Seeing music: The perception of melodic 'ups and downs' modulates the spatial processing of visual stimuli

Carlos Romero-Rivas[a], Fátima Vera-Constán[b,c], Sara Rodríguez-Cuadrado[a,b], Laura Puigcerver[b,d], Irune Fernández-Prieto[b,d,e], Jordi Navarra[b,d,*]

[a] Department of Psychology, Edge Hill University, Ormskirk, UK
[b] Fundació Sant Joan de Déu, Psychiatry and Psychology Service, Hospital Sant Joan de Déu, Esplugues de Llobregat, Barcelona, Spain
[c] Departamento de Metodología y Psicología Básica, Universidad de Murcia, Murcia, Spain
[d] Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona, Spain
[e] Neuropsychology & Cognition Group, Department of Psychology and Research Institute for Health Sciences (iUNICS), University of Balearic Islands, Palma, Spain

ARTICLE INFO

ABSTRACT

Musical melodies have "peaks" and "valleys". Although the vertical component of pitch and music is well-known, the mechanisms underlying its mental representation still remain elusive. We show evidence regarding the importance of previous experience with melodies for crossmodal interactions to emerge. The impact of these crossmodal interactions on other perceptual and attentional processes was also studied. Melodies including two tones with different frequency (e.g., E4 and D3) were repeatedly presented during the study. These melodies could either generate strong predictions (e.g., E4-D3-E4-D3-E4-[D3]) or not (e.g., E4-D3-E4-E4-D3-[?]). After the presentation of each melody, the participants had to judge the colour of a visual stimulus that appeared in a position that was, according to the traditional vertical connotations of pitch, either congruent (e.g., high-low-high-low-[up]), incongruent (high-low-high-low-[down]) or unpredicted with respect to the melody. Behavioural and electroencephalographic responses to the visual stimuli were obtained. Congruent visual stimuli elicited faster responses at the end of the experiment than at the beginning. Additionally, incongruent visual stimuli that broke the spatial prediction generated by the melody elicited larger P3b amplitudes (reflecting 'surprise' responses). Our results suggest that the passive (but repeated) exposure to melodies elicits spatial predictions that modulate the processing of other sensory events.

1. Introduction

The perception of melodies is based on interpreting successive variations in pitch. A parsimonious way to perceive melodies may consist in mapping acoustic frequencies onto vertical coordinates, as if representing them on the stave (Stumpf, 1883). The fact that musical pitch is formally represented vertically and is described using words with spatial connotations in almost all cultures of the World may not be coincidental (see Fernández-Prieto, Spence, Pons and Navarra, 2017; Parkinson et al., 2012). Concurring with this idea, a robust association has been observed, both in adults and in prelinguistic infants, between high- and low-pitched sounds and high and low positions in the external space, respectively (Bernstein and Edelstein, 1971; Cabrera and Morimoto, 2007; Chiou and Rich, 2012; Deroy et al., in press; Lidji et al., 2007; Maeda et al., 2004; Melara and Brien, 1987; Mossbridge et al., 2011; Rusconi et al., 2006; Sonnadara et al., 2009; Spence and

Deroy, 2013). Additionally, previous findings suggest that experience with music improves performance in spatial tasks (the "Mozart effect"; Rauscher, Shaw, and Ky, 1993). In contrast with listeners with no musical expertise, musicians also show crossmodal correspondence effects between pitch and spatial elevation in indirect testing; that is, when the experimental task does not imply any spatial or pitch-based judgment (Rusconi et al., 2006).

By means of repetitive exposure to melodies, we learn to predict their pitch contour (i.e., the ascending-descending dynamics of sound frequency over time). If experience with music influences both their predictive capability and their spatial connotations, it is reasonable to hypothesize that the repeated exposure to melodies will (1) increase their capability to generate melodic predictions, and (2) generate stronger spatial connotations that can even modulate visuospatial attention. In the present study, we investigated whether the passive listening of melodies biases visuospatial attention towards high and low
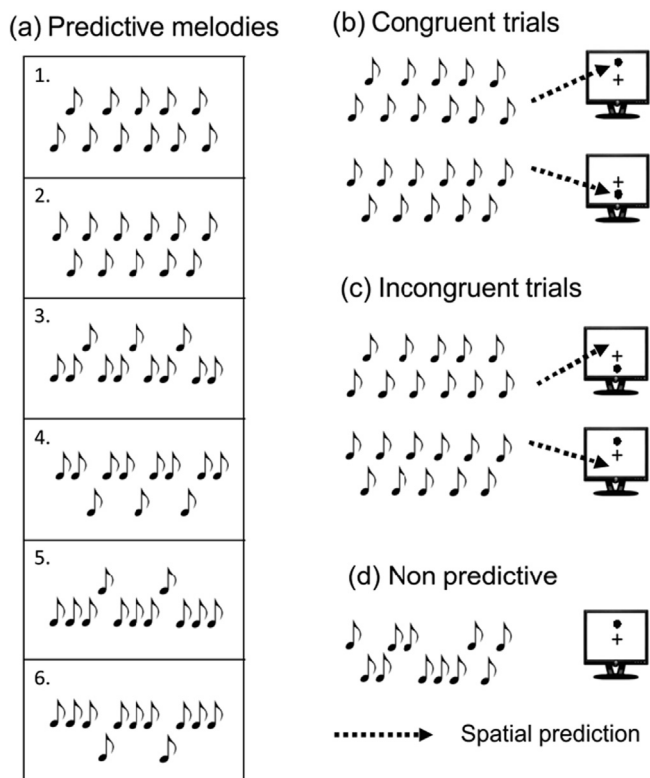
Fig. 1. Schematic representation of the different predictable melodies (a) and experimental conditions. After the presentation of the melody, a visual target (a small filled circle) appeared and the participants performed a speeded colour discrimination task (yellow vs. blue) of the circle. The visual target could either be unpredictable (d) or be crossmodally congruent (b) or incongruent (c) with respect to the possible spatial (vertical) expectancy generated by a predictable melody. Each melody appeared 24 times during the experiment.

spatial positions. Additionally, the role of recent experience to generate these biases was addressed.

Even if the activation of specific space-related areas of the brain such as the intra-parietal sulcus (IPS) is observed when people perform complex musical tasks (see Foster and Zatorre, 2010a, 2010b), it is still unknown whether the spatial connotations of melodic ascents and descents also arise in conditions of repeated passive listening or not. To test this possibility, a group of participants heard simple 11-tone melodies several times during the study. These melodies could or could not generate predictability (e.g., "low-high-low-high-low…" vs. "low-low-high-low-high-high…"; see Fig. 1). After each of the auditory streams, a visual target (either a blue or a yellow filled circle) was presented in an upper or a lower spatial position on a computer screen. No task was performed in relation to the melodies. Regarding the visual target, an indirect (i.e., non-spatial) task, consisting of judging the colour of the stimuli, was performed by the participants.

In the predictable conditions, the visual target could either be crossmodally congruent or else incongruent with respect to the prediction generated by the melody. For example, a visual target appearing on the upper (high) position after the melody "low-low-high-low-low-high-low-low-high-low-low-" led to a congruent trial, while a visual target appearing on the lower spatial position after the presentation of this melody was incongruent. In the unpredictable condition, the participants were not able to predict the pitch changes during the presentation of the melody.

The possible effect of gaining experience with the melodies on their capability to modulate visuospatial processing was explored by comparing the possible crossmodal effects (e.g., faster reaction times -RTs- in congruent trials) in the first and the last trials of the experiment. We

expected that RTs would be faster in congruent trials than in incongruent and unpredictable trials, especially in the last trials of the experiment (see Nobre et al., 2007). This would reflect the effect of recent experience with the melodies on their capacity to modulate spatial attention. On the other hand, no differences were expected, in terms of RTs, between the incongruent and the unpredictable conditions when considering the first and last trials of the experiment, because incongruities in a predictable melody may cause similar response slowness than unpredictable melodies.

In order to address the possible influence of auditory prediction on visuospatial attention, we analysed the event-related potentials (ERPs) from the onset of each visual target. The decision to use ERPs was based on their high temporal resolution, which allowed us to explore the processing of the visual targets after the presentation of the different melodies. If a task-irrelevant rhythmic stream of sounds can bias visuospatial attention and then modulate the processing of forthcoming visual stimuli even in non-musician listeners, larger amplitudes could be expected in attention and memory-mediated components such as P2 (Federmeier and Kutas, 2002), P3a (Guo et al., 2006; McEvoy et al., 2001; Rugg and Doyle, 1992) or P3b (Brázdil et al., 2003, 2001; Conroy and Polich, 2007; Hartikainen and Knight, 2003; Knight, 1996; Kok, 2001; Polich, 2007, 2003; Squire and Kandel, 1999). For instance, the P2 component has been related to the use of contextual cues to prepare for the visual analysis of upcoming stimuli (Federmeier and Kutas, 2002). Hence, we would expect larger P2 amplitudes for predictive than unpredictive melodies, since predictive melodies should allow participants to prepare for the upcoming targets. Somewhat related, the P3a component has been observed to index stimulus representations maintained in memory from previous exposures, being this component larger for recurring than for novel stimuli (Guo et al., 2006; McEvoy et al., 2001; Rugg and Doyle, 1992). Therefore, since our predictive melodies were more easily remembered than the unpredictive melodies, we would expect larger P3a amplitudes for the former. In addition, the P3a amplitude is sensitive to the amount of attentional resources that are recruited during the performance of a task. For instance, as primary task difficulty is increased, the P3a amplitude decreases (Isreal et al., 1980; Kok, 2001; Kramer et al., 1985; Polich, 1987; Wickens et al., 1983). Accordingly, if unpredictable melodies make the processing of the visual target more difficult, we would also expect that this crossmodal factor might modulate the P3a amplitude (i.e., larger amplitude for predictive than for unpredictive melodies). This prediction would be somewhat congruent with previous literature showing that, when processing audiovisual rhythmic streams, the primary visual cortex entrains to the rhythm of the stream, resulting in decreased reaction times to targets (Lakatos, Karmos, Mehta, Ulbert, and Schroeder, 2008), and, importantly, that the degree of entrainment depends on the predictability of the stream (Besle et al., 2011). Finally, while the P3a component is usually related to the engagement of focal attention during stimulus evaluation, the P3b component is supposed to occur when subsequent attentional resources are needed to carry out context-updating operations and subsequent memory storage (Brázdil et al., 2001, 2003; Conroy and Polich, 2007; Hartikainen and Knight, 2003; Knight, 1996; Kok, 2001; Polich, 2003, 2007; Squire and Kandel, 1999). Consequently, we would expect similar P3a amplitudes for congruent and incongruent visual targets (similar focal attention during stimulus processing). However, we would also expect differences between congruent and incongruent visual targets in the P3b amplitude, since incongruent trials would require further context-updating operations.

In addition, we also analysed the N50 and N100 ERP components, as the amplitude of these components is modulated by selective attention on visual processing (e.g., Rugg et al., 1987). Following previous literature, larger amplitudes were expected, in these components, for targets following unpredictable (as compared to predictable) melodies (Rugg et al., 1987).

2. Material and methods

2.1. Participants

Nineteen healthy volunteers (15 females; mean age = 24.89, SD = 5.07; 18 right-handed) with no musical training, normal or corrected-to-normal vision and normal hearing by self-report, took part in the study. Electrophysiological data from one participant and behavioural data from three participants were discarded from the analyses due to technical problems (e.g., software failure). Participants received 20 euros in exchange for their participation. The experiment was non-invasive, was conducted in accordance with the Declaration of Helsinki, and had ethical approval from Hospital Sant Joan de Déu and Parc Sanitari Sant Joan de Déu Ethics Committee. An informed consent was obtained from all of the participants.

2.2. Apparatus and material

Predictable and unpredictable 11-tone auditory streams (or melodies) were created by combining two digitally-generated musical tones with clearly different frequencies (one higher in frequency than the other; see Fig. 1). The melodies were played either in a digital piano (by combining E4 and D3, with 329.6 Hz and 146.8 Hz, respectively) or else in a digital organ (by combining E4 and C3, with 329.6 Hz and 130.8 Hz, respectively). The use of two different instruments was introduced to increase stimulus variation, thus reducing the impact of habituation (e.g., in the ERPs) and boredom. All of the tones composing each melody lasted for 240 ms, had a fade-in of 20 ms and a fade-out of 10 ms and were presented every 250 ms. The auditory stimuli were delivered from two speakers (Altec Lansing V52420, China) located on both sides of an LCD monitor (Acer GD245HQ; 24 in.; refresh rate = 60 Hz).

In order to create patterns that were multiple of 12 (i.e., including 11 tones plus the visual stimulus that would "replace" the expected 12th tone; see Fig. 1), the two musical notes were played alternatively, either alone or presented in groups of two or three notes. The patterns started either with lower-pitched or higher-pitched tones, resulting in 6 different predictable melody sequences (see Fig. 1). Noteworthy, the last (and highly predicted) tone was always replaced with the visual stimulus that could appear above or below a gaze fixation cross. Regarding the unpredictable melodies, random sequences of 11 tones, composed of the two tones played on the same instrument, were arranged in such a way that no logical sequence could be extracted from the pattern. The frequency of the first tone of each melody was counterbalanced across trials.

The visual stimuli consisted of a blue or a yellow filled circle (diameter: 1.7° of visual angle) that appeared 350 ms after the onset of the last tone of each auditory stream (i.e., 100 ms after the offset, in order to have a proper baseline for the ERPs), either 3.6° above or below a central fixation cross (.4°x0.4°) on a screen located 150 cm in front of the participant (see Fig. 1). The spatial location of the visual stimuli (either above or below the central fixation cross) was randomised. The visual target disappeared after participant's response or, automatically, after 2000 ms. The presentation of both the auditory and visual stimuli was controlled by E-Prime 2.0 (Psychology Software Tools Inc., Pittsburg, PA) and run on an Intel Core computer.

2.3. Design

The experiment included 432 trials that appeared in 2 different blocks with a 2-min pause in between that allowed participants to have some rest. Each 216-trial block included 72 predictable/congruent, 72 predictable/incongruent and 72 unpredictable trials that were randomly presented. As it is indicated above, the predictable status of each melody was an intrinsically-defined feature of the melody. In contrast, the congruency status of each trial was defined by the spatial position of the visual stimulus.

Half of the melodies could generate a clear prediction, especially after their repeated presentation, regarding the frequency of a hypothetical 12th tone, the other half could not (*unpredictable condition*; see Fig. 1). The purpose of introducing the unpredictable trials was to provide a direct comparison against predictable trials, in order to explore whether participants predicted the 12th tone in the latter case or not.

After each auditory stream, a visual target appeared (instead of the 12th tone of the melody) randomly in one of two different positions: above (upper position) or below (lower position) a central fixation point. Along with the unpredictable condition, the experiment included two different predictable conditions. Each predictable trial started with one of 6 possible combinations of high and low music tones (melodies; see Fig. 1): (1) In the *congruent condition*, the visual target appeared in an expected position with respect to the hypothetical vertical prediction generated by the melody; (2) in the *incongruent condition*, the position of the target did not coincide with the prediction. Each melody was presented 24 times during the experiment (12 times in each block), thus allowing us to behaviourally analyse the possible effects of previous experience.

Both the colour and the position of the visual targets were counterbalanced across the experimental session, which lasted 40 min approximately.

2.4. Procedure

The participants' task was completely orthogonal with respect to either pitch or space and consisted of judging the colour (blue vs. yellow) of the visual target. Participants were instructed to fixate at a central cross during the whole experimental session and to respond as fast as they could by pressing "n" or "m" in a keyboard. Both the participants' RTs and the percentage of correct responses were recorded. Also, we monitored the participants' behaviour through a webcam to ensure that participants were looking at the computer screen during the whole experimental session.

2.5. EEG acquisition

The ERP data were recorded from 28 Ag/AgCl electrodes according to the international 10–20 system (Electro Cap International, Inc). The recording was referenced to an electrode placed at a central position near vertex (FCZ). The AFZ electrode was used as virtual ground. Two additional electrodes were also placed at the two mastoids (LM and RM). Vertical and horizontal eye movements were monitored with two electrodes; one located at the infraorbital ridge (VEOG), and another one at the lateral (HEOG) edge of the right eye. Electrode impedances were kept below 5kOhm. The electrophysiological signals were filtered on-line with a band-pass of .1–100 Hz and a 50 Hz Notch filter. The sampling rate was of 500 Hz. Off-line, EEG raw data were re-referenced to the mastoids average. Then, the EEG signal was filtered with a band-pass of .53–30 Hz (24 dB/oct). Ocular movement rejection was based on VEOG and HEOG electrodes. Trials containing saccades were removed from the analyses. Data from channels FP1 and FP2 were also removed due to a very noisy signal (possibly due to small muscle movements). Epochs (600 ms long) with EEG exceeding ± 100 µV at any channel were discarded from the analysis. A maximal difference criterion of values was also set (100 µV).

2.6. EEG processing and analyses

Data were averaged and analysed separately for each Block and Condition. A mean of 64.34 valid segments were finally included in the analyses for each participant, condition and block: Congruent, Block 1 = 64.37 (SD = 7.11), Block 2 = 62.42 (SD = 10.67); Incongruent, Block 1 = 65.11 (SD = 6.57), Block 2 = 62.68 (SD = 9.30);

Unpredictable, Block 1 = 63.95 (SD = 7.43), Block 2 = 67.89 (SD = 11.46).

An off-line average was carried out for each stimulus type, block, condition, and participant separately. The baseline was set at 100 ms previous to stimuli onset. Statistical analyses were carried out involving data from 21 electrodes (Frontal Region = F3, Fz, F4, FC1, FC2, F7, and F8; Central Region = C3, Cz, C4, CP1, CP2, CP5, and CP6; Posterior Region = P3, Pz, P4, P7, P8, O1, and O2).

Statistical analyses were performed on five main time windows, based on previous literature: N50 (30–80 ms; see Fellinger et al., 2012), N1 (150–200 ms; see Mangun and Hillyard, 1991), P2 (200–275 ms; see Luck and Hillyard, 1994), P3a (275–350 ms), and P3b (350–500 ms; see Squires et al., 1975). EEG recordings and data processing were conducted using Brain Vision software (Brain Vision Recorder and Analyzer 2.0).

3. Results

The accuracy at judging the colour of the visual targets approached 100% correct in all the experimental conditions (Congruent: $M = 97.5\%$, $SE = .007$; Incongruent: $M = 97.7\%$, $SE = .005$; Unpredictable: $M = 97.7\%$, $SE = .008$). A Repeated Measures ANOVA including accuracy as the dependent variable and condition (Congruent, Incongruent, Unpredictable) as the independent variable showed a non-significant effect of Condition ($F(2,30) = .124$, $p = .793$, $\eta_p^2 = .008$). This result was expected considering the easiness of the task. In a paper questionnaire provided after the experiment, none of the participants reported being aware of any association between the prediction generated in some of the trials and the upper or lower position of the visual target.

Another Repeated Measures ANOVA including RT data from all of the trials revealed a main effect of condition (Congruent, Incongruent, and Unpredictable; $F(2,30) = 9.505$, $p = .001$, $\eta_p^2 = .388$). Further analyses indicated that participants responded significantly slower to visual targets in the Unpredictable condition than in the Congruent ($t(15) = -3.342$, $p = .004$) and the Incongruent ($t(15) = -2.911$, $p = .01$) conditions, suggesting the presence of a general predictability effect. In line with previous studies conducted with non-musicians using indirect tests (Rusconi et al., 2006), no differences between congruent and incongruent trials were observed in this initial analysis.

In order to address possible effects of experience and the predictability of the melodies across the whole experiment, a different Repeated Measures ANOVA was conducted including data from the first and the last 20 trials of each condition separately. The number of trials for this analysis was chosen considering previous studies conducted at our laboratory (e.g., Navarra et al., 2009). Results revealed a significant interaction between Condition (Congruent, Incongruent, and Unpredictable) and Exposure (first vs. last 20 trials) ($F(2,30) = 8.237$, $p = .001$, $\eta_p^2 = .354$). Pairwise comparisons revealed no differences between any of the three conditions in the first 20 trials, but significant differences between the Congruent and the Incongruent condition ($t(15) = -2.151$, $p = .048$) and the Congruent and the Unpredictable condition ($t(15) = -3.776$, $p = .002$) in the last 20 trials (see Fig. 2).

Different Repeated Measures ANOVAs were carried out for the amplitude and the latency of several ERP components separately. These analyses included the factors Condition (Congruent, Incongruent, and Unpredictable), Region (Frontal, Central, and Posterior), and Block (Block 1 and Block 2) as within-subject factors. It is worth highlighting that the 'Exposure' analysis performed on the RTs to elucidate the effects of previous experience with melodies on their capability to modulate spatial attention cannot be compared with the 'Block' ERP analysis. While the RTs analyses only included data from the first 20 and last 20 trials of the experimental session, the ERPs analysis also included trials that fell in the middle of the experiment. The two different experimental blocks were present in our design due to the fact that there was a break for the participants. Therefore, the 'Block' ERP
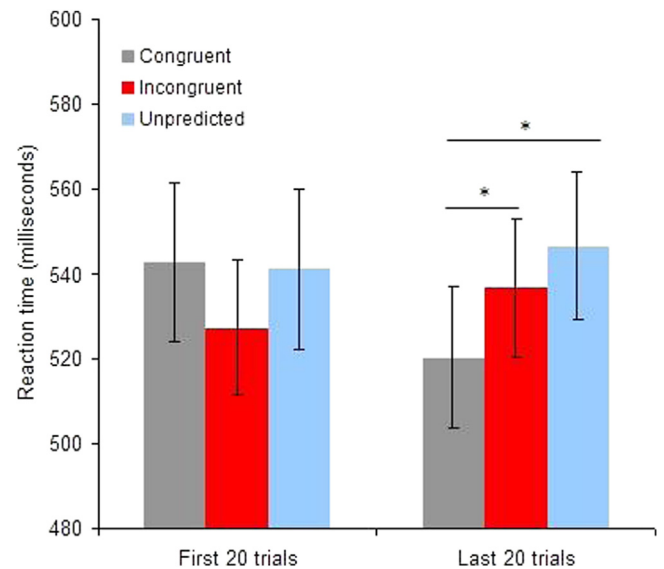

Fig. 2. Behavioural results: Exposure effects on crossmodal congruency between pitch and spatial elevation. Significant differences were found, in RTs, between the congruent and the incongruent condition and the congruent and the unpredictable condition in the last 20 trials of the experiment (3 bars on the right), but not in the first 20 trials (3 bars on the left). Error bars show the standard error of the mean.

analyses included all of the valid trials in these two blocks (not just the first 20 and the last 20 trials). Moreover, 'Block' (i.e., trials before vs. trials after the pause) was introduced as a factor in the ERP analyses just to ensure that the pause did not generate any undesired/uncontrolled effects on the electrophysiological response.

In the analysis of the N50 component, we observed a significant main effect of Condition ($F(2,36) = 4.417$, $p = .019$, $\eta_p^2 = .197$), and a significant interaction between Condition and Region ($F(4,72) = 10.692$, $p < .001$, $\eta_p^2 = .373$). Post-hoc comparisons of the interaction, using Fisher's least significant difference (LSD) method for multiple comparisons (for recent studies using the same approach see Grimaldi et al., 2014; Németh et al., 2014), indicated that Unpredictable trials elicited larger amplitudes than Congruent trials in the frontal ($p = .001$), and central ($p = .014$) regions of the scalp. In addition, Unpredictable trials elicited larger amplitudes than Incongruent trials in the frontal ($p = .005$), and central ($p = .016$) regions of the scalp (see Fig. 3; see also Table 1 for a summary of the results of ERPs analyses). Importantly, there were no significant differences between Congruent and Incongruent trials in any of the three topographic regions.

As for the N1 component, we observed a main effect of Condition ($F(2,36) = 3.807$, $p = .032$, $\eta_p^2 = .175$). LSD post-hoc tests showed that Unpredictable trials elicited larger amplitudes than Congruent trials ($p = .023$). The difference between Incongruent and Unpredictable trials was marginally significant ($p = .082$), with larger amplitudes for Unpredictable than Incongruent trials. Finally, there were no differences between Congruent and Incongruent trials ($p = .458$).

Regarding the P2 component, a significant effect of Condition was found ($F(2,36) = 9.236$, $p = .001$, $\eta_p^2 = .339$), along with a significant interaction between Condition and Region ($F(4,72) = 3.496$, $p = .012$, $\eta_p^2 = .163$). Post-hoc comparisons indicated the general presence of larger amplitudes in Congruent and Incongruent trials than in Unpredictable trials ($p = .005$, and $p = .003$, respectively) over the whole scalp. In frontal regions, the P2 amplitude was larger in the Incongruent than in the Unpredictable condition ($p = .045$; see Fig. 3). The P2 amplitude was also significantly larger both in Congruent and Incongruent than in Unpredicted trials in central and posterior regions of the scalp ($p = .002$, and $p = .001$, respectively, for central
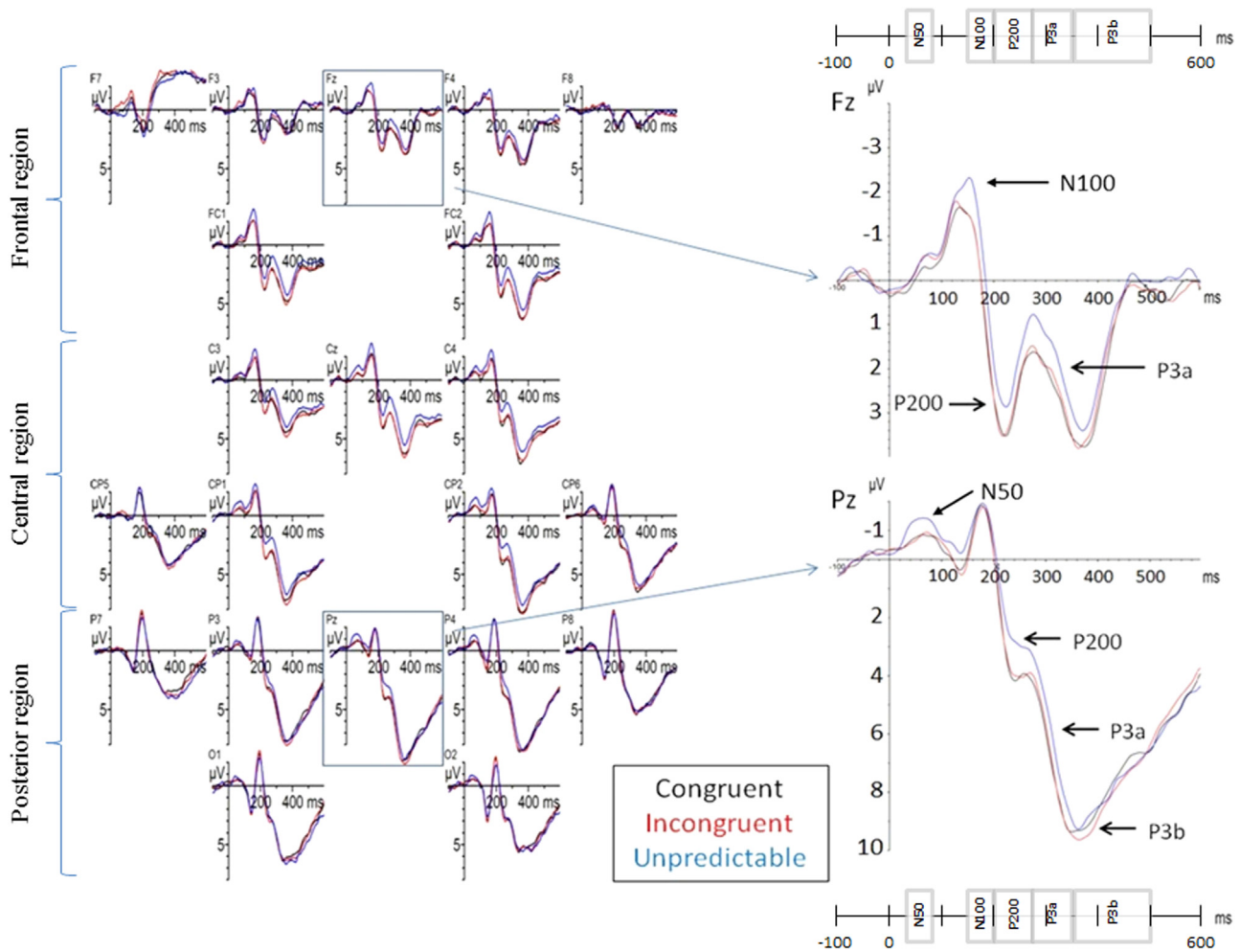
Fig. 3. ERP results: Congruency and Prediction effects. Grand-average ERPs from all of the electrodes are shown. On the right, Fz and Pz electrodes are zoomed, showing Congruent (black line), Incongruent (red line), and Unpredictable (blue line) conditions; visual representations of the time windows for the different ERP components analysed are also shown above and below the zoomed electrodes. Incongruent trials generated larger amplitudes than Congruent trials in the P3b component, in the posterior region of the scalp. Regarding frontal and central areas, the amplitudes of the N50, N100, P2 and P3a components were significantly less positive for Unpredictable than for Predictable (Congruent or Incongruent) trials.

electrodes; and $p = .016$ and $p = .017$, respectively, for posterior electrodes).

Further analyses conducted on the P3a component showed main effects of Condition ($F(2,36) = 17.194$, $p < .001$, $\eta_p^2 = .489$) and Region ($F(2,36) = 9.920$, $p < .001$, $\eta_p^2 = .355$). The interaction between these two factors was also significant ($F(4,72) = 2.907$, $p = .027$, $\eta_p^2 = .139$); revealing larger amplitudes, in frontal ($p = .004$) and central ($p = .024$) regions, in Congruent trials than in Unpredictable trials, as well as larger amplitudes, in the same regions, in Incongruent compared to Unpredictable trials ($p < .001$ and $p < .001$ for frontal

and central regions, respectively).

Finally, in the analysis of the P3b, we observed a main effect of Region ($F(2,36) = 36.396$, $p < .001$, $\eta_p^2 = .669$). No main effects of Condition or Block were observed. A significant interaction was found between Region and Condition ($F(4,72) = 4.491$, $p = .003$, $\eta_p^2 = .200$). Crucially for the purpose of the study, post-hoc analyses revealed a significant difference, in the posterior region of the scalp, between Congruent and Incongruent conditions ($p = .036$); reflecting larger amplitudes of the P3b in the latter (see Figs. 3 and 4). Significantly larger P3b amplitudes were also observed, in central electrodes, in

Table 1
Summary of the results of ERPs analyses.

	N50 (30–80 ms)			N100 (150–200 ms)	P200 (200–275 ms)			P3a (275–350 ms)			P3b (350–500 ms)		
Condition x Region	Frontal	Central	Posterior	Main effect of Condition	Frontal	Central	Posterior	Frontal	Central	Posterior	Frontal	Central	Posterior
Cong. ≠ Incong.													*
Cong. ≠ Unp.	* * *	*		*		* *	*	* *	*			*	
Incong. ≠ Unp.	* *	*		m.s. (p = .08)	*	* * *	*	* * *	* * *			* *	

Overall summary of the results of ERPs analyses. In the first row, the different ERP components and time windows are presented. In the second row, the different topographic regions of analysis are shown (or the main effect of Condition for time windows in which we did not observe an interaction between Condition and Region). In the first column, the tested post-hoc analyses are presented. Cong. = Congruent condition; Incong. = Incongruent condition; Unp. = Unpredictable condition; * = $p < .05$; ** = $p < .01$; *** = $p \leq .001$; m.s. = marginally significant.
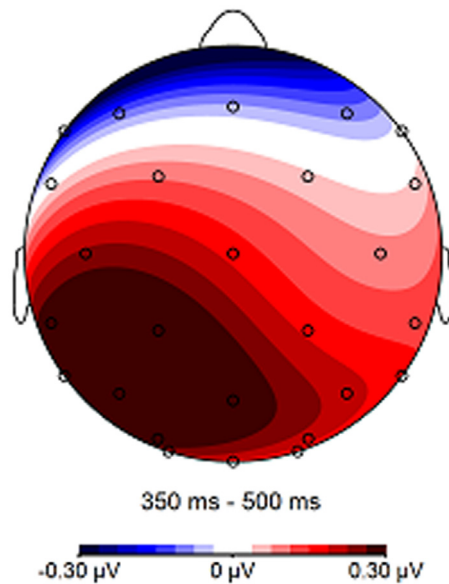
350 ms - 500 ms

-0.30 µV　　　0 µV　　　0.30 µV

Fig. 4. ERP results: Congruency effects in the P3b time window. Topographic map showing mean amplitude differences for the P3b component (350–500 ms time window) between Congruent and Incongruent conditions. Differences between conditions appeared in posterior electrodes.

Congruent and Incongruent conditions compared to the Unpredictable condition ($p = .040$, and $p = .009$, respectively).

Taken together, these results indicate that the presentation of un-predictable melodies elicited larger amplitudes in the N50 and N1 ERP components than the presentation of predictable melodies. Also, the presentation of predictable melodies elicited larger amplitudes in the P2 and P3a components than the presentation of unpredictable me-lodies. These effects were prominently observed in frontal and central regions of the scalp. Most importantly, predictable melodies followed by a vertically-incongruent visual target elicited larger P3b amplitudes in posterior regions of the scalp, as compared to a vertically-congruent visual target. No significant differences between conditions were ob-served in the latency analyses for any of the ERP components.

Additionally, and following the same logic as in the behavioural data analysis, we carried out several Repeated Measures ANOVAs for the amplitude and latency of the different ERP components, only in-cluding data from the first and the last 20 trials of each condition se-parately. These analyses were run in order to explore possible effects of experience and predictability of the melodies on the electro-physiological responses across the experiment. In these analyses we included the factors Condition (Congruent, Incongruent, and Unpredictable), Region (Frontal, Central, and Posterior), and Exposure (first vs. last 20 trials). Nevertheless, we were unable to obtain any significant main effect or interaction, probably because of an in-sufficient number of trials per condition.

4. Discussion

In the present study, we aimed at investigating whether the passive, but repeated, listening of melodies can modulate spatial attention, having a subsequent influence on perceptual processes (i.e., facilitating the following processing of visual stimuli appearing in a high or low position in space). Our behavioural results revealed that participants responded faster to congruent visual stimuli than to incongruent or unpredicted visual stimuli in the last 20 trials of the experiment, but not in the first 20 trials. These results suggest that recent experience with the melodies is of great importance in order to modulate visuospatial attention and generate stronger predictions. Relevantly, spatial atten-tion was modulated by the melodies even if they were not relevant for

the task at hand.

It is important to note that our results are congruent with previous literature on multisensory interactions. Thus, multisensory interactions between the information from different sensory modalities would constantly occur in the brain to analyse whether signals from different modalities are matching or in conflict (e.g., Kayser and Logothetis, 2007, 2009). Importantly, the outcome of these interactions would be particularly relevant for different cognitive processes, such as attention, perception, or controlling actions (e.g., Kayser and Logothetis, 2007). Also, multisensory interactions are thought to occur even when there is no meaningful relationship between the different sensory inputs, and seem to modulate early sensory processing (as discussed in Ghazanfar and Schroeder, 2006). Furthermore, and particularly relevant for in-terpreting our results, there is increasing electrophysiological (e.g., Giard and Peronnet, 1999; Molholm et al., 2002) and functional ima-ging evidence (e.g., Foxe et al., 2002) suggesting that multisensory interactions emerge at very early stages of the cortical processing hierarchy. More concretely, the circuitry between the thalamus, the thalamic reticular nucleus (TRN), and prefrontal cortical areas im-plicated in selective attention would be involved in the dynamic se-lection of the leading sensory modality that may be modulating the perception of accompanying stimuli in other modalities (Lakatos et al., 2009). Hence, attention would play a key role in determining the leading sense based on the relative salience of stimuli across modalities, and that would enable an efficient merge of the information from dif-ferent modalities at higher levels of processing (Lakatos et al., 2009). Also, some studies suggest that areas conventionally considered unim-odal, such as some areas of the auditory cortex along the superior temporal gyrus, can also exhibit multisensory interaction/integration features (e.g., Foxe et al., 2002). Therefore, these previous studies would suggest that information processed by the early auditory cortices may contribute to early multisensory interactions, and therefore pro-vide a possible explanation for our results. More specifically, the task-irrelevant rhythmic streams of sounds would be modulating visuospa-tial attention from very early stages of the cortical processing hierarchy (as we mention below when discussing the results concerning the ear-liest ERP components).

In line with the behavioural results of our study, the amplitude of the P3b in posterior areas was larger in incongruent trials than con-gruent trials. The P3b component is usually observed in posterior areas of the scalp, whereas the P3a component has a wider distribution (Brázdil et al., 2003, 2001; Conroy and Polich, 2007; Knight, 1996, 1984; Knight et al., 1995; Squire and Kandel, 1999). Thus, since the difference between congruent and incongruent trials only appeared during the later ERP time window, and only in posterior electrodes, we interpret these results as indicating the presence of a P3b component. Larger amplitudes on this time window have previously been linked to the processing of unexpected changes in the sensory modality (visual vs. auditory) of stimuli in a cueing paradigm (Sutton et al., 1965) and often indicate a 'surprise' response to unexpected visual (Courchesne et al., 1975) or auditory stimulus changes (Squires et al., 1975). Also, the processing of improbable task-relevant events usually increases the amplitude on this time window (Donchin, 1981). Therefore, our find-ings could easily be interpreted as the attentional system's reaction to a visual stimulus that appeared in an unpredicted spatial position. Bearing in mind that the 'spatial prediction' was induced by means of melodies with highly-predictable changes in pitch, our ERP results suggest that the passive listening of highly predictable music modulates spatial attention (mainly along the vertical axis), subsequently affecting the perceiver's expectancies regarding visual events. Notably, while in the behavioural data the differences between congruent and incon-gruent trials only appeared during the last trials of the experiment, this difference was present in the ERP data (P3b time window) in-dependently of the experimental block. Hence, even if the brain pro-cessed congruent and incongruent trials in a different way throughout the experiment, participants only showed a behavioural benefit at the

end of the study. Nevertheless, we acknowledge that future studies with a more fine-grained design should further explore what is the precise moment in time when congruent and incongruent stimuli produce different ERP patterns.

The conceptualization of the P3b component as reflecting processes of memory update also provides us with an adequate framework to explain the pattern of results found in the present study (Donchin et al., 1986; Polich, 2007). The P3b component has shown to capture the mechanisms underlying context-updating and subsequent memory processes (Brázdil et al., 2003, 2001; Conroy and Polich, 2007; Hartikainen and Knight, 2003; Knight, 1996; Kok, 2001; Polich, 2007, 2003; Squire and Kandel, 1999). The tones included in the melodies would have progressively generated a prediction as they were serially stored in memory. Therefore, the appearance of a spatially-incongruent visual stimulus would contradict this previous experience, perhaps requiring a representational update. Thus, in our experiment, the P3b effect observed for incongruent trials may reveal the difficulties of integrating the incoming information with the mismatching information already stored in memory (after repeated exposure to the melodies).

The fact that participants did not perform any spatial- or pitch-related task during the experiment, and, also, that they were not aware of any possible association between the auditory prediction and the location of the visual target, supports the idea that the spatial recoding of pitch can occur even when the listener is not engaged in any spatial or pitch-based task. This seems to be true even for non-musicians. Furthermore, the use of an orthogonal task (colour classification) allowed us to reduce possible confounds at a decisional level. Since the participants were not judging the visual stimuli in terms of their verticality, a strategic vertical remapping of the melodies could not give them any clue regarding the colour of the visual stimulus. Therefore, our results fit well into a view of crossmodal correspondence where an explicit, task-relevant processing of pitch or space is not needed for correspondence effects to happen, as it seems to be the case in pre-linguistic infants (Dolscheid et al., 2014; Fernández-Prieto et al., 2015; Lewkowicz and Minar, 2014; Walker et al., 2010).

The presence of a congruency effect in the visual P3b ERP component, but not in earlier components, favours the idea that pitch-based spatial prediction may occur at a relatively late stage of processing, and perhaps in associative areas of the brain where information from different sensory modalities can converge (Chiou and Rich, 2012). This hypothesis also fits with recent attempts to describe the representation of magnitudes (numbers, size, height, pitch, etc.) as being the result of common neural processes (Walsh, 2003). Further implications of this theory may be that these mechanisms occur in multisensory areas of the brain (e.g., in the right parietal cortex (Foster et al., 2013; Foster and Zatorre, 2010a, 2010b; Zatorre et al., 2009). The modulation of spatial attention during the perception of predictable music may easily explain why most (if not all) of the graphical representations of music across different cultures are (or have been, historically) based on the idea that high and low pitch corresponds to upper and lower positions, respectively (e.g., the stave in Western music). In fact, this crossmodal association appears to occur even in non-western cultures where the linguistic label used to describe pitch is not associated with verticality (Parkinson et al., 2012; Shayan et al., 2011).

Regarding the effects of auditory predictability on the processing of visual targets, we observed clearly distinctive behavioural and electrophysiological patterns for predictable and unpredictable melodies. More specifically, we observed that participants took more time to respond to visual stimuli following unpredictable than predictable melodies. This effect was particularly robust during the last trials of the experiment, suggesting that participants probably learnt to recognise predictable melodies after exposure to the different patterns. However, it is important to note that there were no behavioural differences between unpredictable and incongruent trials at the end of the experiment. This may be so because incongruities in a predictable melody would have caused response slowness similar to that caused by

unpredictable melodies.

In the present study, we observed larger amplitudes in the N50 and N1 ERP components for visual stimuli following unpredictable melodies (as compared to predictable melodies), and larger amplitudes in the P2 and P3a components for targets following predictable (as compared to unpredictable) melodies, at frontal and central regions of the scalp. It is important to note that we cannot discard that the difference between predictable and unpredictable conditions observed on the late components are a post-effect of the differences already observed on earlier time windows. Nevertheless, all of these components are supposed to index the recruitment of attentional resources in order to successfully carry out the task, and therefore this possibility should not affect our interpretation of the results. For instance, the N50 and N100 ERP components index selective effects of attention on visual processing; as an example, attention to targets in attended vs. unattended visual fields modulate the amplitude of the N1 component (e.g., Rugg et al., 1987). Also, the P2 component has been related to the use of contextual information to prepare for the visual analysis of upcoming stimuli (Federmeier and Kutas, 2002). Moreover, the P3a amplitude has been associated with the engagement of attention when orienting to changes in the environment (Polich, 2007, 2003). Hence, our results would suggest that predictable melodies made participants anticipate the spatial position of visual stimuli and perhaps pay more attention to them. Here, it is important to note that there were always 11 notes in the task-irrelevant auditory stream preceding the visual stimulus, and therefore the presentation of visual targets could have been anticipated by the participants. Nevertheless, anticipation only occurred when participants were presented with predictable melodies. Consequently, it was the content of the auditory sequences (i.e., whether they were predictable or not) what seemed to matter the most for anticipating the visual stimuli. Therefore, our results extend these anticipatory effects found in previous literature to the crossmodal case where predictability generated in one modality (audition) modulates the processing of information in another modality (vision).

To sum up, our results show that the passive listening of melodies with highly predictable 'ups and downs' in pitch modulates the response to visual stimuli presented in high or low spatial positions. Moreover, this modulatory effect seems to depend on the previous experience of the listener with the melodies. The most likely explanation of this effect implies that the acoustic 'ups and downs' can spontaneously modulate spatial attention (along the vertical axis), having a direct impact on the subsequent visuospatial processing of other stimuli.

Acknowledgements

References

Bernstein, I.H., Edelstein, B.A., 1971. Effects of some variations in auditory input upon visual choice reaction time. J. Exp. Psychol. 87, 241–247. http://dx.doi.org/10.1037/h0030524.

Besle, J., Schevon, C.A., Mehta, A.D., Lakatos, P., Goodman, R.R., McKhann, G.M., Emerson, R.G., Schroeder, C.E., 2011. Turning of the human neocortex to the temporal dynamics of attended events. J. Neurosci. 31, 3176–3185.

Brázdil, M., Rektor, I., Daniel, P., Dufek, M., Jurák, P., 2001. Intracerebral event-related potentials to subthreshold target stimuli. Clin. Neurophysiol. 112, 650–661.

Brázdil, M., Roman, R., Daniel, P., Rektor, I., 2003. Intracerebral somatosensory event-related potentials: effect of response type (button pressing versus mental counting) on P3-like potentials within the human brain. Clin. Neurophysiol. 114, 1489–1496. http://dx.doi.org/10.1016/S1388-2457(03)00135-4.

Cabrera, D., Morimoto, M., 2007. Influence of fundamental frequency and source elevation on the vertical localization of complex tones and complex tone pairs. J. Acoust.

Soc. Am. 122, 478–488. http://dx.doi.org/10.1121/1.2736782.

Chiou, R., Rich, A.N., 2012. Cross-modality correspondence between pitch and spatial location modulates attentional orienting. Percept. 41 (3), 339–353. http://dx.doi.org/10.1068/p7161.

Conroy, M.A., Polich, J., 2007. Normative variation of P3a and P3b from a large sample: gender, topography, and response time. J. Psychophysiol. 21, 22–32. http://dx.doi.org/10.1027/0269-8803.21.1.22.

Courchesne, E., Hillyard, S.A., Galambos, R., 1975. Stimulus novelty, task relevance and the visual evoked potential in man. Electroencephalogr. Clin. Neurophysiol. 39, 131–143. http://dx.doi.org/10.1016/0013-4694(75)90003-6.

Deroy, O., Fernández-Prieto, I., Navarra, J., Spence, C., in press. Unravelling the paradox of spatial pitch. In: Hubbard, T.L.(Ed.), Spatial Biases in Perception and Cognition. Cambridge University Press, Cambridge.

Dolscheid, S., Hunnius, S., Casasanto, D., Majid, A., 2014. Prelinguistic infants are sensitive to space-pitch associations found across cultures. Psychol. Sci. 1–6. http://dx.doi.org/10.1177/0956797614528521.

Donchin, E., 1981. Surprise!… surprise? Psychophysiol. 81 (5), 493–513.

Donchin, E., Kramer, A., Wickens, C., 1986. Applications of event-related brain potentials to problems in engineering psychology. Psychophysiol. Syst. Process. Appl. New Y. Guilford Press.

Federmeier, K.D., Kutas, M., 2002. Picture the difference: electrophysiological investigations of picture processing in the two cerebral hemispheres. Neuropsychologia 40, 730–747. http://dx.doi.org/10.1016/S0028-3932(01)00193-2.

Fellinger, R., Gruber, W., Zauner, A., Freunberger, R., Klimesch, W., 2012. Evoked traveling alpha waves predict visual-semantic categorization-speed. Neuroimage 59 (4), 3379–3388.

Fernández-Prieto, I., Navarra, J., Pons, F., 2015. How big is this sound? Crossmodal association between pitch and size in infants. Infant Behav. Dev. 38, 77–81. http://dx.doi.org/10.1016/j.infbeh.2014.12.008.

Fernández-Prieto, I., Spence, C., Pons, F., Navarra, J., 2017. Does language influence the vertical representation of auditory pitch and loudness? i-Percept. 8 (3). http://dx.doi.org/10.1177/2041669517716183.

Foster, N.E.V., Zatorre, R.J., 2010a. Cortical structure predicts success in performing musical transformation judgments. Neuroimage 53, 26–36. http://dx.doi.org/10.1016/j.neuroimage.2010.06.042.

Foster, N.E.V., Zatorre, R.J., 2010b. A role for the intraparietal sulcus in transforming musical pitch information. Cereb. Cortex 20, 1350–1359. http://dx.doi.org/10.1093/cercor/bhp199.

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. http://dx.doi.org/10.1016/j.neuroimage.2013.02.044.

Foxe, J.J., Wylie, G.R., Martinez, A., Schroeder, C.E., Javitt, D.C., Guilfoyle, D., Murray, M.M., 2002. Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J. Neurophysiol. 88 (1), 540–543.

Ghazanfar, A.A., Schroeder, C.E., 2006. Is neocortex essentially multisensory? Trends Cogn. Sci. 10 (6), 278–285.

Giard, M.H., Peronnet, F., 1999. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J. Cogn. Neurosci. 11 (5), 473–490.

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., Brattico, E., 2014. Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: a behavioral and ERP study. Front Hum. Neurosci. 8, 279. http://dx.doi.org/10.3389/fnhum.2014.00279.

Guo, C., Duan, L., Li, W., Paller, K.A., 2006. Distinguishing source memory and item memory: brain potentials at encoding and retrieval. Brain Res. 1118, 142–154. http://dx.doi.org/10.1016/j.brainres.2006.08.034.

Hartikainen, K.M., Knight, R.T., 2003. Lateral and orbital prefrontal cortex contributions to attention. In: Detection of Change. Springer, pp. 99–116.

Isreal, J.B., Chesney, G.L., Wickens, C.D., Donchin, E., 1980. P300 and tracking difficulty: evidence for multiple resources in dual-task performance. Psychophysiology 17 (3), 259–273.

Kayser, C., Logothetis, N.K., 2007. Do early sensory cortices integrate crossmodal information? Brain Struct. Funct. 212, 121–132.

Kayser, C., Logothetis, N.K., 2009. Directed interactions between auditory and superior temporal cortices and their role in sensory integration. Front. Integr. Neurosci. 3, 7.

Knight, R., 1996. Contribution of human hippocampal region to novelty detection. Nature 383, 256–259. http://dx.doi.org/10.1038/383256a0.

Knight, R.T., 1984. Decreased response to novel stimuli after prefrontal lesions in man. Electroencephalogr. Clin. Neurophysiol. Evoked Potentials 59, 9–20. http://dx.doi.org/10.1016/0168-5597(84)90016-9.

Knight, R.T., Grabowecky, M.F., Scabini, D., 1995. Role of human prefrontal cortex in attention control. Adv. Neurol. http://dx.doi.org/10.1212/WNL.45.10.1952-a.

Kok, A., 2001. On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology 38, 557–577. http://dx.doi.org/10.1016/S0167-8760(98)90168-4.

Kramer, A.F., Wickens, C.D., Donchin, E., 1985. Processing of stimulus properties: evidence for dual-task integrality. J. Exp. Psychol.: Human. Percept. Perform. 11 (4), 393.

Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of neural oscillations as a mechanism of attentional selection. Science 320, 110–113.

Lakatos, P., O'Connell, M.N., Barczak, A., Mills, A., Javitt, D.C., Schroeder, C.E., 2009.

The leading sense: supramodal control of neurophysiological context by attention. Neuron 64, 419–430.

Lewkowicz, D.J., Minar, N.J., 2014. Infants are not sensitive to synesthetic cross-modality correspondences: a comment on Walker et al. (2010). Psychol. Sci. 25, 832–834. http://dx.doi.org/10.1177/0956797613516011.

Lidji, P., Kolinsky, R., Lochy, A., Morais, J., 2007. Spatial associations for musical stimuli: a piano in the head? J. Exp. Psychol. Hum. Percept. Perform. 33, 1189–1207. http://dx.doi.org/10.1037/0096-1523.33.5.1189.

Luck, S.J., Hillyard, S. a., 1994. Electrophysiological correlates of feature analysis during visual search. Psychophysiology. http://dx.doi.org/10.1111/j.1469-8986.1994.tb02218.x.

Maeda, F., Kanai, R., Shimojo, S., 2004. Changing pitch induced visual motion illusion. Curr. Biol. 14, R990–R991. http://dx.doi.org/10.1016/j.cub.2004.11.018.

Mangun, G.R., Hillyard, S.A., 1991. Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. J. Exp. Psychol.: Human. Percept. Perform. 17, 1057–1074.

McEvoy, L.K., Pellouchoud, E., Smith, M.E., Gevins, A., 2001. Neurophysiological signals of working memory in normal aging. Cogn. Brain Res. 11, 363–376. http://dx.doi.org/10.1016/S0926-6410(01)00009-X.

Melara, R.D., Brien, T.P.O., 1987. Interact. Synesthetically Corresp. Dimens. 116, 323–336.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2002. Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cogn. brain Res. 14 (1), 115–128.

Mossbridge, J.A., Grabowecky, M., Suzuki, S., 2011. Changes in auditory frequency guide visual – spatial attention. Cognition 121, 133–139. http://dx.doi.org/10.1016/j.cognition.2011.06.003.

Navarra, J., Hartcher-O'Brien, J., Piazza, E., Spence, C., 2009. Adaptation to audiovisual asynchrony modulates the speeded detection of sound. Proc. Natl. Acad. Sci. Usa. 106, 9169–9173. http://dx.doi.org/10.1073/pnas.0810486106.

Németh, K., Kovács, P., Vakli, P., Kovács, G., Zimmer, M., 2014. Phase noise reveals early category-specific modulation of the event-related potentials. Front. psychol. 5, 367.

Nobre, A., Correa, A., Coull, J., 2007. The hazards of time. Curr. Opin. Neurobiol. 17, 465–470. http://dx.doi.org/10.1016/j.conb.2007.07.006.

Parkinson, C., Kohler, P.J., Sievers, B., Wheatley, T., 2012. Associations between auditory pitch and visual elevation do not depend on language: Evid. remote. popul. 41 (7), 854–861.

Polich, J., 1987. Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. Electroencephalogr. Clin. Neurophysiol./Evoked Potentials Sect. 68 (4), 311–320.

Polich, J., 2003. Detection of change: event-related potential and fMRI findings. Springer Science & Business Media.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148. http://dx.doi.org/10.1016/j.clinph.2007.04.019.

Rauscher, F.H., Shaw, G.L., Ky, K.N., 1993. Music and spatial task performance. Nature 365, 611.

Rugg, M.D., Doyle, M.C., 1992. Event-related potentials and recognition memory for low- and high-frequency words. J. Cogn. Neurosci. 4, 69–79. http://dx.doi.org/10.1162/jocn.1992.4.1.69.

Rugg, M.D., Milner, A.D., Lines, C.R., Phalp, R., 1987. Modulations of visual event-related potentials by spatial and non-spatial visual selective attention. Neuropsychologia 25, 85–96.

Rusconi, E., Kwan, B., Giordano, B.L., Umilta, C., Butterworth, B., 2006. Spatial representation of pitch height: the SMARC effect. Cognit. 99 (2), 113–129.

Shayan, S., Ozturk, O., Sicoli, M.A., 2011. The thickness of pitch: crossmodal metaphors in Farsi, Turkish, and Zapotec. Senses Soc. 6, 96–105. http://dx.doi.org/10.2752/174589311×12893982233911.

Sonnadara, R.R., Gonzalez, D.A., Hansen, S., Elliott, D., Lyons, J.L., 2009. Spatial properties of perceived pitch. N.Y. Acad. Sci. 1169 (1), 503–507. http://dx.doi.org/10.1111/j.1749-6632.2009.04858.x.

Spence, C., Deroy, O., 2013. How automatic are crossmodal correspondences? Conscious. Cogn. 22, 245–260. http://dx.doi.org/10.1016/j.concog.2012.12.006.

Squire, L.R., Kandel, E.R., 1999. From mind to molecules. Sci. Am. Libr. Free. NY.

Squires, N.K., Squires, K.C., Hillyard, S.A., 1975. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Electroencephalogr. Clin. Neurophysiol. 38, 387–401. http://dx.doi.org/10.1016/0013-4694(75)90263-1.

Stumpf, C., 1883. Tonpsychologie 1 Leipzig, Hirzel.

Sutton, S., Braren, M., Zubin, J., John, E.R., 1965. Evoked-potential correlates of stimulus uncertainty. Science 150, 1187 (80-.).

Walker, P., Bremner, J.G., Mason, U., Spring, J., Mattock, K., Slater, A., Johnson, S.P., 2010. Preverbal Infants' Sensitivity to Synaesthetic Crossmodality Correspondences. Psychol. Sci. 21, 21–25. http://dx.doi.org/10.1177/0956797609354734.

Walsh, V., 2003. A theory of magnitude: common cortical metrics of time, space and quantity. Trends Cogn. Sci. 7 (11), 483–488.

Wickens, C., Kramer, A., Vanasse, L., Donchin, E., 1983. Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. Science 221 (4615), 1080–1082.

Zatorre, R.J., Halpern, A.R., Bouffard, M., 2009. Mental reversal of imagined melodies: a role for the posterior parietal cortex. J. Cogn. Neurosci. 22, 775–789. http://dx.doi.org/10.1162/jocn.2009.21239.