



The Spatial Musical Association of Response Codes does not depend on a normal visual experience: A study with early blind individuals

Zaira Cattaneo^{1,2} · Carlotta Lega¹ · Luca Rinaldi¹ · Micaela Fantino³ · Chiara Ferrari¹ · Lotfi B. Merabet⁴ · Tomaso Vecchi^{2,3}

Published online: 26 February 2018
© The Psychonomic Society, Inc. 2018

Abstract

Converging evidence suggests that the perception of auditory pitch exhibits a characteristic spatial organization. This pitch–space association can be demonstrated experimentally by the Spatial Musical Association of Response Codes (SMARC) effect. This is characterized by faster response times when a low-positioned key is pressed in response to a low-pitched tone, and a high-positioned key is pressed in response to a high-pitched tone. To investigate whether the development of this pitch–space association is mediated by normal visual experience, we tested a group of early blind individuals on a task that required them to discriminate the timbre of different instrument sounds with varying pitch. Results revealed a comparable pattern in the SMARC effect in both blind participants and sighted controls, suggesting that the lack of prior visual experience does not prevent the development of an association between pitch height and vertical space.

Keywords Spatial cognition · Audition · Multisensory processing

Converging evidence suggests that the perception of auditory pitch exhibits a characteristic spatial organization. A compelling demonstration of this resides in the Spatial Musical Association of Response Codes (SMARC) effect, referring to the observation that response times to low-pitched tones are faster when keys are located in the lower part of the space and responses to high-pitched tones are faster when keys are located in the upper part of the space (Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). This is also the case when pitch is irrelevant for the task being considered (as, for instance, when the

task is about discriminating between musical instruments; Lidji et al., 2007; Rusconi et al., 2006), leading researchers to suggest that the pitch–space association may happen in an automatic (i.e., without requiring attentional control) fashion (Lidji et al., 2007). Hence, high tones seem to be mentally linked to upper locations and low tones to lower locations in space. Pitch can also be mapped along a left-to-right continuum, but such horizontal mapping is weaker than vertical mapping, being more influenced by one’s musical and instrumental playing experience (Lidji et al., 2007; Rusconi et al., 2006; Timmers & Li, 2016). Interestingly, listening to sounds of a different pitch may also affect orienting of spatial attention in peripersonal space (e.g., Akiva-Kabiri, Linkovski, Gertner, & Henik, 2014; Bernardi et al., 2015; Bodak, Malhotra, Bernardi, Cocchini, & Stewart, 2014; Fernández-Prieto, Vera-Constán, García-Morera, & Navarra, 2012; Ishihara et al., 2013; Lega, Cattaneo, Merabet, Vecchi, & Cucchi, 2014).

The mapping between pitch height and vertical location is quite robust (Dolscheid, Shayan, Majid, & Casasanto, 2013; Roffler & Butler, 1968) and is already present in early infancy (Dolscheid, Hunnius, Casasanto, & Majid, 2014; Mondloch & Maurer, 2004; Walker et al., 2010). Although it may also rely on linguistic associations (Dolscheid et al., 2013), it has been reported in prelingual

✉ Zaira Cattaneo
zaira.cattaneo@unimib.it

¹ Department of Psychology, University of Milano-Bicocca, Milano, Italy

² Brain Connectivity Center, IRCCS Mondino Foundation, Pavia, Italy

³ Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy

⁴ The Laboratory for Visual Neuroplasticity, Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, USA

infants (Dolscheid et al., 2014; Walker et al., 2010) and in remote isolated populations who do not use linguistic spatial metaphors (i.e., “high” and “low”) to describe pitch (Parkinson, Kohler, Sievers, & Wheatley, 2012). Indeed, pitch–space mapping may be rooted in auditory scene statistics that reveal a clear association between frequency and elevation, with sounds coming from higher elevations tending to have higher frequencies than those coming from lower elevations (Parise, Knorre, & Ernst, 2014).

Sound sources are often visible, and the availability of visual information improves sound localization estimates (e.g., Tabry, Zatorre, & Voss, 2013). Indeed, when visual and auditory inputs providing spatial information are both available, the final multisensory estimate tends to be more precise than either unisensory estimate, with vision playing a major role in localization tasks (e.g., Alais & Burr, 2004; Shelton & Searle, 1980; see also Stein, Stanford, & Rowland, 2014). Visual experience may thus play a crucial role in associating the pitch of natural sounds to the locations they generally originate from, and therefore in internalizing audio-spatial correlations in the environment (Deroy, Fasiello, Hayward, & Auvray, 2016). In this regard, it is interesting that the first evidence about pitch–space correspondence comes from experiments manipulating visual stimuli. In one of the earliest investigations on this topic, Bernstein and Edelstein (1971) found that a visual stimulus was classified faster as high or low when it was accompanied by a tone that was congruent rather than incongruent (e.g., high pitch with high position rather than low). This result has been replicated in many subsequent studies (e.g., Ben-Artzi & Marks, 1995; Evans & Treisman, 2010; Melara & O’Brien, 1987; Patching & Quinlan, 2002), also investigating other pitch-visual feature correspondences (such as lightness, brightness, shape, size).

If pitch–space correspondence arises through frequent associations in everyday experience (where vision plays a critical role as the dominant modality we rely on in spatial localization; e.g., Alais & Burr, 2004; Shelton & Searle, 1980; Tabry et al., 2013), we could suspect that blind individuals who lack prior visual experience may not show the same learned perceptual association. Moreover, while the blind typically make superior use of spectral cues for localization in the horizontal plane, this compensatory behavior may come at the cost of a reduced ability to use these cues for localization in the vertical plane (Voss, Tabry, & Zatorre, 2015). This impaired ability to localize sounds in the vertical plane may also (indirectly) affect pitch-vertical space mapping in the blind.

Previous studies have demonstrated that the lack of prior visual experience may affect but not prevent the use of mental spatial representations (e.g., Afonso et al., 2010; Cattaneo et al., 2011; Cattaneo, Vecchi, Monegato,

Pece, & Cornoldi, 2007; Postma, Zuidhoek, Noordzij, & Kappers, 2007; Röder, Kusmirek, Spence, & Schicke, 2007). For instance, blind individuals are likely to represent numerical information spatially along a left-to-right-oriented mental number line, as sighted individuals typically do (e.g., Castronovo & Seron, 2007; Cattaneo et al., 2011; Cattaneo, Fantino, Tinti, Silvanto, & Vecchi, 2010; Rinaldi, Vecchi, Fantino, Merabet, & Cattaneo, 2015; Szűcs, & Csépe, 2005; but see Pasqualotto, Taya, & Proulx, 2014). Moreover, past and future events seem to be mapped on a left (past) to right (future) spatial continuum, regardless of visual experience (Bottini, Crepaldi, Casasanto, Crollen, & Collignon, 2015; Santiago, Lupiáñez, Pérez, & Funes, 2007); although blindness may affect the mapping of past and future on the sagittal plane (Rinaldi, Vecchi, Fantino, Merabet, & Cattaneo, 2017). Furthermore, in verbal memory tasks (in which sighted individuals tend to organize information spatially), blind individuals may use preferentially alternative nonspatial strategies (Bottini, Mattioni, & Collignon, 2016). Specifically for pitch processing, a recent study reported a preferred association between tones increasing in pitch and upward tactile movements and tones decreasing in pitch and downward tactile movements in sighted individuals, supporting a pitch–space mapping, whereas blind individuals did not show any preferential association (Deroy et al., 2016). These findings suggest that the normal development of vision may be critical in mediating this pitch–space correspondence (Deroy et al., 2016).

To shed further light on this issue, we carried out a study in which we presented a group of early blind and sighted individuals with an implicit (timbre judgment) SMARC-like task measuring the mapping of pitch height along a bottom-to-top vertical dimension. In accordance to existing literature, we expected sighted participants to be faster at categorizing the timbre of various musical instruments when they responded to low pitches using a response key located in lower space, and to high pitches using a response key located in upper space, despite the fact that pitch was irrelevant for the purposes of carrying out the task (e.g., Lidji et al., 2007; Rusconi et al., 2006). Importantly, if the mapping between pitch height and spatial location is mediated by normal visual development, responses of early blind individuals to high and low pitch tones should not be affected by the vertical location of the response key. This would suggest that prior visual experience is needed for learning the pitch–vertical space correspondence. Alternatively, if mapping of pitch in space depends more on auditory-motor experience, early blind individuals should also associate high and low pitches with different positions in vertical space, suggesting that the source of pitch–space mapping can be accounted for by mechanisms other than visual binding.

Method

Participants

Twenty-three right-handed early blind participants (14 males; mean age = 42.22 years, $SD = 13.00$, range: 18–65, mean education: 14.39 years, $SD = 3.87$) and 23 right-handed sighted participants (12 males; mean age = 38.61 years, $SD = 12.86$, range: 22–65, mean education: 15.83 years, $SD = 2.70$) took part in the experiment. The participants had little musical training. Specifically, blind participants had on average 2.60 years ($SD = 2.72$) and sighted participants 1.50 ($SD = 2.35$) years of musical training beyond basic music education obtained in primary school; with no significant difference between the two groups, $t(44) < 1$, $p = .31$. No participant had any history of neurological disorders or motor dysfunction. All sighted participants had normal (or corrected to normal) vision. All the blind participants were profoundly blind due to various ocular causes (see Table 1 for individual demographics), and all were proficient Braille readers.

Table 1 Characteristics of early blind participants

Gender	Age	Education	Cause of blindness	Light perception
M	48	18	Leber's congenital amaurosis	–
M	42	13	Optic nerve damage	+
M	28	13	Optic nerve damage	+
M	53	8	Optic nerve damage	+
F	20	13	Leber's congenital amaurosis	+
M	54	18	Retinopathy of prematurity	–
F	18	13	Congenital glaucoma	–
F	65	8	Optic nerve damage	+
F	33	18	Optic nerve damage	–
F	36	18	Retinal detachment	–
M	46	8	Optic nerve damage	+
M	39	18	Optic nerve damage	+
M	55	8	Optic nerve damage	+
F	28	18	Congenital glaucoma	+
F	40	18	Microphthalmia	–
F	27	16	Congenital glaucoma	–
M	58	16	Retinitis pigmentosa	–
M	60	16	Optic nerve damage	–
F	42	16	Optic nerve damage	+
M	57	16	Optic nerve damage	–
M	44	18	Optic nerve damage	–
M	45	8	Congenital glaucoma	–
M	33	15	Retinitis pigmentosa	–

Note. + = present; – = absent

Stimuli

The stimuli consisted of two low-pitched tones (C2 and E2; 65.41 Hz and 82.41 Hz, respectively) and two high-pitched tones (F5 and A5; 698.46 Hz and 880.00 Hz, respectively). These pitches were chosen in order to achieve the same semitone distance between the two low and the two high tones (i.e., four semitones between C2 and E2 and four semitones between F5 and A5; see Lidji et al., 2007, for study design similarities). Each tone was synthesized with piano and piano-keyboard timbre (belonging to the keyboard instrument family) and clarinet and saxophone timbre (belonging to the wind instrument family), for a total of 16 different stimuli. All stimuli were presented for a duration of 500 ms and were normalized in loudness at 0 dB (Audacity software, <http://audacity.sourceforge.net/>). Although all the sounds used were normalized at 0 dB, auditory stimuli may still differ in terms of perceived loudness. To rule out this possible confound, we conducted a preliminary experiment on 10 participants (not taking part in the main experiment) and confirmed that all sounds were perceived as equal in terms of loudness (for a similar procedure, see Rinaldi, Lega, Cattaneo, Girelli, & Bernardi, 2016).

Procedure

Participants were seated comfortably in a dimly lit room. Sighted participants were blindfolded throughout the entire experiment, as typically done in studies comparing blind and sighted performance on different perceptual tasks. This was done to avoid the potential effects of visual input (even if task irrelevant) on the performance of the sighted group (e.g., Tabry et al., 2013). Stimuli were binaurally delivered using professional headphones (Sennheiser HD 280 Pro headphone). Figure 1 shows the experimental setting and procedure. In each trial, the target auditory stimulus was presented for 500 ms. Participants were asked to judge whether the tone was played by a wind or a keyboard instrument (timbre judgment task) by pressing one of two vertically aligned response keys; one at the bottom (the space bar key) and one at the top (the key corresponding to the number 6) of a standard keyboard (see Lidji et al., 2007; Rusconi et al., 2006, for a similar procedure). There was no time limit for response, but task instructions emphasized both speed and accuracy. Even if participants could respond during the 500-ms stimulus presentation, the sound was always presented for its entire duration. After stimulus offset, a silent interval of 650 ms preceded the presentation of the next auditory stimulus. Participants took part in two experimental blocks, in which the association between instrument category (wind vs. keyboard) and response key

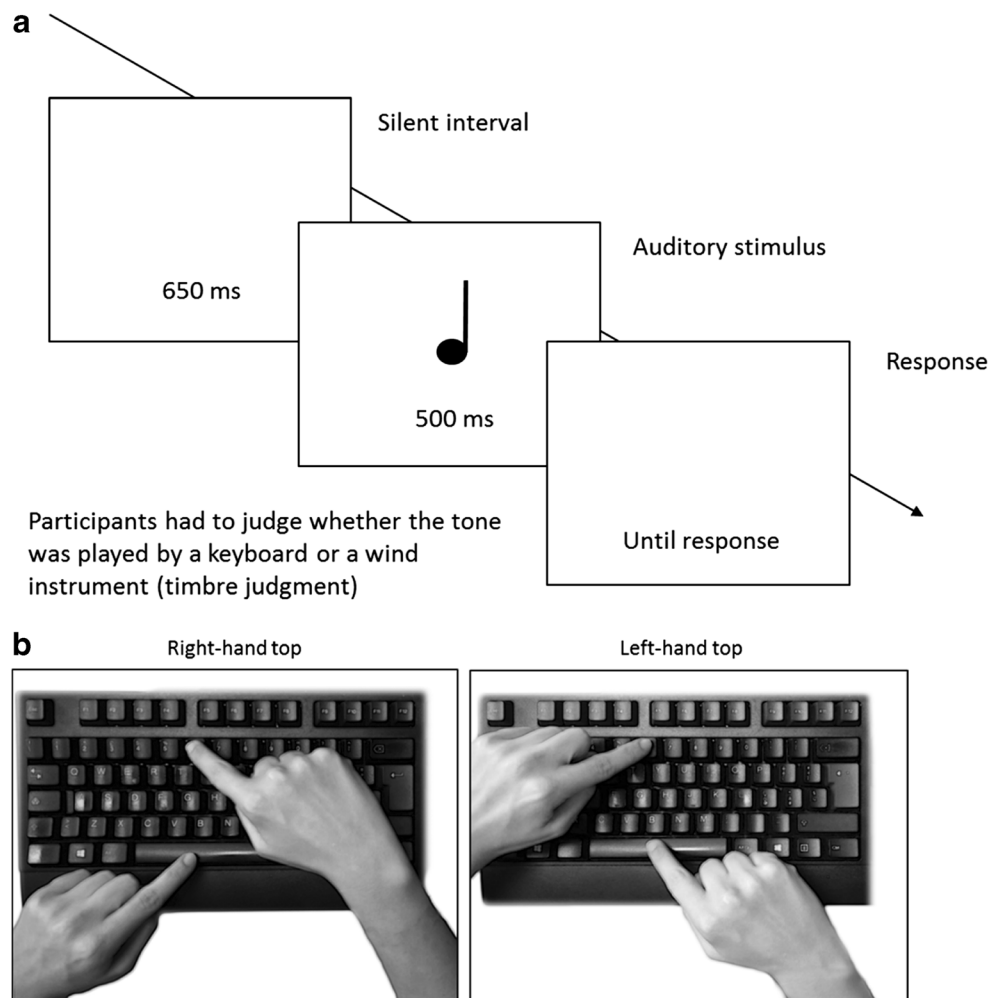


Fig. 1 **a** Experimental timeline. A silent interval of 650 ms preceded the presentation of the 500-ms auditory stimulus. In each trial, participants were asked to judge whether the tone was played by a keyboard instrument or by a wind instrument (timbre judgment task). **b** Hand

assignment. Participants performed two experimental blocks: in one block, they pressed the top key with their right hand and the bottom key with their left hand. In the other block hand position was reversed

was kept constant. However, in one block participants used their right hand to press the up response key and the left hand to press the bottom response key, while in the other block, the hand position was reversed (see Fig. 1b). Hand assignment to top/bottom response keys was manipulated in light of prior findings suggesting that hand position may affect the SMARC effect in the vertical plane (Lidji et al., 2007; but see Rusconi et al., 2006) and in light of prior evidence showing that blind individuals tend to rely more on body–hand-centered reference frames compared with sighted people (e.g., Cattaneo et al., 2008; Crollen, Dormal, Seron, Lepore, & Collignon, 2013; Noordzij, Zuidhoek, & Postma, 2006, 2007; Pasqualotto et al., 2014; Rinaldi et al., 2015; for recent evidence in auditory localization, see Vercillo, Tonelli, & Gori, 2018). The order of the two experimental blocks and the response-key–timbre assignment were counterbalanced across participants. Each of the 16

stimuli (i.e., four tones played by four different instruments; see above) was presented six times for a total of 96 trials in each block. Within each block, trials were presented in random order, with the only constraint being that the same identical tone (same pitch played by the same instrument) never occurred consecutively. Before the first experimental block, participants listened to two tones from each instrument used during the task in order to familiarize themselves with the auditory stimuli, and performed eight practice trials with tones of instruments not used in the real experiment (but still belonging to the keyboard or wind family). Participants did not train again before starting the second block, but were allowed a few minutes break. The software E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was used for stimuli presentation and data collection. The entire experiment lasted approximately 1 hour, including instructions, short breaks between the two blocks, and debriefing.

Table 2 Mean correct reaction times (ms) as a function of hand position (right-hand top, left-hand top) and pitch (high, low) in blind and sighted participants

Hand position		Blind		Sighted	
		High	Low	High	Low
Right-hand top	Bottom	845	794	968	854
	Top	794	879	894	937
	dRT (top–bottom)	–51	85	–74	83
Left-hand top	Bottom	930	810	913	821
	Top	838	907	877	893
	dRT (top–bottom)	–92	97	–36	72

Data analysis

We first carried out a repeated-measures analysis of variance (ANOVA) on mean correct reaction times (RT, recorded from onset of the auditory stimulus) and on mean error scores to look for possible group and block effects on performance. Hence, the SMARC effect was analyzed as in prior studies (Lidji et al., 2007; Weis, Estner, & Lachmann, 2016) computing the difference in mean correct reaction times (dRTs) and percentage of error rates (dErrors) between top-key and bottom-key responses. Accordingly, positive values indicate faster responses and fewer errors for bottom-key responses, and negative values indicate faster responses and fewer errors for the top-key responses. Values for dRT and dError were analyzed by means of a repeated-measures ANOVA with

pitch (low and high) and hand assignment (right-hand top/left hand bottom vs. right hand bottom/left hand top) as within-subjects variables and group (blind vs. sighted) as the between-subjects variable.

Results

RT analyses

Mean response latencies for correct responses are reported in Table 2. The ANOVA on mean correct RT with block as a within-subjects variable and group as a between-subjects variable revealed that participants were overall faster in the second block, $F(1, 44) = 15.89, p < .01, \eta_p^2 = .26$, likely reflecting practice effects (note that timbre-key assignment was the same in the two blocks; what changed was only the hand-key assignment). Neither the main effect of group, $F(1, 44) < 1, p = .49, \eta_p^2 = .01$, nor the interaction group by block, $F(1, 44) < 1, p = .44, \eta_p^2 = .01$, were significant.

The ANOVA on the dRTs with pitch and hand assignment (i.e., right hand top/left hand bottom vs. right hand bottom/left hand top) as within-subjects variables and group as the between-subjects variable yielded a significant main effect of pitch, $F(1, 44) = 11.13, p = .002, \eta_p^2 = .20$, with positive dRTs in response to low tones and negative dRT in response to high tones, consistent with the characteristics of a SMARC effect (see Fig. 2). Neither the main effect of hand assignment, $F(1, 44) < 1, p = .99, \eta_p^2 = .00$, nor the main effect of group, $F(1, 44) < 1, p = .97, \eta_p^2 = .00$, were significant. None of the interactions reached significance (all $ps > .20$).

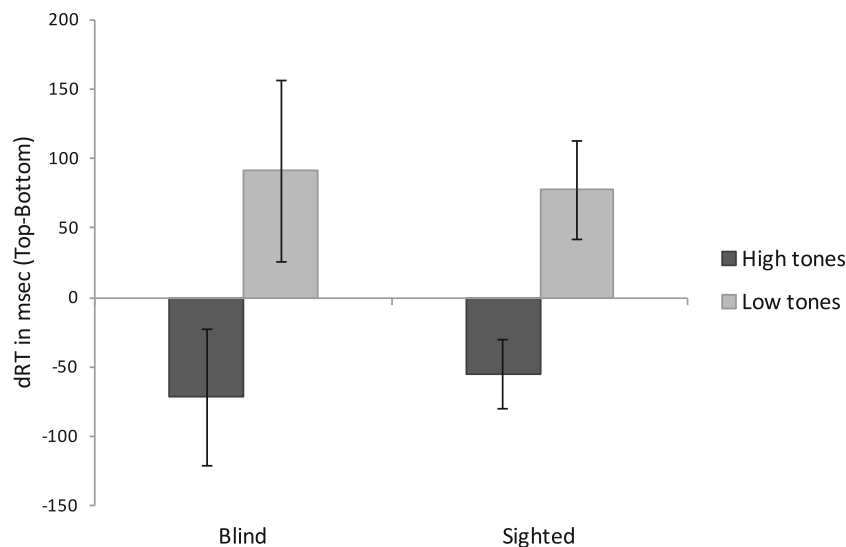


Fig. 2 Differences in reaction times (dRTs) between bottom and top responses as a function of pitch (high, low) and group (blind, sighted). Positive dRTs indicate faster bottom-key than top-key responses. A

significant SMARC effect emerged in both early blind and sighted individuals. Error bars represent $\pm 1SEM$

Table 3 Mean error rates (%) as a function of hand position (right-hand top, left-hand top) and pitch (high, low) in blind and sighted participants

Hand position		Blind		Sighted	
		High	Low	High	Low
Right-hand top	Bottom	15	13	20	7
	Top	10	14	16	13
	dError (top–bottom)	–5	1	–4	6
Left-hand top	Bottom	12	8	20	9
	Top	12	15	15	13
	dError (top–bottom)	0	7	–5	4

Error analyses

Mean error rates for sighted and blind participants in the different conditions are reported in Table 3. The ANOVA on mean error rates with block as the within-subjects variable and group as the between-subjects variable showed that participants improved with practice, with fewer errors made in the second than in the first block, $F(1, 44) = 7.96, p = .007, \eta_p^2 = .15$. Neither the main effect of group $F(1, 44) < 1, p = .57, \eta_p^2 = .00$, nor the interaction group by block, were significant.

Analysis on the dErrors with pitch and hand assignment (i.e., right hand top/left hand bottom vs. right hand bottom/left hand top) as within-subjects variables and group as the between-subjects variable showed no significant main effects of hand assignment, $F(1, 44) = 1.25, p = .26, \eta_p^2 = .02$, or group, $F(1, 44) < 1, p = .87, \eta_p^2 = .00$. Although the pattern of dErrors was consistent with that of the dRT (see Fig. 3), with

participants performing better when responding to low tones by pressing the bottom key and high tones by pressing the top key, the main effect of pitch failed to reach significance, $F(1, 44) = 1.71, p = .19, \eta_p^2 = .03$. None of the interactions reached significance (all $ps > .07$).

Discussion

In this study, we tested the potential effect of normal visual development on the association between pitch and space by comparing sighted and early blind individuals in a stimulus–response compatibility task. We found a consistent SMARC effect with respect to response latencies (see Lidji et al., 2007) in both blind and sighted participants, with all participants responding faster to low tones when these were associated with a bottom-response key press and to high tones when these were associated with a top-response key press. Importantly, in neither group was the SMARC effect modulated by hand assignment (i.e., left or right hand used to press the bottom or top key). Overall, these results extend previous findings suggesting that tones are associated with a vertical spatial continuum, demonstrating that the pitch–space association does not require normal visual experience to develop.

Our findings suggest that the spatial representation of auditory pitch can develop even in the absence of vision, likely deriving from other sensorimotor (and possibly also verbal) experiences. First, it is reasonable that auditory scene statistics for which sounds coming from higher elevations tend to have higher frequencies than those coming from lower elevations (Parise et al., 2014) can be experienced also when the visual

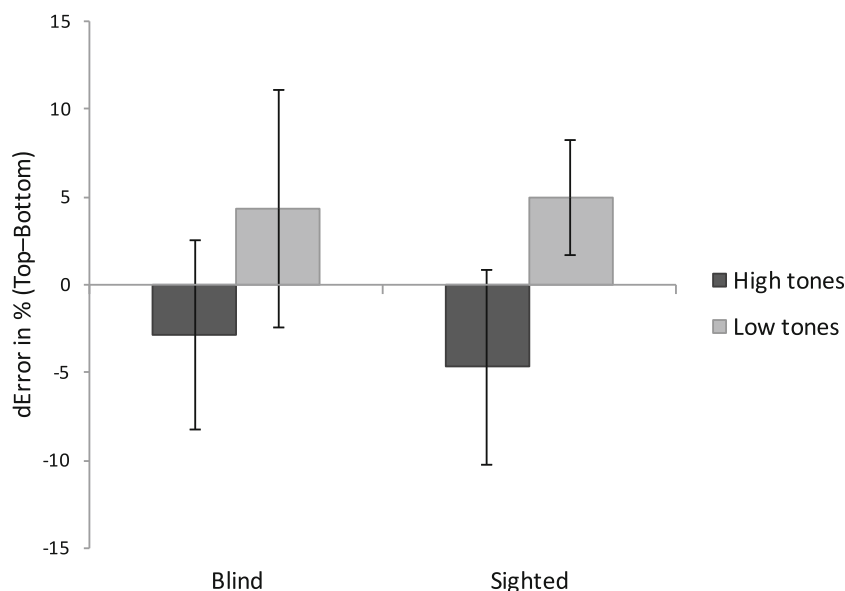


Fig. 3 Differences in percentage of errors (dErrors) between bottom and top responses as a function of pitch (high, low) and group (blind, sighted). Positive dErrors indicate fewer errors for bottom-key responses. Error bars represent $\pm 1SEM$

input is not available. Moreover, the association between pitch and elevation in space may derive from anatomical features related to auditory processing and vocal emission. In particular, Parise et al. (2014) demonstrated that the anatomical properties of the outer ear evolved to mirror the statistical regularities of the external world, therefore suggesting that the mapping between pitch and elevation is at least partially embodied and independent from audio-visual binding. Moreover, low tones tend to resonate in the lower portion of the chest compared to high-pitch sounds (Zbikowski, 1998; see also Shayan, Ozturk, & Sicoli, 2011), and when people produce higher voice frequencies, the larynx moves upward in the throat, whereas it moves downward when they produce lower frequencies (see Connell, Cai, & Holler, 2013). Taken together, these observations suggest an important role of motor and bodily experiences in reinforcing the spatial representation of pitch. In this view, normal visual experience may not be strictly needed (but rather is only supportive) in experiencing the association between pitch and vertical space. Moreover, it is worth mentioning that blind individuals make use of visuo-spatial linguistic metaphors used by the sighted in describing pitch (e.g., “high” and “low” for high-pitched and low-pitched tones; see, for instance Antović, Bennett, & Turner, 2013; Eitan, Omoy, & Granot, 2012; Walker, 1985; Welch, 1991). Thus, the spatial mapping of pitch may also be learned (or reinforced) in the blind by the explicit (verbal) conceptualization of pitch in terms of spatial elevation.

Our results may appear at odds with those reported by Deroy et al. (2016) that failed to find a cross-modal correspondence between auditory pitch and the perceived direction of movement of a tactile stimulation in early blind participants. However, in Deroy et al. (2016), participants were presented with an implicit association task (Greenwald, McGhee, & Schwartz, 1998) in which they had to categorize either a sound or a tactile stimulus delivered to one hand (auditory and tactile stimuli presented in different blocks) by means of a left–right key press using two fingers of the unstimulated hand. In our task, the position of the response keys (and thus of the two hands) was directly related to the mapping of sounds along the vertical dimension, this creating a direct (and possibly more detectable) conflict between the motor response required and the mental spatial mapping of pitch whenever low pitches required a response with top keys and high pitches with bottom keys. Moreover, whereas Deroy et al. (2016) used pure tones of linearly increasing and decreasing pitch, in our study we asked participants to classify single tones of different timbres. It is possible that the type of experimental stimuli used (i.e., pure tones vs. rising/descending sounds) lead to different effects on the saliency of the critical dimension, that is, the perceived distance between low and high tones (Chiou & Rich, 2012; Fernández-Prieto et al., 2012; Mossbridge, Grabowecky, & Suzuki, 2011; see Spence & Deroy, 2013, for a discussion).

In our task, hand assignment (right hand assigned to the top response key or vice versa) did not modulate performance in either blind or sighted participants. There is evidence that pitch may also be mapped (especially by musicians) along a horizontal left-to-right oriented continuum, with low pitches on the left and high pitches on the right (e.g., Lidji et al., 2007; Rusconi et al., 2006). Because of stimulus–response compatibility effects, typically reporting an up-right/down-left advantage (Lippa & Adam, 2001; Weeks, Proctor, & Beyak, 1995), the hand used to respond to low and high pitches may then also affect performance beyond the position of the response keys per se. Accordingly, Lidji et al. (2007) found a tendency toward a stronger vertical SMARC effect in participants having the right hand on the top key and the left hand on the bottom key compared to participants with the reverse mapping (but see Rusconi et al., 2006). We hypothesized that because blind individuals preferentially rely on body–hand centered over external spatial auditory representations (Vercillo et al., 2018), stimulus–response compatibility effects may be more evident in blind than in sighted participants. However, we did not find any hand position effect in either blind or sighted participants, suggesting that the mapping between pitch and vertical space (at least with a paradigm as the one used here) is not affected by the possible activation of hand-centered auditory representations (note further that the effect reported in Lidji et al., 2007, was mainly driven by participants with musical training). Our results suggest that the mapping between pitch and vertical space occurs regardless of the possible activation of hand-centered coordinates in the blind. Moreover, whether pitch is mapped along the horizontal dimension in the blind and whether this possible mapping is anchored to hand position has not been experimentally investigated and deserves further consideration.

In conclusion, while the origin and underlying mechanisms of the association between pitch and spatial elevation is still a topic of debate, our findings show for the first time that early onset profound blindness does not prevent the development of a pitch–space correspondence.

Acknowledgements This work was supported by the Italian Ministry of Education, University and Research (PRIN 2015WXAXJF to Z.C., FIRB FR12F0BD to Z.C., and PRIN 2015AR52F9 to T.V.).

References

- Afonso, A., Blum, A., Katz, B. F., Tarroux, P., Borst, G., & Denis, M. (2010). Structural properties of spatial representations in blind people: Scanning images constructed from haptic exploration or from locomotion in a 3-D audio virtual environment. *Memory & Cognition*, 38(5), 591–604.
- Akiva-Kabiri, L., Linkovski, O., Gertner, L., & Henik, A. (2014). Musical space synesthesia: Automatic, explicit and conceptual connections between musical stimuli and space. *Consciousness & Cognition*, 28, 17–29.

- Antović, M., Bennett, A., & Turner, M. (2013). Running in circles or moving along lines: Conceptualization of musical elements in sighted and blind children. *Musicae Scientiae*, *17*(2), 229–245.
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.
- Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, *57*(8), 1151–1162.
- Bernardi, N. F., Cioffi, M. C., Ronchi, R., Maravita, A., Bricolo, E., Zigiotta, L., . . . & Vallar, G. (2015). Improving left spatial neglect through music scale playing. *Journal of Neuropsychology*, *11*(1), 135–158.
- 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.
- Bodak, R., Malhotra, P., Bernardi, N. F., Cocchini, G., & Stewart, L. (2014). Reducing chronic visuo-spatial neglect following right hemisphere stroke through instrument playing. *Frontiers in Human Neuroscience*, *8*, 413.
- Bottini, R., Crepaldi, D., Casasanto, D., Crollen, V., & Collignon, O. (2015). Space and time in the sighted and blind. *Cognition*, *141*, 67–72.
- Bottini, R., Mattioni, S., & Collignon, O. (2016). Early blindness alters the spatial organization of verbal working memory. *Cortex*, *83*, 271–279.
- Castronovo, J., & Seron, X. (2007). Semantic numerical representation in blind subjects: The role of vision in the spatial format of the mental number line. *The Quarterly Journal of Experimental Psychology*, *60*(1), 101–119.
- Cattaneo, Z., Fantino, M., Tinti, C., Pascual-Leone, A., Silvanto, J., & Vecchi, T. (2011). Spatial biases in peripersonal space in sighted and blind individuals revealed by a haptic line bisection paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1110.
- Cattaneo, Z., Fantino, M., Tinti, C., Silvanto, J., & Vecchi, T. (2010). Crossmodal interaction between the mental number line and peripersonal haptic space representation in sighted and blind individuals. *Attention, Perception, & Psychophysics*, *72*(4), 885–890.
- Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., & Pietrini, P. (2008). Imagery and spatial processes in blindness and visual impairment. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1346–1360.
- Cattaneo, Z., Vecchi, T., Monegato, M., Pece, A., & Comoldi, C. (2007). Effects of late visual impairment on mental representations activated by visual and tactile stimuli. *Brain Research*, *1148*, 170–176.
- Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, *41*(3), 339–353.
- Connell, L., Cai, Z. G., & Holler, J. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and Cognition*, *81*(1), 124–130.
- Crollen, V., Dormal, G., Seron, X., Lepore, F., & Collignon, O. (2013). Embodied numbers: The role of vision in the development of number–space interactions. *Cortex*, *49*(1), 276–283.
- Deroy, O., Fasiello, I., Hayward, V., & Auvray, M. (2016). Differentiated audio-tactile correspondences in sighted and blind individuals. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1204–1214.
- 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.
- Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, *24*(5), 613–621.
- Eitan, Z., Ormoy, E., & Granot, R. Y. (2012). Listening in the dark: Congenital and early blindness and cross-domain mappings in music. *Psychomusicology: Music, Mind, and Brain*, *22*(1), 33–45.
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1). doi: <https://doi.org/10.1167/10.1.6>
- Fernández-Prieto, I., Vera-Constán, F., García-Morera, J., & Navarra, J. (2012). Spatial recoding of sound: Pitch-varying auditory cues modulate up/down visual spatial attention. *Seeing and Perceiving*, *25*, 150–151.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.
- Ishihara, M., Revol, P., Jacquin-Courtois, S., Mayet, R., Rode, G., Boisson, D., . . . & Rossetti, Y. (2013). Tonal cues modulate line bisection performance: Preliminary evidence for a new rehabilitation prospect?. *Frontiers in Psychology*, *4*, 704.
- Lega, C., Cattaneo, Z., Merabet, L. B., Vecchi, T., & Cucchi, S. (2014). The effect of musical expertise on the representation of space. *Frontiers in Human Neuroscience*, *8*, 250.
- 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.
- Lippa, Y., & Adam, J. J. (2001). An explanation of orthogonal SR compatibility effects that vary with hand or response position: The end-state comfort hypothesis. *Attention, Perception, & Psychophysics*, *63*(1), 156–174.
- Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, *116*(4), 323–336.
- Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133–136.
- Mossbridge, J. A., Grabowecy, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual–spatial attention. *Cognition*, *121*(1), 133–139.
- Noordzij, M. L., Zuidhoek, S., & Postma, A. (2006). The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. *Cognition*, *100*(2), 321–342.
- Noordzij, M. L., Zuidhoek, S., & Postma, A. (2007). The influence of visual experience on visual and spatial imagery. *Perception*, *36*, 101–112.
- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, *111*(16), 6104–6108.
- 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. *Perception*, *41*(7), 854–861.
- Pasqualotto, A., Taya, S., & Proulx, M. J. (2014). Sensory deprivation: Visual experience alters the mental number line. *Behavioural Brain Research*, *261*, 110–113.
- Patching, G. R., & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology–Human Perception and Performance*, *28*(4), 755–774.
- Postma, A., Zuidhoek, S., Noordzij, M. L., & Kappers, A. M. (2007). Differences between early-blind, late-blind, and blindfolded-sighted people in haptic spatial-configuration learning and resulting memory traces. *Perception*, *36*(8), 1253–1265.
- Rinaldi, L., Lega, C., Cattaneo, Z., Girelli, L., & Bernardi, N. F. (2016). Grasping the sound: Auditory pitch influences size processing in motor planning. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(1), 11–22.

- Rinaldi, L., Vecchi, T., Fantino, M., Merabet, L. B., & Cattaneo, Z. (2015). The effect of hand movements on numerical bisection judgments in early blind and sighted individuals. *Cortex*, *71*, 76–84.
- Rinaldi, L., Vecchi, T., Fantino, M., Merabet, L. B., & Cattaneo, Z. (2017). The ego-moving metaphor of time relies on visual experience: No representation of time along the sagittal space in the blind. *Journal of Experimental Psychology-General*. doi:<https://doi.org/10.1037/xge0000373>
- Röder, B., Kusmirek, A., Spence, C., & Schicke, T. (2007). Developmental vision determines the reference frame for the multisensory control of action. *Proceedings of the National Academy of Sciences*, *104*(11), 4753–4758.
- Roffler, S. K., & Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *The Journal of the Acoustical Society of America*, *43*(6), 1255–1259.
- 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.
- Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M. J. (2007). Time, also flies from left to right. *Psychonomic Bulletin & Review*, *14*, 512–516.
- Shayan, S., Ozturk, O., & Sicoli, M. A. (2011). The thickness of pitch: Crossmodal metaphors in Farsi, Turkish, and Zapotec. *The Senses and Society*, *6*(1), 96–105.
- Shelton, B. R., & Searle, C. L. (1980). The influence of vision on the absolute identification of sound-source position. *Attention, Perception, & Psychophysics*, *28*(6), 589–596.
- Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness & Cognition*, *22*(1), 245–260.
- Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, *15*(8), 520–535.
- Szűcs, D., & Csépe, V. (2005). The parietal distance effect appears in both the congenitally blind and matched sighted controls in an acoustic number comparison task. *Neuroscience Letters*, *384*(1), 11–16.
- Tabry, V., Zatorre, R. J., & Voss, P. (2013). The influence of vision on sound localization abilities in both the horizontal and vertical planes. *Frontiers in Psychology*, *4*, 932.
- Timmers, R., & Li, S. (2016). Representation of pitch in horizontal space and its dependence on musical and instrumental experience. *Psychomusicology: Music, Mind, and Brain*, *26*(2), 139.
- Vercillo, T., Tonelli, A., & Gori, M. (2018). Early visual deprivation prompts the use of body-centered frames of reference for auditory localization. *Cognition*, *170*, 263–269.
- Voss, P., Tabry, V., & Zatorre, R. J. (2015). Trade-off in the sound localization abilities of early blind individuals between the horizontal and vertical planes. *Journal of Neuroscience*, *35*(15), 6051–6056.
- Walker, A. R. (1985). Mental imagery and musical concepts: Some evidence from the congenitally blind. *Bulletin of the Council for Research in Music Education*, *85*, 229–237.
- 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.
- Weeks, D. J., Proctor, R. W., & Beyak, B. (1995). Stimulus–response compatibility for vertically oriented stimuli and horizontally oriented responses: Evidence for spatial coding. *The Quarterly Journal of Experimental Psychology*, *48*(2), 367–383.
- Weis, T., Estner, B., & Lachmann, T. (2016). When speech enhances Spatial Musical Association of Response Codes: Joint spatial associations of pitch and timbre in nonmusicians. *The Quarterly Journal of Experimental Psychology*, *69*(9), 1687–1700.
- Welch, G. F. (1991). Visual metaphors for sound: A study of mental imagery, language and pitch perception in the congenitally blind. *Canadian Journal of Research in Music Education*, *33*(1), 215–222.
- Zbikowski, L. M. (1998). Metaphor and music theory: Reflections from cognitive science. *Music Theory Online*, *4*(1), 1–8.