

The higher the pitch the larger its crossmodal influence on visuospatial processing

Psychology of Music

1–12

© The Author(s) 2017

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0305735616684205

journals.sagepub.com/home/pom



Irune Fernández-Prieto and Jordi Navarra

Abstract

High-pitched sounds generate larger neural responses than low-pitched sounds. We investigated whether this neural difference has implications, at cognitive level, for the “vertical” representation of pitch. Participants performed a speeded detection of visual targets that could appear at one of four different spatial positions. Rising or falling frequency sweeps were randomly presented before the visual target. Faster reaction times to visual targets appearing above (but not below) a central fixation point were observed after the presentation of rising frequencies. No significant effects were found for falling frequency sweeps and visual targets presented below fixation point. These results suggest that the difference in the level of arousal between rising and falling frequencies influences their capacity for generating spatial representations. The fact that no difference was found, in terms of crossmodal effects, between the two upper positions may indicate that this “spatial representation of pitch” is not specific for any particular spatial location but rather has a widespread influence over stimuli appearing in the upper visual field. The present findings are relevant for the study of music performance, the design of musical instruments, and research in areas where visual and auditory stimuli with certain complexity are combined (music in advertisements, movies, etc.).

Keywords

auditory perception, auditory pitch, crossmodal correspondences, spatial encoding of pitch, visuospatial attention

Auditory pitch seems to be mapped onto vertical coordinates in certain conditions (see Occelli, Spence, & Zampini, 2009; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006; Sonnadara, Gonzalez, Hansen, Elliott, & Lyons, 2009). Higher frequencies, for example, appear to be perceived as originating from higher positions in space (Bregman & Steiger, 1980; Pratt, 1930). Since the ancient Greeks, or even before, music notation has followed similar pitch-to-space

Fundació Sant Joan de Déu, Barcelona, Spain

Corresponding author:

Irune Fernández-Prieto, Fundació Sant Joan de Déu, Santa Rosa, 39–57, Esplugues del Llobregat, Barcelona, 08015, Spain.

Email: irunefernandez@gmail.com

mappings. For instance, the discovery of the Seikilos epitaph in Aydin (Turkey), from around 100 AD, gave us an example of how the visual representation of pitch in music notation followed this “crossmodal congruency” even in the ancient world.

The perceptual overlap between pitch and spatial elevation has been widely reported in the literature. Melara and O'Brien (1987) demonstrated that these two sensory dimensions are integrated in such a way that any variation perceived in one of them has a direct impact on a classification task involving the other. In this previous study, participants had to rapidly classify, by means of two different response buttons, two different tones (high vs. low) according to their pitch. Therefore, pitch was the task-relevant perceptual dimension. Each tone was presented together with a dot that could appear either above or below the visual middle line. The position of this visual stimulus was irrelevant for the task. The authors observed that participants' reaction times (RTs) were faster when the position of the dot was crossmodally congruent with the tone (e.g., the dot appeared above the middle line and was presented together with the high tone) than when they were incongruent. Moreover, the responses when classifying stimuli according to the relevant dimension slowed down as a consequence of variation perceived in the irrelevant dimension (Garner, 1974). Similar results were found when the participants had to judge the position of the dot and ignore any change in pitch.

Recent studies have also shown that reaction times (RTs) at judging differences in pitch between auditory stimuli can be modulated by the spatial location of the response button. The response to a sound that is higher in frequency with respect to a reference sound is faster when it implies an upward movement (Rusconi et al., 2006; see also Sonnadara et al., 2009). Similar results have been obtained using a large variety of experimental methods (see Occelli et al., 2009; Parise & Spence, 2009; Sonnadara et al., 2009), including indirect tasks (Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi et al., 2006). Further studies also suggest that the spatial representation of pitch can even modulate visuospatial attention (see Chiou & Rich, 2012; Mossbridge, Grabowecky, & Suzuki, 2011).

The fact that infants prefer to look at a visual stimulus that moves coherently with respect to a tone that progressively increases or decreases in frequency (Walker et al., 2010) suggests the presence of this perceptual association between auditory pitch and spatial elevation from the very first steps of life. In Walker et al.'s study (2010), 3- to 4-month-old infants looked longer at visual stimuli that moved towards the upper part of the screen when they were presented together with a sound containing an ascending frequency sweep than when they were presented with a sound with descending frequency. Dolscheid, Hunnius, Casasanto, and Majid (2014) followed the same procedure as Walker et al. (2010) and found similar results showing crossmodal correspondences in prelinguistic babies. In contrast, however, Lewkowicz and Minar (2014) failed to replicate these effects after conducting five experiments with 4-, 6-, and 8-month-old infants, using both identical and different methods with respect to Walker et al. (2012; see also Walker et al., 2014, for a response to Lewkowicz & Minar's study).

Perceptual and cognitive differences between rising and falling pitch

Why is the music in climax scenes (e.g., in terror movies) so high-pitched? High frequencies are often perceived as being louder and more salient than low frequencies when they are presented at the same physical intensity (see Fletcher & Munson, 1933). In classic studies by Deutsch (1976, 1978), participants listened to two different sounds played simultaneously, each one presented at a different ear, with frequencies of 400 and 800Hz. They reported to hear a fused tone at the ear where the higher tone was presented. This phenomenon could be taken as evidence suggesting that the higher frequencies have a larger influence on the spatial perception

(i.e., lateralization) of the stimuli than the lower frequencies (see also Deutsch & Roll, 1976; von Békésy, 1963).

Several studies conducted with infants indicate a preference for high frequencies at early stages of life. Infants tend to show a predilection for listening to high-pitched than low-pitched speech (Patterson, Muir, & Hains, 1997) and songs (Trainor & Zacharias, 1998). Furthermore, the discrimination between high frequencies seems to precede, during the maturation of infants' auditory system, the discrimination between low frequencies (Olsho, 1984; Olsho, Koch, & Halpin, 1987; Trehub, Schneider, & Endman, 1980). At a neural level, several studies have revealed, using electroencephalography (EEG), larger mismatch negativity (MMN) for higher than for lower deviant tones (Näätänen, 1990; Näätänen, Gaillard, & Mantysalo, 1978; Ruusuvirta & Astikainen, 2012). Previous research using tones that contained frequency sweeps (i.e., dynamically covering a range of frequencies) has revealed a better performance at detecting an increase in frequency than a decrease (Kishon-Rabin, Roth, Dijk, Yinon, & Amir, 2004). In the so-called Doppler Shift, an increase in acoustic frequency is perceived for sounds that are associated with visual stimuli that approach us, even when these sounds are presented at a constant frequency (see Neuhoff, McBeath, & Wanzie, 1996; see also Hassett & Feth, 1999; McBeath & Neuhoff, 2002). This effect suggests the presence of a phenomenological relation between the perception of rising pitch and the alertness generated by objects that move towards us. This illusory percept has been related to the *Doppler Effect* (Doppler, 1842), in which the frequency of a sound generated by a moving object is perceived to be higher, identical and lower than the emitted frequency as the object approaches, passes by and recedes an observer, respectively. For example, the perceived sound of an ambulance siren is perceived as being higher or lower in frequency than the emitted frequency when the ambulance approaches or moves away from the perceiver, respectively. Thus, there seems to be an association between rising frequencies, approaching objects and, arguably, an increase of alertness. In contrast, dynamic sounds with descending frequencies seem to be more related to objects that move away, perhaps inducing a reduction of the level of alertness. This evidence could easily lead to the hypothesis that rising (i.e., low-to-high) frequency sweeps have a larger impact on arousal and/or alertness than descending (i.e., high-to-low) frequency sweeps. An interesting question, addressed in the present study, refers to the possibility that ascending frequency sweeps (or high tones) also have larger inherent spatial properties than descending frequency sweeps (or low tones). This has not been directly explored in previous studies, where possible crossmodal effects have been addressed including collapsed data from both high (or ascending) and low (or descending) pitch (e.g., Rusconi et al., 2006; Sonnadara et al., 2009).

Due to the physical properties of the sound and the spatial separation between the ears' pinnae, humans and other animals can localize high frequencies more accurately than low frequencies (see Masterton, Heffner, & Ravizza, 1969). Unlike low frequencies, which are characterized by long wavelengths, high frequencies have short wavelengths that allow sound localization based on interaural time difference (i.e., the interval between the different moments at which an acoustic signal reaches each ear). As a consequence, the spatial localization of sound relies more on high frequencies than on low frequencies. This fact, combined with the idea that high frequency sounds generate a larger physiological response, could make us expect larger crossmodal effects in sounds with rising pitch than in sounds with descending pitch.

In the present study, we investigated whether tones with rising frequencies are better suited for generating spatial representations than tones with falling frequencies. For this purpose, an adaptation of the Posner cueing paradigm (Posner, 1980) was used. This paradigm was originally designed to study spatial attention. Two boxes were presented on the left and the right of the screen. One of these two boxes was briefly highlighted (spatial cue). After a certain

stimulus-onset asynchrony (SOA), an asterisk (visual target) appeared in the same box as the spatial cue (i.e., valid or congruent trial) or else in the box at the opposite side (i.e., invalid or incongruent trial). Participants were faster at detecting the asterisk in congruent trials than in incongruent trials. Therefore, the detection of a visual target could be facilitated by a previous visuospatial cue that oriented spatial attention towards a specific area of the visual field.

In our modification of the Posner paradigm, rising and falling frequency sweeps were used as spatial cues, under the assumption that they may differ in their capability to modulate (1) the perceiver's arousal and alertness (see Tomatis, 1978), and consequently, (2) visuospatial attention. More specifically, we used the Posner cueing paradigm to test the hypothesis that the detection of visual targets would be more affected by rising frequency sweeps than by falling frequency sweeps. While the majority of studies addressing the spatial representation of pitch have used pure tones, presented either in isolation or else embedded in melodies, only a few of them used pitch-varying (dynamic) stimuli such as frequency sweeps (see Mossbridge et al., 2011; Walker et al., 2010 for exceptions). Because of the possible capability of frequency dynamic sweeps to induce "spatial directionality", they are particularly useful for the study of the spatial representation of pitch.

Vertical and horizontal crossmodal effects between pitch and spatial elevation

While the vertical representation of pitch has received most of the attention in research on crossmodal correspondences, only a few studies have addressed the possible perceptual correspondence between pitch and space in the horizontal axis (see Lidji et al., 2007; Rusconi et al., 2006). The possible horizontal representation of pitch could perhaps be related to the ability of the brain to create a mental representation of quantities (e.g., with small numbers being located on the left side and the large numbers on the right side; see Dehaene, Bossini, & Giraux, 1993; see also Hubbard, Piazza, Pinel, & Dehaene, 2005, for a review). Following an auditory-based reinterpretation of this metaphorical association, low frequencies would be located on the left side (in Western cultures, at least; see Dehaene et al., 1993) and high frequencies on the right. This possible crossmodal association may perhaps explain why musical instruments are often designed following this "left-low/right-high rule" (e.g., piano).

Rusconi and colleagues (2006) reported crossmodal correspondence effects between pitch and horizontal space in experienced musicians. In one of the experiments included in their study, participants completed a speeded pitch discrimination task where they had to compare the frequency of a probe and a reference sound. The results showed that the response times (RTs) were modulated by the spatial location of the response button in the horizontal axis: faster reaction times were observed for "higher" and "lower" responses when participants had to press a key located at the right and at the left side of the keyboard, respectively. However, the fact that no effects were observed in non-musicians that also participated in this study suggests that horizontal representations of pitch are largely driven by experience (e.g., musical training; see also Mossbridge et al., 2011; Chiou & Rich, 2012). According to this speculative interpretation, pitch would preferentially be represented vertically. The experimental approach adopted in the present study (see Figure 1) allowed us to investigate this possible vertical-over-horizontal preference in the spatial representation of pitch in non-musicians.

How specific is the vertical representation of pitch?

Using an adaptation of the Posner cueing paradigm (Posner, 1980), Mossbridge and colleagues (2011) recently showed that perceiving low-to-high (rising) and high-to-low (falling) frequency sweeps facilitated the subsequent detection of a visual stimulus appearing in a spatial position

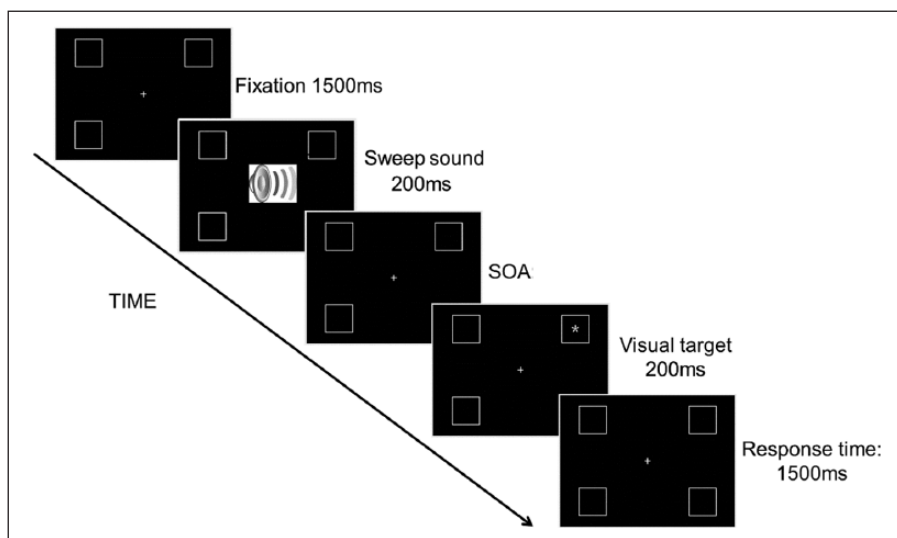


Figure 1. Sequence and timing of the events in a trial. One of two different frequency sweeps (one varying from 200Hz to 700Hz or another from 700Hz to 200Hz) were used as “spatial” cues in a modified version of Posner cueing paradigm.

(upper or lower) that was crossmodally congruent with the sound. Interestingly, this cross-modal correspondence effect vanished when the visual stimulus appeared in one of 4 different positions (left-up corner, right-up corner, left-down corner or right-down corner of the computer screen; see Figure 2), instead of centrally (i.e., above or below a central fixation point). This pattern of results supports a “local”, rather than “global”, account of the spatial representation of pitch, as sounds only influenced the processing of visual stimuli that appeared at a relatively small area immediately above or below the gaze’s fixation point.

Another relevant aspect of Mossbridge et al.’s study (2011) is that the frequency sweeps only ranged from 300 to 450Hz. If pitch can effectively be mapped onto spatial coordinates, a possibility may be that a larger variation (e.g., 500Hz instead of just 150Hz) could “cover” a larger area of the visual space, thus inducing larger cueing effects. Indeed, a plausible hypothesis, tested in the present study, could be that frequency sweeps covering a larger range of sound frequencies (e.g. 200 to 700Hz or 700 to 200Hz) generate a “path” between two relatively specific spatial positions (e.g., between positions B1 and A2 in; see Figure 2).

Keeping in mind that the optimal vertical remapping sound seems to take a certain amount of processing time (i.e., more than 300 ms; see Chiou et al., 2012), two relatively long stimulus onset asynchronies (SOAs: 400 ms and 550 ms) were also used, between the tone and the visual stimulus, to allow for a complete spatial remapping of sound, thus facilitating the appearance of spatial cueing effects.

Methods

Participants

In the current study the inclusion criterion for non-musicians was to have no musical experience as a professional, music student, or high-level amateur (e.g. more than 3 years). Sixteen right-handed non-musician participants (12 females, average age: 21.5), with normal hearing,

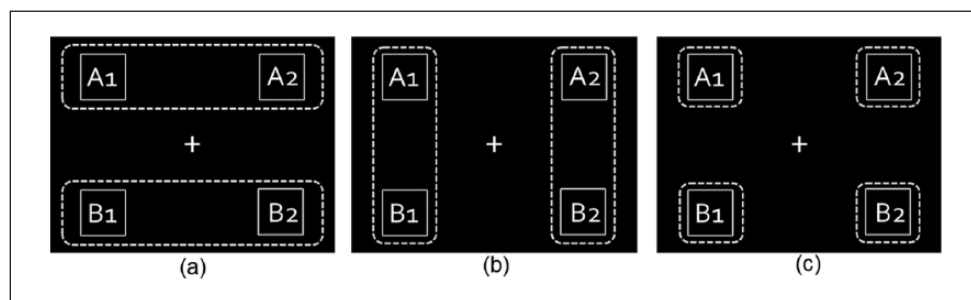


Figure 2. Global vs. positional analysis. (a) In the global vertical cueing analysis, reaction times (RTs) were grouped in two categories: high positions of the screen (including RTs to targets appearing in positions A1 and A2) and low positions (including RTs to targets appearing in B1 and B2). (b) For the global horizontal cueing analysis, RTs were grouped in the two categories left (including A1 and B1) and right (including A2 and B2). (c) A positional cueing analysis was conducted for all of the four different spatial positions (A1, A2, B1 and B2) separately.

and normal or corrected-to-normal vision, took part in the study, and received 6 euros for their participation. None of the participants had received musical training since elementary school. The experiment was conducted in accordance with the Declaration of Helsinki, and had ethical approval from the Hospital Sant Joan de Deú Ethics Committee. The participants provided written informed consent to participate in the study.

Materials

An Intel Core computer and a 15-inch CRT monitor (Philips 107-E Monitor, 85Hz) were used for testing. The experimental procedure was run using E-Prime 2.0 (Psychology Software Tools Inc., Pittsburg, PA) in a dark and soundproof room. The participants sat at a table in front of the monitor at an approximate distance of 60cm. Two loudspeakers (Phillips A 1.2 Fun Power, 7510704863, China) were located at each side of the computer screen.

Procedure

In the variation of the Posner cueing paradigm (Posner, 1980) employed in the present study, the participant had to detect a visual target (a white asterisk of 1.3 cm of diameter) as quickly and as accurately as possible in 320 trials. The visual target could appear at one of four different spatial positions (two above and two below the fixation point; see Figure 2). Rising (200–700Hz) or falling (700–200Hz) frequency sweeps were randomly presented either 400 or 550 ms before the onset of a visual target. These two different SOAs were selected based on previous literature (Chiou & Rich, 2012) and were randomly presented to avoid temporal predictability between the auditory and the visual stimuli. Participants were instructed to press, as fast as possible, using the index finger of their right hand, a key on a computer keyboard after detecting a visual target. The participants' index finger of the right hand rested on the response key during the testing session.

Each trial began with a fixation display (1.5×1.5 cm), consisting of a white central cross flanked by four square-shaped placeholders of 16 cm^2 (two above and two below the fixation point; see Figure 2). After 1500 ms, one of the two possible frequency sweeps (rising or falling; constantly varying from 200 to 700Hz or from 700 to 200Hz, respectively) was presented for 210 ms (with a 5 ms fade-in and fade-out to avoid clicks) at 75dB(A). The

visual target appeared, for 200 ms, after a stimulus-onset asynchrony (SOA) of either 400 or 550 ms, at one of four different spatial positions inside a placeholder (up-right, up-left, down-right or down-left; i.e., positions A1, A2, B1, and B2 in Figure 2). The display (fixation cross and placeholders) remained visible until participants' response with a time limit of 1500 ms.

Results

Reaction times (RTs) faster than 150 ms were considered anticipatory responses and were not included in the statistical analyses. This decision was motivated by the fact that the time needed to process the perceptual information and execute the motor response cannot physiologically be shorter than 150ms. Several statistical analyses were performed to address possible audio-visual cueing effects along the vertical axis (A1 + A2 vs. B1 + B2; see Figure 2), along the horizontal axis (A1 + B1 vs. A2 + B2; see Figure 2), as well as at specific positions on each of the different placeholders (A1 vs. A2 vs. B1 vs. B2; see Figure 2).

“Vertical” analyses

Following previous literature (see Spence, 2011, for a review), we understood the factor “Spatial Congruency” as follows: a target that appeared in an upper position of the screen (A1 or A2 positions) after the presentation of a rising frequency sweep was considered as congruent. Visual targets appearing at lower positions (B1 or B2) after rising frequency sweeps were considered as incongruent. In contrast, visual targets appearing at a lower position (B1 or B2) or at a higher position (A1 or A2), after the presentation of a falling frequency sweep, were considered as congruent and incongruent, respectively (see Figure 3).

A repeated-measures analysis of variance (ANOVA), including the within-subjects factors “Congruency” (congruent vs. incongruent condition), “SOA” (400 vs. 550 ms) and “Direction of the Frequency Sweep” (ascending vs. descending frequency sweep), revealed no interaction between this three factors, $F(1, 15) = .835, p = .375, \eta_p^2 = .053$. Significant effects of congruency and direction of the frequency sweep were found, $F(1, 15) = 5.996, p = .027, \eta_p^2 = .236$; $F(1, 15) = 5.325, p = .036, \eta_p^2 = .262$, respectively. A subsequent ANOVA including only the factors “congruency” and “Direction of the Frequency Sweep” revealed a significant interaction between them, $F(1, 15) = 6.695, p = .018, \eta_p^2 = .318$. No significant differences were found between the two SOAs.

Further analyses with Bonferroni post-tests, including collapsed data from both SOAs and conducted only with trials that contained rising frequency sweeps, revealed significantly faster RTs in the congruent condition than in the incongruent condition, $t(15) = -4.513, p < .01$. No significant differences were found between the congruent and the incongruent condition for falling frequency sweeps, $t(15) = .230, p = .822$ (see Table 1 and Figure 4).

“Horizontal” analyses

In the “Horizontal” analyses, and following previous literature (see Rusconi et al., 2006), targets appearing on the left (A1 and B1) and right (A2 and B2) sides of the screen were considered as congruent after the presentation of falling and rising frequency sweeps, respectively (see Figure 3). Targets appearing on the opposite sides were considered as incongruent. None of the significant effects found in the previous analyses were observed in the “Horizontal” analyses (see Table 1).

Table 1. Participants’ mean reaction times and standard deviations for all conditions. Student’s *t*-tests with a Bonferroni correction were performed between congruent and incongruent conditions.

	Direction of the frequency sweep	Congruent condition	Incongruent condition	Student’s <i>t</i> -test
Vertical cueing effects	Rising pitch	261.917ms (<i>SD</i> = 41.276ms)	272.939ms (<i>SD</i> = 43.992ms)	$t(15) = -4.513$, $p < .01$
	Falling pitch	262.620ms (<i>SD</i> = 40.408ms)	261.800ms (<i>SD</i> = 40.408ms)	$t(15) = .230$, $p = .822$
Horizontal cueing effects	Rising pitch	263.978ms (<i>SD</i> = 42.010ms)	261.356ms (<i>SD</i> = 41.062ms)	$t(15) = 1.314$ $p = .208$
	Falling pitch	263.375ms (<i>SD</i> = 38.747ms)	263.096ms (<i>SD</i> = 43.382ms)	$t(15) = -.657$, $p = .521$
Positional cueing effects	Rising pitch	263.636ms (<i>SD</i> = 40.490ms)	268.692ms (<i>SD</i> = 43.402ms)	$t(15) = -1.684$, $p = .113$
	Falling pitch	264.668ms (<i>SD</i> = 40.312ms)	261.391ms (<i>SD</i> = 37.899ms)	$t(15) = 1.042$, $p = .314$

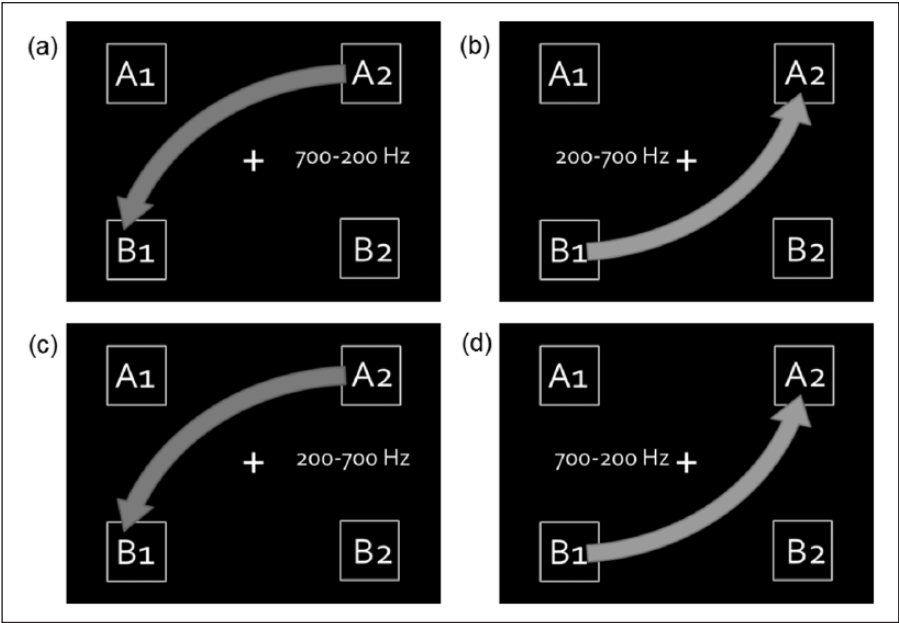


Figure 3. The arrow represents four different hypothetical representations of the auditory cue as spatially directional. (a) Congruent condition (valid cue) with a falling frequency sweep ranging from a high tone to a low tone. (b) Congruent condition (valid cue) with a rising frequency sweep ranging from a low tone to a high tone. (c) Incongruent condition (invalid cue) with a falling frequency sweep ranging from a high tone to a low tone. (d) Incongruent condition (invalid cue) with a rising frequency sweep ranging from a low tone to a high tone.

“Positional” analyses

An ANOVA including the within-subjects factors “SOA”, “Position” (A1, A2, B1, B2; see Figure 3) and “Direction of the Frequency Sweep” revealed an absence of interaction between the

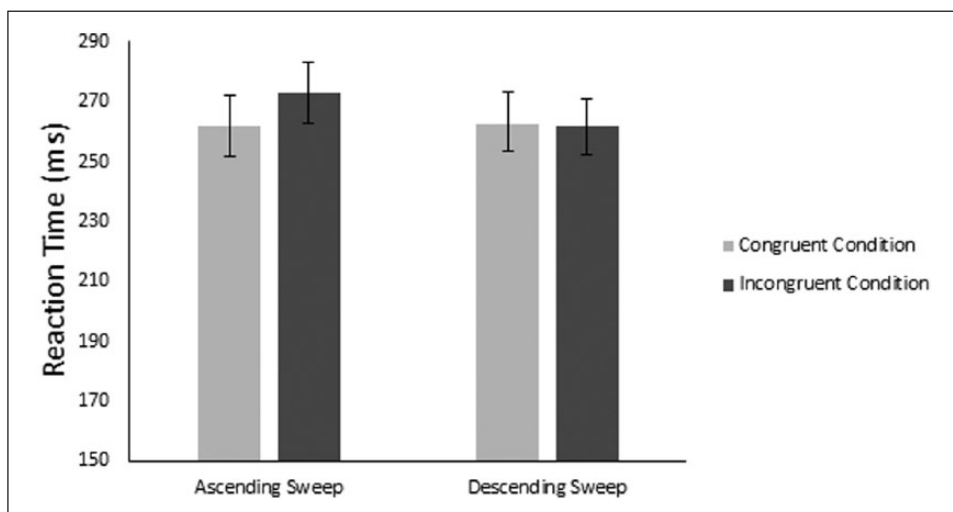


Figure 4. Average of the participants' RTs (in milliseconds) in each condition (congruent and incongruent) for the two different frequency sweeps (rising and falling). Error bars indicate standard error of the mean.

three factors, $F(1, 15) = 1.455$, $p = .272$, $\eta_p^2 = .251$. However, significant differences by position were revealed, $F(1, 15) = 6.134$, $p = .008$, $\eta_p^2 = .686$; reflecting the widely-observed tendency of upward movements to be faster than downward movements.

In an attempt to see whether the spatial cueing effects were “global” (i.e., taking place on the upper or lower wide areas of the screen, including A1 + A2 and B1 + B2, respectively) or “local” (i.e., for specific spatial positions, e.g., A2), more analyses were carried out. *T*-tests, conducted separately for rising and falling frequency sweeps, revealed no significant differences between the RTs in each of the four different positions and the average of RTs in all of the other positions. These analyses allowed us to see whether RTs to visual targets were significantly faster or slower (see Table 1), when preceded by a specific auditory cue, in a particular position (e.g., A2; see Figure 3b) than in the other three positions (A1, B1 and B2). Note that the tests for a possible “global” account of our results (i.e., considering collapsed data from the two upper positions and the two lower positions separately) are presented above (see “Vertical” analyses).

Discussion

Along with previous studies (Chiou & Rich, 2012; Rusconi et al., 2006; Sonnadara et al., 2009) our results suggest that auditory stimuli have inherent spatial properties that can modulate the subsequent spatial processing of visual stimuli by means of spatial cueing. However, this cueing effect was only observed, in our study, in certain conditions:

- (1) Rising frequency sweeps elicited faster responses to visual targets presented on the superior part of the screen (see Figure 2) than to visual targets presented on the inferior part of the screen. Falling frequency sweeps did not elicit comparable effects (i.e., faster RTs for visual stimuli presented below fixation point). The fact that crossmodal correspondences occurred for rising but not for falling frequency sweeps may be related to basic differences between them in terms of modulating the perceivers' physiological response (Näätänen, 1990; Näätänen et al., 1978; Ruusuvirta & Astikainen, 2012). Following

our initial hypotheses, this difference between rising and falling frequency sweeps may percolate into their capacity to generate spatial representations and cueing effects based on crossmodal correspondences.

- (2) Frequency sweeps do not modulate, in non-musicians, the detection of visual target appearing at the right or the left side of the visual field (horizontal axis). Crossmodal correspondences between pitch and space occurred along the vertical axis but not along the horizontal axis, suggesting that pitch is preferentially represented vertically, rather than horizontally.
- (3) The spatial cueing effect generated by the rising frequency sweeps was, in our study, not specific for any particular position (e.g., right-up corner, or A2 in our experiment). This result supports a more “global” (i.e., for “up” and “down” positions, in general) than “local” (i.e., for a particular position in space) account of the spatial representation of pitch.
- (4) Rising frequency sweeps seemed to slow down the participants’ responses to visual stimuli that appeared in a crossmodally incongruent spatial position (i.e., below fixation point). However, a neutral (or baseline) condition would be needed to further test this hypothesis and see whether RT effects can be seen in congruent trials (i.e., shorter RTs with respect to neutral trials), incongruent trials (i.e., slower RTs) or in both types of trial.

A possible explanation of why the cueing effects took place only for upper positions may be that low-to-high (rising) frequency sweeps have a larger impact over the perceiver’s arousal than falling sweeps, and also that this effect interacts with the crossmodal (spatial) representation of pitch. In line with classical sound lateralization studies (Deutsch, 1976, 1978; Deutsch & Roll, 1976), the auditory system presents a perceptual bias for high frequencies over lower frequencies. Higher frequencies seem to drive sound localization: competitive sounds are usually perceived in the ear that received the highest frequency. We believe that crossmodal correspondences occur more preponderantly for ascending frequency sweeps due to the particular properties of high (and, arguably, ascending) frequencies: they can, for example, increase the perceiver’s alertness (Tomatis, 1978). The fact that high-pitched sounds generate more psycho-physiological response (as measured in EEG) than low-pitched tones (see Näätänen, 1990; Näätänen et al., 1978; Ruusuvirta & Astikainen, 2012) may also support our interpretation of the results. At an early age, our auditory system seems to be more tuned to perceive high frequencies than low frequencies (Olsho, 1984; Olsho et al., 1987; Trehub et al., 1980).

Our data did not indicate the presence of any spatial representation of pitch along the horizontal axis in non-musicians. Despite the fact that intense musical training can modify and increase the spatial encoding along the horizontal axis (see Rusconi et al., 2006), our results indicate that pitch is predominantly encoded vertically in the absence of intensive musical training. Furthermore, and considering evidence from a previous study by Lidji and collaborators (2007), the use of indirect speeded tasks (e.g., detecting a visual target, as in the present study) does not seem to produce crossmodal correspondence effects between pitch and space in the horizontal plane in non-musicians. Further results obtained by Chiou and Rich (2012), in which no evidence was found indicating the presence of crossmodal correspondence along the horizontal axis in non-musicians, may provide further support for the idea that pitch is preferentially represented vertically in listeners with no musical expertise.

As Figure 3 reveals, the frequency sweep used in the present study could plausibly have originated a “sense of direction”, perhaps moving the focus of attention to a specific area of the superior or inferior visual field (e.g., a rising frequency sweep cueing position A2). However, our data suggests that the effects of the spatial representation of rising pitch are widespread and non-directional rather than location-specific and directional.

Finally, the longer RTs observed in the incongruent condition are in line with another recent study conducted in our laboratory, in which a larger amplitude of the P3b visual-evoked potential was observed as a consequence of a mismatch between a “spatial expectation” generated by a highly predictable melody and the spatial location of a visual target (Puigcerver et al., 2016). Taken together, the results of both studies may suggest that auditory stimuli containing changes in pitch (e.g., sounds with rising pitch) can modulate the perceptual system’s reaction to upcoming visual targets that appear in particular spatial positions. Thus, these findings could perhaps have a significant impact in several disciplines related to sound processing and music. The design of musical instruments, loudspeakers, or digital platforms for music production and edition may perhaps take our results into account to balance (or take advantage of) the different psychological effects of perceiving low and high frequencies.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Bregman, A., & Steiger, H. (1980). Auditory streaming and vertical localization: Interdependence of “what” and “where” decisions in audition. *Perception & Psychophysics*, 28(6), 539–546.
- Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, 41(3), 339–353.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and numerical magnitude. *Journal of Experimental Psychology: General*, 122, 371–396.
- Deutsch, D. (1976). Lateralization by frequency in dichotic tonal sequences as a function of interaural amplitude and time differences. *Journal of the Acoustical Society of America*, 60(S1), S50–S50.
- Deutsch, D. (1978). Lateralization by frequency for repeating sequences of dichotic 400- and 800-Hz tones. *Journal of the Acoustical Society of America*, 63(1), 184–186.
- Deutsch, D., & Roll, P. L. (1976). Separate “what” and “where” decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance*, 2(1), 23–29.
- 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.
- Doppler, C. (1842). *Ueber das farbige Licht der Doppelsterne und einiger anderer Gestirne des Himmels: Versuch einer das Bradley'sche Aberrations-Theorem als integrierenden Theil in sich schliessenden allgemeineren Theorie*. Prague: Borrosch & André.
- Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America*, 5(2), 82–108.
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum.
- Hassett, S. M., & Feth, L. L. (1999). Just discriminable change of velocity of a simulated moving sound source. *Journal of the Acoustical Society of America*, 106(4), 2209–2209.
- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, 6(6), 435–448.
- Kishon-Rabin, L., Roth, D. A. E., Dijk, B. V., Yinon, T., & Amir, O. (2004). Frequency discrimination thresholds: The effect of increment versus decrement detection of frequency. *Journal of Basic and Clinical Physiology and Pharmacology*, 15(1–2), 29–40.
- 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.
- 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.
- Masterton, B., Heffner, H., & Ravizza, R. (1969). The evolution of human hearing. *Journal of the Acoustical Society of America*, 45(4), 966–985.

- McBeath, M. K., & Neuhoﬀ, J. G. (2002). The Doppler Eﬀect is not what you think it is: Dramatic pitch change due to dynamic intensity change. *Psychonomic Bulletin & Review*, 9(2), 306–313.
- Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116(4), 323–336.
- Mossbridge, J. A., Grabowecky, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual-spatial attention. *Cognition*, 121(1), 133–139.
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, 13(2), 201–233.
- Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention eﬀect on evoked potential reinterpreted. *Acta Psychologica*, 42(4), 313–329.
- Neuhoﬀ, J. G., McBeath, M. K., & Wanzie, W. C. (1999). Dynamic frequency change influences loudness perception: A central, analytic process. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4), 1050.
- Ocelli, V., Spence, C., & Zampini, M. (2009). Compatibility eﬀects between sound frequency and tactile elevation. *Neuroreport*, 20(8), 793–797.
- Olsho, L. W. (1984). Infant frequency discrimination. *Infant Behavior and Development*, 7(1), 27–35.
- Olsho, L. W., Koch, E. G., & Halpin, C. F. (1987). Level and age eﬀects in infant frequency discrimination. *Journal of the Acoustical Society of America*, 82(2), 454–464.
- 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.
- Patterson, M., Muir, D. W., & Hains, S. (1997). *Infant sensitivity to perturbations in adult infant-directed speech during social interactions with mother and stranger*. Poster presented at the Society for Research in Child Development, New Orleans, LA.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3–25.
- Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, 13, 278–285.
- Puigcerver, L., Vera-Constán, F., Rodríguez-Cuadrado, S., Romero-Rivas, C., Fernández-Prieto, I., & Navarra, J. (2016, June). *Seeing music: The perception of melodic ups and downs modulates the spatial processing of visual stimuli*. Oral communication presented at the Barcelona Computational, Cognitive and Systems Neuroscience (BARCCSYN), Barcelona, Spain.
- 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.
- Ruusuvirta, T. T., & Astikainen, P. (2012). Mismatch negativity of higher amplitude for melodic ascendance than descendance. *Neuroreport*, 23(4), 220–223.
- Sonnadara, R. R., Gonzalez, D. A., Hansen, S., Elliott, D., & Lyons, J. L. (2009). Spatial properties of perceived pitch: Influence on reaching movements. *Annals of the New York Academy of Sciences*, 1169, 503–507.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4), 971–995.
- Tomatis, A. A. (1978). Presentation at International Kodaly Symposium in Paris.
- Trainor, L. J., & Zacharias, C. A. (1998). Infants prefer higher-pitched singing. *Infant Behavior and Development*, 21(4), 799–805.
- Trehub, S. E., Schneider, B. A., & Endman, M. (1980). Developmental changes in infants' sensitivity to octave-band noises. *Journal of Experimental Child Psychology*, 29(2), 282–293.
- von Békésy, G. (1963). Three experiments concerned with pitch perception. *Journal of the Acoustical Society of America*, 35(4), 602–606.
- 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.
- 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.