10.1037/0096-1523.13.3.384

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, *116*(4), 323–336. https://doi.org/10.1037/0096-3445.116.4.323

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1989). Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neuroscience Letters*, *98*(2), 217–221. https://doi.org/10.1016/0304-3940(89)90513-2

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, *395*(6698), 123–124. https://doi.org/10.1038/25862

Occelli, V., Spence, C., & Zampini, M. (2009). Compatibility effects between sound frequency and tactile elevation. *NeuroReport*, *20*(8), 793–797. https://doi.org/10.1097/WNR.0b013e32832b8069

Parise, C. V., & Spence, C. (2009). 'When Birds of a Feather Flock Together': Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes. *PLoS ONE*, *4*(5), e5664. https://doi.org/10.1371/journal.pone.0005664

Parise, C. V. (2016). Crossmodal Correspondences : Standing Issues and Experimental Guidelines. *Multisensory Research*, *29*(1–3), 7–28. https://doi.org/10.1163/22134808-00002502

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between Auditory Pitch and Visual Elevation Do Not Depend on Language: Evidence from a Remote Population. *Http://Dx.Doi.Org/10.1068/P7225*. https://doi.org/10.1068/P7225

Phillips, J. G., & Triggs, T. J. (2001). Characteristics of cursor trajectories controlled by the computer mouse. *Ergonomics*, *44*(5), 527–536. https://doi.org/10.1080/00140130121560

Rinne, T., Särkkä, A., Degerman, A., Schröger, E., & Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research*, *1077*(1), 135–143. https://doi.org/10.1016/j.brainres.2006.01.043

Root, R. T., & Ross, S. (1965). Further Validation of Subjective Scales for Loudness and Brightness by Means of Cross-Modality Matching. *The American Journal of Psychology*, *78*(2), 285. https://doi.org/10.2307/1420502

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*(2), 113–129. https://doi.org/10.1016/j.cognition.2005.01.004

Schröger, E. (1996). The influence of stimulus intensity and inter-stimulus interval on the detection of pitch and loudness changes. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *100*(6), 517–526. https://doi.org/10.1016/S0168-5597(96)95576-8

Sluming, V., Brooks, J., Howard, M., Downes, J. J., & Roberts, N. (2007). Broca's Area Supports Enhanced Visuospatial Cognition in Orchestral Musicians. *Journal of Neuroscience*, *27*(14), 3799–3806. https://doi.org/10.1523/JNEUROSCI.0147-07.2007

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, *24*(1), 99–142. https://doi.org/10.1016/0010-0285(92)90004-L

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*, *22*(1), 245–260. https://doi.org/10.1016/j.concog.2012.12.006

Stevens, J., & Marks, L. (1965). Cross-modality matching of brightness and loudness. In *Proceedings of the National Academy of Sciences* (Vol. 54.2, pp. 407–411).

Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., & Västfjäll, D. (2010). Embodied auditory perception: The emotional impact of approaching and receding sound sources. *Emotion*, *10*(2), 216–229. https://doi.org/10.1037/a0018422

Tillmann, B., Jolicœur, P., Ishihara, M., Gosselin, N., Bertrand, O., Rossetti, Y., & Peretz, I. (2010). The Amusic Brain: Lost in Music, but Not in Space. *PLoS ONE*, *5*(4), e10173. https://doi.org/10.1371/journal.pone.0010173

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences. *Psychological Science*, *21*(1), 21–25. https://doi.org/10.1177/0956797609354734

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2014). Preverbal Infants Are Sensitive to Cross-Sensory Correspondences: Much Ado About the Null Results of Lewkowicz and Minar (2014). *Psychological Science*, *25*(3), 835–836. https://doi.org/10.1177/0956797613520170

Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Perception & Psychophysics*, *42*(5), 491–502. https://doi.org/10.3758/BF03209757

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002

Walsh, V., Gallistell, R. C., Gellman, R., Brannon, E. M., Roitman, J. D., Rossetti, Y., … Fischer, M. H. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. https://doi.org/10.1016/J.TICS.2003.09.002

Wicker, F. W. (1968). Mapping the Intersensory Regions of Perceptual Space. *The American Journal of Psychology*, *81*(2), 178. https://doi.org/10.2307/1421262

Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory Distance Perception in Humans: A Summary of Past and Present Research. *Acta Acustica United with Acustica*, *91*(3), 409–420.

(Manuscript received: 20 September 2018; accepted: 12 June 2019)