Perception, 2015, volume 44, pages 169–182

doi:10.1068/p7699

# Volitional mechanisms mediate the cuing effect of pitch on attention orienting: The influences of perceptual difficulty and response pressure

Rocco Chiou[1,2]§, Anina N Rich[2,3]§
[1]The Neuroscience and Aphasia Research Unit (NARU), School of Psychological Sciences, University of Manchester, UK; [2] Perception in Action Research Centre and Department of Cognitive Science, Faculty of Human Sciences, Macquarie University, Australia; [3]ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Australia; e-mail: roccochiou@gmail.com, anina.rich@mq.edu.au
Received 25 January 2014, in revised form 26 December 2014

**Abstract.** Our cognitive system tends to link auditory pitch with spatial location in a specific manner (ie high-pitched sounds are usually associated with an upper location, and low sounds are associated with a lower location). Recent studies have demonstrated that this cross-modality association biases the allocation of visual attention and affects performance despite the auditory stimuli being irrelevant to the behavioural task. There is, however, a discrepancy between studies in their interpretation of the underlying mechanisms. Whereas we have previously claimed that the pitch-location mapping is mediated by volitional shifts of attention (Chiou & Rich, 2012, *Perception*, **41**, 339–353), other researchers suggest that this cross-modal effect reflects automatic shifts of attention (Mossbridge, Grabowecky, & Suzuki, 2011, *Cognition*, **121**, 133–139). Here we report a series of three experiments examining the effects of perceptual and response-related pressure on the ability of nonpredictive pitch to bias visual attention. We compare it with two control cues: a predictive pitch that triggers voluntary attention shifts and a salient peripheral flash that evokes involuntary shifts. The results show that the effect of nonpredictive pitch is abolished by pressure at either perceptual or response levels. By contrast, the effects of the two control cues remain significant, demonstrating the robustness of informative and perceptually salient stimuli in directing attention. This distinction suggests that, in contexts of high perceptual demand and response pressure, cognitive resources are primarily engaged by the task-relevant stimuli, which effectively prevents uninformative pitch from orienting attention to its cross-modally associated location. These findings are consistent with the hypothesis that the link between pitch and location affects attentional deployment via volitional rather than automatic mechanisms.

**Keywords:** cross-modality mapping, multisensory, automaticity, voluntary attention, involuntary attention

## 1 Introduction

Despite the constant stream of sensory input that far exceeds our cognitive processing capacity, the human brain seems adept at selecting information relevant to our behavioural goals and filtering out irrelevant signals. Attention is the mechanism that performs this remarkable feat. In the visual domain attention converts ‘looking’ into ‘seeing’, allowing us to prioritise various locations and objects in a visual scene. Owing to its significance, the processes underlying this crucial cognitive function have been intensively studied (eg Carrasco & Yeshurun, 2009; Prinzmetal & Landau, 2008).

There is considerable evidence that there are two distinct types of attentional systems: an involuntary mechanism that triggers automatic orienting in response to salient external stimuli, and a voluntary one that directs attention wilfully. A sudden salient stimulus (eg a flash) captures attention involuntarily, which in turn can affect detection or discrimination of visual targets

§ Correspondence should be addressed to either author at Department of Cognitive Science, Macquarie University, Sydney, NSW 2019, Australia. Please note Dr Chiou’s new address.

(Posner, 1980). A variety of other stimuli also orient attention although they carry no information about target location, such as arrows (Jonides, 1981), eye-gaze (Ristic, Friesen, & Kingstone, 2002), numerals (Fischer, Castel, Dodd, & Pratt, 2003), and auditory pitch (Chiou & Rich, 2012). The capability of eliciting attention shifts despite them being nonpredictive is consistent with involuntary attention. However, they differ fundamentally from the typical flash-type stimuli in how they summon attention. Whereas flashes seize attention through 'bottom-up' perceptual saliency, other stimuli rely upon an observer shifting attention volitionally, reflecting more 'top-down' control. The reliance on top-down processes is consistent with what we know about voluntary attention, although some stimuli, such as arrows and gazes, have been demonstrated to be less voluntary than other cues known to elicit entirely wilful shifts of attention (see Ristic & Kingstone, 2006). Studying the extent to which different cues engage the two systems of attention informs us about the mechanisms whereby the human brain directs its attentional focus based on internal and external information.

We recently demonstrated that spatially nonpredictive auditory pitch modulates attention orienting in a visual detection task (Chiou & Rich, 2012): high tones biased attention upwards, whereas low tones biased attention downwards. Similar effects of pitch were observed in another study using a visual matching task (Mossbridge et al., 2011): ascending tones guided attention to an upper location, whereas descending tones guided it to a lower location, which helped discern colour patches (also see Evans & Treisman, 2010). Mossbridge et al. (2011) argue that the influence of pitch was automatic rather than volitional because the sounds were spatially uninformative of the upcoming target, and their participants were instructed to ignore the sounds. In our study, however, we observed psychophysical traits that were more consistent with an effect on voluntary attention. First, sounds of the same pitch could direct attention to opposite locations, depending on whether the stimuli were interpreted as high or low tones (based on context). Second, the pitch effect could be overridden by top-down control, unlike the impact of salient stimuli that robustly resists volitional suppression. Thus, it seems that the pitch-induced cuing has features of both voluntary and involuntary attentional effects.

Here we present a more direct examination of the psychophysical characteristics of the pitch effect to clarify the roles of voluntary and involuntary mechanisms. In addition to a better understanding of the attentional system, such investigations may also shed light on the automaticity of cross-modality correspondences (eg the reliable tendency to associate high-pitched sounds with visually higher, smaller, and brighter stimuli; see Spence, 2011).

We report the results from three experiments that test how pressure on perceptual and response-related processes modulates the pitch effect on attention. We compared nonpredictive pitch (ie the pitch–location mapping) with two types of control cues: (1) predictive pitch that drives voluntary attention shifts due to its relevance to the behavioural goal; and (2) nonpredictive flashes that evoke involuntary shifts due to high salience. Although there is evidence that increasing perceptual/cognitive load attenuates the size of involuntary cuing effects, most studies find that such cuing remains robust despite load (Spence, 2010). This design allowed us to assess the impact of salience and task-relevance by comparing the nonpredictive pitch cue with cues known to involve purely voluntary and involuntary attention. If the pitch effect was automatic, it would persist despite tones being uninformative and task demands occupying cognitive resources.

To preempt the results, we found that there was no effect of nonpredictive pitch for tasks requiring either easy or hard visual discrimination. By contrast, the effects of predictive pitch and nonpredictive flash remained robust. This is consistent with attentional biasing by uninformative pitch being mediated by voluntary rather than involuntary attention. Only when pitch is predictive of location does this information influence attention under conditions of challenging visual discrimination.

## 2 Experiment 1

We modified the design of Prinzmetal, Zvinyatskovskiy, Gutierrez, and Dilem (2009) to explore the effect of perceptual difficulty on attentional cuing by pitch. In their study predictive cues had a greater cuing effect in a perceptually difficult than an easy context (for similar findings see also Dufour, 1999). With nonpredictive cues, Prinzmetal et al. (2009) found their cuing effect persisted and was greater in an easy than in a hard context. In our experiment we presented an auditory cue either high or low in pitch, followed by a visual target in either an upper or lower location. The task required visual discrimination of letter targets. We manipulated perceptual difficulty by using stimuli that either differed by simple visual features (easy) or by conjunctions of features (hard). There were two types of cues: for one group of participants pitch did not predict target location and was task-irrelevant; for the other (control) group it was predictive and prompted voluntary shifts of attention to a probable target location. This comparison enabled us to explore the effects of task-relevance in contexts of higher versus lower perceptual difficulty.

Increasing perceptual difficulty should reduce the available resources for processing the pitch cue (especially when it carries no task-relevant message). If nonpredictive pitch drives automatic attention shifts, it would result in a cuing effect that perseveres even under a condition in which a visually more complex target engages most cognitive resources, leaving no room for volitional processing of the cue. We also expected that the results of predictive pitch (the control group) should replicate Prinzmetal et al. (2009) such that the cuing effect would be larger in the hard than easy condition.

### 2.1 *Methods*

2.1.1 *Participants.* Twenty undergraduates (age range: 20–25 years) at Macquarie University participated for course credit or monetary reward. Half of them (five females) were in the nonpredictive group and the other half (seven females) were in the predictive group. All of them reported normal (or corrected-to-normal) vision and hearing. All gave consent before taking part in the research. The study was reviewed and approved by the Macquarie University Research Ethics Committee.

2.1.2 *Apparatus, stimuli, and design.* A Pentium III computer was used for stimulus presentation and response collection. Stimuli were displayed on a 17-inch CRT monitor. The experimental procedure was controlled by Matlab 7.5 with Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The participants sat at a viewing distance of ~57 cm with a chin-rest to stabilise head position.

The trial procedure is shown in figure 1. The stimuli appeared in white against a black background. Each trial began with a fixation display containing a central cross (1 deg × 1 deg) and two placeholders. The two square placeholders subtended 3.5 deg and were positioned 5.5 deg above and below the central cross. After 500 ms, a sinusoidal tone (cue) of either high (2500 Hz) or low (150 Hz) pitch was presented for 50 ms, while the fixation display remained on the screen. Cues were presented at a comfortable hearing level and came from two loudspeakers to the left and right of the screen. After a cue–target onset asynchrony (SOA) of 350 ms, the visual target appeared in one of the two placeholders for 150 ms. This SOA was based on previously reported data (experiment 4a, Chiou & Rich, 2012), which revealed that the pitch cueing effect did not emerge until 350 ms after the onset of a tone (see supplementary materials at http://dx.doi.org/10.1068/p7699 for additional details). The target was composed of a target letter (F or T) flanked by two distractor letters. In the low difficulty (feature) condition, the target letter was flanked by the letter O (eg OFO). In the high difficulty (conjunction) condition it was flanked by the letter H (eg HFH). Participants responded whether the display contained F or T by pressing one of two designated keys.

The display (without the target) remained visible until a response was made or 2.5 s had elapsed. Feedback regarding accuracy was presented during the 750 ms intertrial interval. On congruent trials a high tone preceded an upper target and a low tone preceded a lower target. On incongruent trials it was the reverse mapping. The two congruency conditions were randomly interleaved within a block. Participants were required to fixate at the central cross and respond as quickly and accurately as possible.

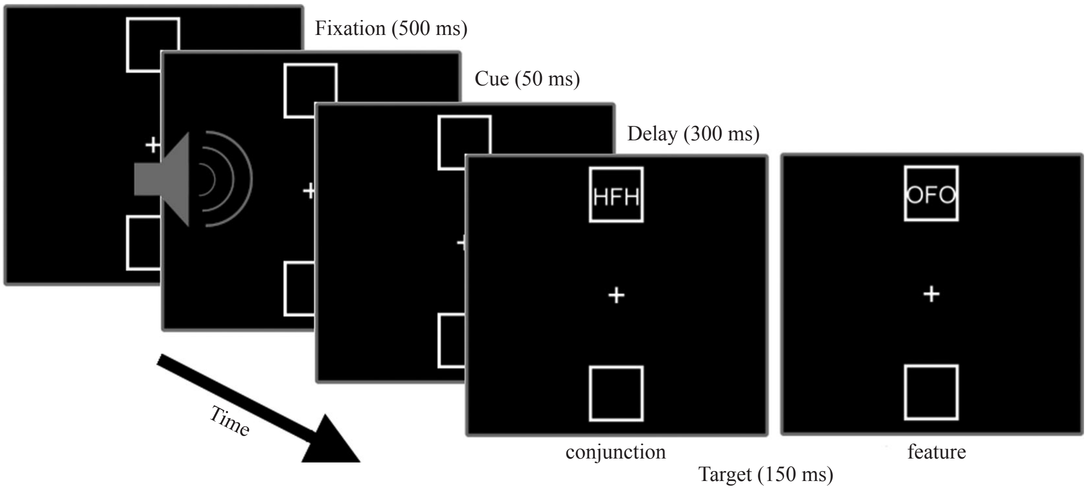


**Figure 1.** Sample stimuli and time course of events in a trial in experiment 1. The auditory cues were nonpredictive of target location for one group of participants (50% congruent and 50% incongruent) and predictive for the other group (80% congruent and 20% incongruent). The task required discriminating whether the stimuli contained F or T. In the conjunction condition the stimuli were either HFH or HTH. In the feature condition the stimuli were either OFO or OTO. Not to scale.

2.1.3 *Procedure.* Similar to the procedure of Prinzmetal et al. (2009), each participant began with two practice blocks of 40 trials, containing only the feature-type stimuli with a target display duration of 150 ms. The ‘feature’ practice blocks were followed by two practice blocks with only conjunction targets. In the first ‘conjunction’ block target duration was initially set at 360 ms for the first 20 trials and shortened to 240 ms in the remaining 20 trials. This was then followed by a complete block of 40 practice trials with 150 ms duration.

Data were collected from ten experimental blocks of 40 trials, with feature and conjunction trials separated into different blocks (5 blocks of each). The order of feature and conjunction blocks was counterbalanced across participants such that half of participants completed five blocks of feature trials prior to five blocks of conjunction trials; the other half of participants had the inverse order. For participants in the nonpredictive group there was an equal proportion (50 : 50) of congruent and incongruent trials in a block. This gave 100 trials per condition. For those in the predictive group, 80% of the trials in a block were congruent and the remaining 20% were incongruent. This gave 160 congruent trials and 40 incongruent trials in total. This ratio of trials was the same in the feature and conjunction blocks. Participants were informed of the relative proportions and hence whether the auditory cue was predictive or nonpredictive.

## 2.2 *Results and discussion*

After excluding errors (nonpredictive: 11%; predictive: 12.8%) and reaction time (RT) outliers (RTs > 3 SD above the condition mean; nonpredictive: 0.1%; predictive: 1%), we analysed the mean correct RTs using a mixed design ANOVA with perceptual difficulty (feature vs conjunction) and congruency (congruent vs incongruent) as *within*-participant

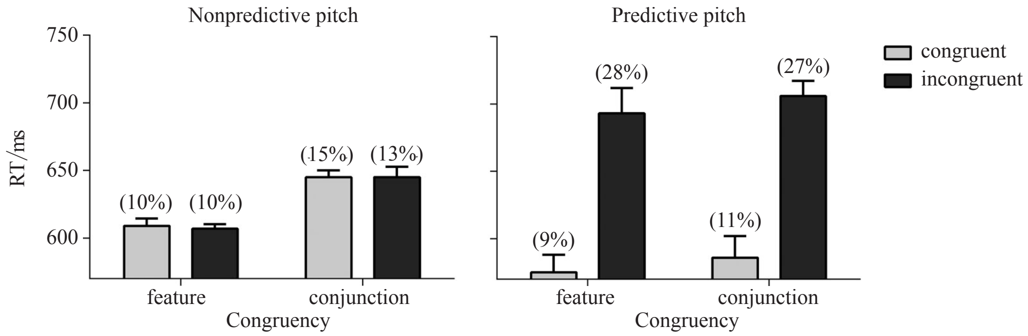


**Figure 2.** The mean correct reaction times (RTs) for each condition in experiment 1. Predictive and nonpredictive pitch data are from separate groups of participants. Error bars represent one repeated-measure standard error of the mean. Mean error rates (% incorrect) are presented in parentheses above each condition's RT bar.

factors and predictability (predictive vs nonpredictive) as a *between*-participant factor. The same analysis was also conducted on the mean error rates. The mean RTs and error rates of all conditions are reported in figure 2.

Analyses of RTs showed significant main effects of perceptual difficulty ($F_{1,18} = 18.51$, $p < 0.001$, $\eta^2 = 0.50$) and congruency ($F_{1,18} = 15.78$, $p = 0.001$, $\eta^2 = 0.46$) but no between-group main effect of predictability ($p = 0.62$). The congruency × predictability interaction was significant ($F_{1,18} = 16.35$, $p = 0.001$, $\eta^2 = 0.47$). A posteriori comparisons by group revealed that RTs in the incongruent condition were slower than those in the congruent condition for the predictive group ($p < 0.001$), but *not* for the nonpredictive group ($p = 0.96$). The perceptual difficulty × predictability interaction was also significant ($F_{1,18} = 4.60$, $p = 0.04$, $\eta^2 = 0.20$). A posteriori comparisons showed that RTs in the conjunction condition were slower than those in the feature condition only for the nonpredictive group ($p < 0.001$), not for the predictive group ($p = 0.14$). No other statistics were significant (all $p$s > 0.85).

Analyses of error rates were basically consistent with the pattern of RTs: the main effects of predictability ($F_{1,18} = 6.96$, $p = 0.01$, $\eta^2 = 0.27$) and congruency ($F_{1,18} = 22.97$, $p < 0.001$, $\eta^2 = 0.56$) were significant, but there was no main effect of perceptual difficulty ($F_{1,18} = 2.96$, $p = 0.10$, $\eta^2 = 0.14$). The perceptual difficulty × predictability interaction showed a weak trend ($F_{1,18} = 3.13$, $p = 0.09$, $\eta^2 = 0.14$). The congruency × predictability interaction was significant ($F_{1,18} = 31.70$, $p < 0.001$, $\eta^2 = 0.63$). A posteriori comparisons by group revealed that error rates in the incongruent condition were greater than those in the congruent condition only for the predictive group ($p < 0.001$), not for the nonpredictive group ($p = 0.56$). No other statistics were significant (all $p$s > 0.10).

The major finding was that task-relevance potently affected the presence of the cuing effect. Whereas predictive pitch caused a significant congruency effect, nonpredictive pitch did not orient attention and modulate performance for either the easy or hard discrimination task. We also observed that, although perceptual difficulty affected RT overall in the nonpredictive group, there was no evidence that it modulated the size of the congruency effect in either group. We speculate that this pattern is due to the task difficulty being set at an inappropriate level. Our reasoning is elaborated separately for the two types of cues below.

We failed to replicate the cuing effect of nonpredictive pitch that we previously obtained with a simple detection task (Chiou & Rich, 2012). This might be due to perceptual difficulty being too high, even in the feature case, for the irrelevant sound to be processed. A major difference between visual detection and discrimination is that the latter requires processing finer-grained perceptual details and actively maintaining the stimulus-response mapping rules in memory during the task. Thus, the present discrimination task may be too demanding (even in the feature condition), leaving no spare resources for processing the task-irrelevant sounds. If task difficulty makes a difference, then reducing the overall perceptual demand might release some resources for processing of the task-irrelevant tones.

We also failed to replicate the findings of Prinzmetal et al. (2009) in our predictive cuing condition: perceptual difficulty did not modulate the magnitude of the cuing effect. Although one difference in design is that Prinzmetal et al. used predictive flash cues whereas we used predictive tones, it seems more likely that the differences in results reflect the load of the task. Specifically, Prinzmetal et al. presented distractors at nonpredicted locations, resulting in overall more demanding contexts for visual processing than our less cluttered displays. These distractors might exert a greater influence in the conjunction than the feature condition, thus causing a modulation of the cuing effect. If the lack of modulatory effect was due to task being not challenging enough, then increasing the difficulty by presenting "distractors at an unexpected location" might be helpful in replicating their effect.

In the subsequent experiment the issue of perceptual difficulty was tackled separately for predictive and nonpredictive cues. We reduced overall difficulty for nonpredictive pitch and increased difficulty for predictive pitch.

## 3 Experiment 2a and 2b

We carried out two experiments to further test the modulatory effect of perceptual difficulty on predictive and nonpredictive pitch cues. In experiment 2a we presented nonpredictive cues and removed the flanking distractors from the visual targets to reduce perceptual load. We tested whether the removal of distractors would free up enough volitional resources to reveal the biasing effect of pitch on attention orienting. In experiment 2b we used predictive cues and added distractors in the noncued/unexpected location (ie opposite to the one suggested by the cue) to increase perceptual load. The distractors of the conjunction condition were visually more complex than those of the feature condition. We tested whether the inclusion of distractors at the uncued location would reveal a modulation of perceptual difficulty on the size of attention cuing effect.

### 3.1 *Methods*

3.1.1 *Participants.* Fourteen undergraduates (nine females; age range: 20–25 years) participated in experiment 2a (conducted first), and ten (five females; age range: 20–25 years) participated in 2b (as the aim of 2b was to replicate the predictive cue condition of experiment 1 of Prinzmetal et al., 2009, we used a similar sample size to that experiment, which had twelve participants). All of them reported normal (or corrected-to-normal) vision and hearing. All gave consent before taking part in the research and received payment for their participation.

3.1.2 *Design.* The apparatus, time frame, and procedure of experiment 2a were identical to those of the *nonpredictive* group of experiment 1, with the following modifications: we removed the flanking letters in the target displays and used letters O and X in the feature condition and letters T or L in the conjunction condition. Participants did one practice block of 40 trials, followed by ten experimental blocks of 40 trials (5 feature and 5 conjunction blocks presented in counterbalanced order), giving 100 trials per condition. Participants were informed that sounds did not predict target location.

The design of experiment 2b was the same as that of the *predictive* group of experiment 1, with the following changes: we presented distractors OOO (feature condition) or HHH (conjunction condition) at the nonpredicted location (ie the placeholder opposite to the one suggested by the pitch cue). The targets were OFO or OTO in the feature condition and HFH or HTH in the conjunction condition. The task required discrimination between F and T.

Participants did two practice blocks of 40 trials, followed by ten experimental blocks of 40 trials (5 feature and 5 conjunction blocks; counterbalanced). Participants were informed of the proportion of trials (congruent: 80%; incongruent: 20%) and were encouraged to use the pitch information to improve performance.

### 3.2 *Results and discussion*

We excluded errors (2a: 3%; 2b: 16%) and RT outliers (2a: 0.3%; 2b: 0.8%) prior to analysis of the RT data. We analysed the mean correct RTs of experiment 2a and 2b separately using repeated-measure ANOVAs with perceptual difficulty (feature vs conjunction) and congruency (congruent vs incongruent) as within-participant factors. The same analysis was also conducted on the mean error rates.

The mean RTs and error rates of experiment 2a are illustrated in figure 3a. The results showed a significant main effect of perceptual difficulty, with conjunction RTs being slower than feature RTs ($F_{1,13} = 4.98$, $p = 0.04$, $\eta^2 = 0.27$). There was a very weak trend for the main effect of congruency ($F_{1,13} = 3.31$, $p = 0.09$, $\eta^2 = 0.20$) and no significant interaction between perceptual difficulty and congruency ($F_{1,13} = 1.24$, $p = 0.28$, $\eta^2 = 0.08$). The simple main effect of congruency for each individual is shown in figure 3b. Although nine of the individuals (64%, red dashed lines) had numerical differences in the direction consistent with a cuing effect (note mostly extremely small; 4 of the 9 showing a difference smaller than 10 ms), five participants (36%, green solid lines) showed the inverse pattern or no difference. Due to the substantial proportion of participants showing no effect and the negligibly small difference, we conclude that there is really very little evidence of a cuing effect of nonpredictive pitch. Analogous analyses of error rates showed no significant effects (all $p$s > 0.14).

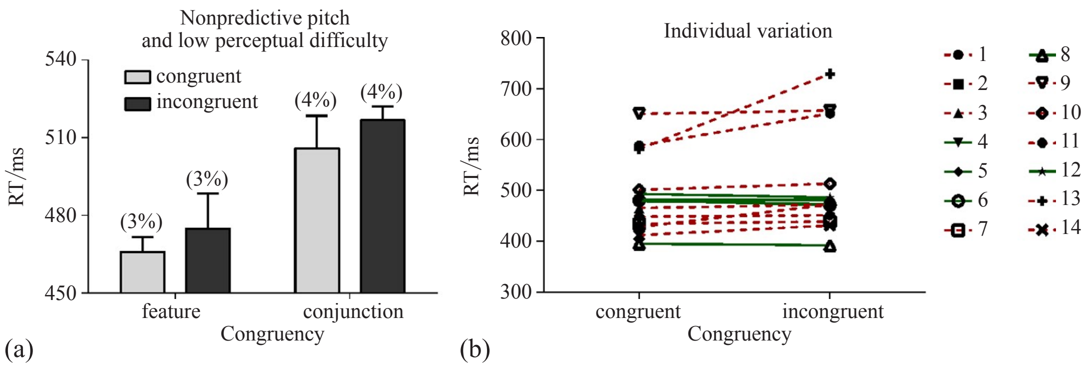


**Figure 3.** [In colour online, see http://dx.doi.org/10.1068/p7699] (a) The mean correct reaction times (RTs) and error rates for each condition in experiment 2a. Error bars represent one repeated-measure standard error of the mean. Mean error rates (% incorrect) are presented in parentheses above each condition's RT bar. (b) The simple main effect of congruency (as there were no significant interactions involving congruency), expressed as the individual congruent and incongruent RTs of each participant. A dashed line (red online) represents a faster congruent RT than an incongruent one. A solid line (green online) represents the reverse pattern or no difference between congruent and incongruent RTs.

The mean correct RTs and error rates of each condition in experiment 2b are reported in figure 4a. Analysis of RTs showed significant main effects of perceptual difficulty ($F_{1,9} = 11.22$, $p = 0.009$, $\eta^2 = 0.55$) and congruency ($F_{1,9} = 36.72$, $p < 0.001$, $\eta^2 = 0.80$). Crucially, there was a significant perceptual difficulty × congruency interaction ($F_{1,9} = 5.60$, $p = 0.04$, $\eta^2 = 0.38$; illustrated in figure 4b). A posteriori comparisons revealed that the incongruent RTs were significantly slower than congruent ones at both levels of perceptual difficulty (both $p$s < 0.05). However, the interaction suggests that it was significantly larger in the conjunction (262 ms) than in the feature condition (182 ms). Analysis of the error rates was consistent with the pattern of RTs. There was a significant main effect of congruency

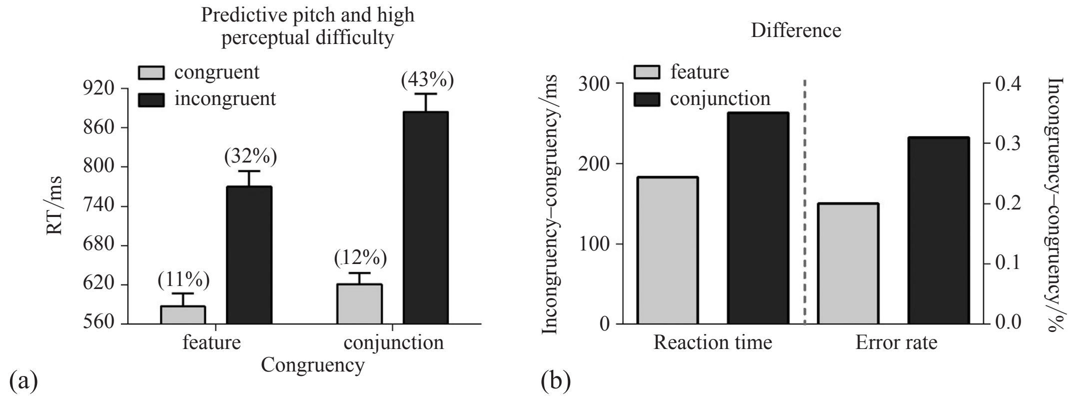


**Figure 4.** (a) The mean reaction time (RT) of all conditions in experiment 2b. Mean error rates (% incorrect) are presented in parentheses above each condition's RT bar. (b) The size of the congruency effect in RT and error rate of experiment 2b, expressed as the mean RT difference between the congruent and incongruent condition.

($F_{1,9} = 56.48$, $p < 0.001$, $\eta^2 = 0.86$) and a trend for a main effect of perceptual difficulty ($F_{1,9} = 4.18$, $p = 0.07$, $\eta^2 = 0.31$). There was also a strong trend for the perceptual difficulty × congruency interaction ($F_{1,9} = 4.30$, $p = 0.06$, $\eta^2 = 0.32$; figure 4b). Further tests showed that error rates were significantly higher in the incongruent than congruent condition in both difficulty levels (both *p*s < 0.05), with the difference being greater in the conjunction (32%) than the feature condition (20%).

Experiment 2a did not reveal a significant effect of pitch. There was a very slight hint of a possible congruency effect ($p = 0.09$), but the considerable individual variability suggested any such effect was extremely weak. Overall, the data suggest that the congruency effect is *not* robust for visual discrimination tasks, unlike our previous results with simple detection (Chiou & Rich, 2012). Despite reducing perceptual load (ie the removal of flankers), there is still a fundamental difference in cognitive load of visual discrimination versus detection. A discrimination task requires participants to hold the stimuli-to-response mapping in working memory, whereas the far simpler detection task requires an identical repetitive response to every target (like our design, Mossbridge et al., 2011, showed a pitch cuing effect also using identical button-press to every target). The need to maintain the task rules in mind may reduce the available resources for processing the task-irrelevant pitch, which in turn greatly attenuates or completely eliminates the congruency effect.

Experiment 2b showed that the effect of informative pitch was significantly larger in the hard than easy condition, replicating the RT and error rate results of Prinzmetal et al. (2009). This pattern suggests that distractors (at the noncued/unexpected location) play a key role in mediating the modulation of perceptual difficulty on the cuing effect of voluntary attention. Specifically, when voluntary attention was misled by the predictive cues on incongruent trials, participants needed to reorient their attention to the other location. Reorientation was more difficult in the conjunction than in the feature condition because the conjunction distractors (HHH) engaged attention for longer than did the feature distractors (OOO). It might take longer for participants to discover that a distractor stimulus did not contain the target (due to similarity in feature composition), relative to a feature stimulus. As a result, the cuing effect of voluntary attention was larger when perceptual difficulty was high.

Taken together, the two experiments demonstrate that perceptual load modulates the attentional effects of predictive and nonpredictive pitch differently. Whereas the effect of predictive (task-relevant) pitch was robust to the load and even had a greater impact in a high-load condition, there was no significant effect of nonpredictive (task-irrelevant) pitch when the task was a relatively more demanding visual discrimination (compared with the robust effect found with simple visual detection; see Chiou & Rich, 2012, and supplementary materials). This is consistent with a prominent theory of attentional selection that processing of task-irrelevant stimuli is lessened in conditions of high perceptual load (Lavie, 2005). It also concurs with the interpretation that the effect of task-irrelevant tones on attention does not reflect automatic shifts of attention.

In experiment 3 we approached this automaticity issue of the pitch effect from a different angle—by moving the load to a response execution level. Furthermore, we compared the pitch cue with a peripheral flash cue, which has been previously demonstrated to be highly robust despite response pressure (Prinzmetal, Ha, & Khani, 2010).

## 4 Experiment 3

To explore the effect of speed pressure on attention, we adopted a similar design to that of Prinzmetal and colleagues (Prinzmetal et al., 2010). These authors used a variant of the attention-cuing paradigm that contained a no/no-go procedure, peripheral flashes as cues, and highly discriminable visual targets. Participants discriminated targets on go trials and had to withhold a response on no-go trials. They made the task challenging by imposing a 'deadline' of RT: participants were required to make a visual discrimination within 150 ms. The results showed a significant cuing effect, suggesting that the impact of involuntary attention was robust to speed pressure. In addition, the authors found that false alarm (FA) rates were significantly higher in the congruent/cued location than in the incongruent/noncued location. They interpreted this as involuntary attention biasing the selection of the to-be-attended location but not affecting perceptual representation *per se*.

The speed pressure presumably increases the demand on cognitive resources, in this case at a level of response execution. If the pitch-induced cuing effect is automatic, in the same way as that induced by a peripheral flash, the pitch would induce similar effects to those of flashes. Namely, it would cause a cuing effect and give higher FA rates on the congruent location. If, in contrast, pitch cuing relies on volitional mechanisms, speed pressure may eliminate all effects. We also used the ubiquitous involuntary cuing effect of a peripheral flash as a baseline for comparison.

### 4.1 *Methods*

4.1.1 *Participants.* Fourteen undergraduates (five females; age range: 20–25 years) participated in experiment 3. All of them reported normal (or corrected-to-normal) vision and hearing. All gave consent before taking part in the research and received payment for their participation.

4.1.2 *Design.* The apparatus and procedure of experiment 3 were identical to those of experiment 2a, with the following modifications: the task was a combination of an attention cuing paradigm and a go/no-go procedure. On 75% of the trials one of the two easily discernible letters (O or X) was presented, and participants had to discriminate the letter by pressing a designated button. On 25% of the trials a nontarget letter (I) was present, and participants had to withhold their response. We introduced speed pressure by requiring participants to make a response within 300 ms.[(1)] If they failed to react within the time limit, the computer presented the words 'TOO SLOW' in the 750 ms intertrial interval. This was the only feedback given; there was no feedback about accuracy. In the pitch condition the cue was either a high or a low tone. In the flash condition the cue was a salient flash (a thickening of the frame) at the upper or lower placeholder. Both types of cues were 50 ms

[(1)] Initially we intended to use the same time limit (150 ms) as Prinzmetal et al. (2010). However, the results of a pilot experiment showed that participants tended to respond nearly randomly under such a high speed pressure. We therefore adopted a time limit of 300 ms.

in duration and uninformative. On the basis of previous research, we used cue–target onset asynchronies suitable for each cue: 350 ms in the pitch condition (identical to our previous experiments; see supplementary materials) and 100 ms in the flash condition (Posner, 1980). Note that the time-course of the two cuing effects does not overlap: the involuntary cuing of flashes arises within 100 ms and rapidly dissipates as SOA increases (Posner, 1980), whereas the earliest onset of the pitch cuing is ~350 ms (Chiou & Rich, 2012). On the basis of these known time courses, and the practical consideration of the length of the experiment doubling if we included both SOAs for each cue type, we used distinct SOAs for each cue to gauge attention cuing.(2) The congruency of the pitch condition was defined identically to previous experiments. In the flash condition the congruent trials contained an upper flash followed by an upper target and a lower flash followed by a lower target whereas the incongruent trials had the inverse mappings. The two types of cues were presented in separate blocks. Participants did two practice blocks of 64 trials, followed by twelve experimental blocks of 64 trials (6 pitch and 6 flash blocks; counterbalanced). Prior to the experimental blocks, participants were informed that neither of the two types of cues carried information about probable location of targets.

### 4.2 *Results and discussion*

We excluded errors (20%) and RT outliers (0.1%) prior to analyses on RTs. The mean correct RTs were analysed using a repeated-measure ANOVA with cue (pitch vs flash) and congruency (congruent vs incongruent) as within-participant factors. The same analysis was conducted on the mean error rates and mean FA rates (ie responses in the no-go trials).

Figure 5a illustrates the mean RTs and error rates of all conditions. The analyses of RTs showed a significant main effect of congruency ($F_{1,13} = 10.05$, $p = 0.007$, $\eta^2 = 0.43$) and no main effect of cue ($F < 1$, $p > 0.77$). Importantly, there was a significant cue × congruency interaction ($F_{1,13} = 12.91$, $p = 0.003$, $\eta^2 = 0.49$). A posteriori comparisons showed that the congruency effect was significant only in the flash condition ($p < 0.001$), not in the pitch condition ($p = 0.97$). The results of mean error rates only showed a trend of cue ($F_{1,13} = 3.88$, $p = 0.07$, $\eta^2 = 0.23$). There were no other significant effects on the error rates (both $p$s > 0.20).

Figure 5b illustrates the mean FA rates of all conditions. The analyses of mean FA rates showed significant main effects of cue ($F_{1,13} = 6.47$, $p = 0.02$, $\eta^2 = 0.33$) and congruency ($F_{1,13} = 12.60$, $p = 0.004$, $\eta^2 = 0.49$). The cue × congruency interaction was not significant ($F_{1,13} = 1.26$, $p = 0.28$). The interaction failed to reach significance because the pattern was

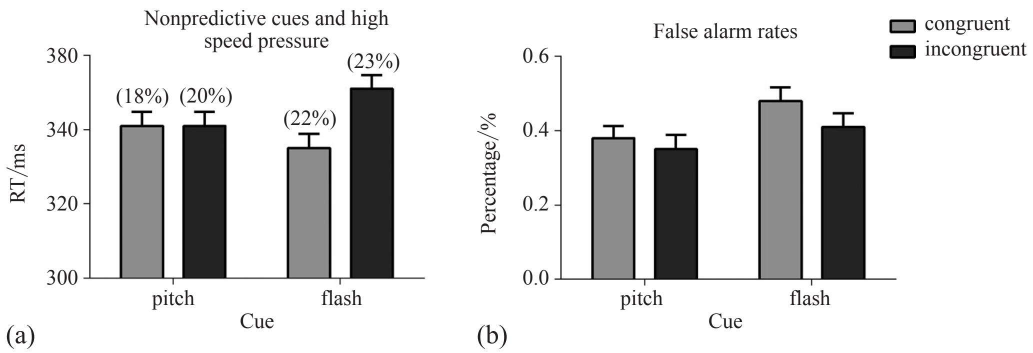


**Figure 5.** (a) The mean correct reaction times (RTs) and error rates for all conditions in experiment 3. (b) The false alarm rates for all conditions in experiment 3. Error bars represent one repeated-measure standard error of the mean. Mean error rates (% incorrect) are presented in parentheses above each condition's RT bar.

(2) We report an additional analysis examining the cuing effects of pitch and flash cues separately in supplementary materials.

similar in the pitch and flash conditions. As mentioned, Prinzmetal et al. (2010) used flashes as cues and found that FA rates were significantly higher in the congruent location than in the incongruent location. While acknowledging that we do not have sufficient statistical power to draw out a significant interaction, we checked to see if we replicated the result of Prinzmetal et al. (2010) by conducting a planned comparison between mean congruent and incongruent FA rates for the flash condition. For completeness, we did the same for the pitch cue condition (note that $\alpha$ level = 0.025; Bonferroni corrected to control the possible inflation of Type I-error, making the tests more conservative). The tests revealed that the difference was significant only in the flash condition ($p = 0.004$, with FA rates being higher in the congruent than incongruent condition), not in the pitch condition ($p = 0.14$). Despite some differences in experimental parameters, our result replicated Prinzmetal et al. (2010).

In experiment 3 we kept the load of perceptual processing low by using easily discernible stimuli but set the load of response execution high by introducing a speed deadline. To allow the effects of voluntary and involuntary attention to emerge at their respective optimum time frames, we used different cue–target delay durations previously shown to be suitable for flash and auditory pitch. Despite response pressure, peripheral flashes caused a robust cuing effect indicative of involuntary attention. By contrast, auditory pitch induced no effect. This suggests that speed pressure led participants to ignore auditory pitch, eliminating its effect.

## 5 General discussion

Our cognitive system has a reliable tendency to associate auditory pitch with location on the vertical axis (Eitan & Timmers, 2010; Spence, 2011). In our previous study (Chiou & Rich, 2012) we report results suggesting the association between pitch and location relies on the volitional system of attention, despite it occurring when the pitch is nonpredictive. Other studies (Evans & Treisman, 2010; Mossbridge et al., 2011), however, interpret such pitch-elicited effects as involuntary phenomena. Here, we investigated this issue more directly, by examining the effects of cognitive and perceptual demands on the attentional shifts driven by nonpredictive pitch. Our exploration includes the contrasts with a classic cue that triggers involuntary attention shifts, the peripheral flash, and with a cue that encourages clearly voluntary shifts, the predictive pitch. The results of three experiments consistently showed that imposing perceptual or response pressure eliminated any effect of nonpredictive pitch, presumably by reducing volitional resources. By contrast, informative pitch cues and uninformative flash cues were able to produce robust cuing despite the load on perception or response. This illustrates the importance of task-relevance and salience: under a context of high cognitive load, attentional cues affect performance only when they are informative (which motivates processing of the cues) or when they consist of salient bottom-up incoming signals which capture attention. Although we must be cautious in interpreting any null effect, we found that the nonpredictive pitch possessed neither of these characteristics, failing to affect attention when we manipulated either the perceptual or response load of the task. As load insensitivity has been suggested as one of the critical diagnostic criteria of an automatic process (Moors & De Houwer, 2006), the present results are consistent with the hypothesis that the biasing effect of the pitch–location mapping is mediated by the volitional attention system.

Our results raise a cautionary note to the interpretation that cross-modality correspondences reflect automatic processing. Two recent studies reported significant effects of task-irrelevant (uninformative) pitch on visual discrimination/matching that the authors interpret as evidence for automaticity. Evans and Treisman (2010) used a task that presented concomitant auditory cue and visual target and required discrimination of grating orientation; they found a pitch–location congruency effect. Mossbridge et al. (2011), like our experiments, presented preceding auditory cues before a visual target (they called it a ‘probe’) and used a colour matching task; they also observed a congruency effect. Both papers interpret these results as favouring the view that the pitch–location mapping occurs automatically at a perceptual level. However, our findings suggest otherwise: although pitch biases attention orienting despite it conveying no information about possible target location (Chiou & Rich, 2012), this biasing effect cracks under the pressure of perceptual and response demands, indicating that it does not depend on automatic mechanisms. Considered alongside our previous findings that show the impact of cognitive interpretation and volitional suppression on the cuing effect (experiments 3 and 4b, Chiou & Rich, 2012), the overall pattern of data favours the view that the pitch cuing phenomenon relies on voluntary attention.

Here we were specifically interested in testing attention orienting and how it is guided cross-modally by the pitch–location mapping. Other researchers have looked at the nature of cross-modal mappings from different perspectives, such as examining the effects of multisensory integration using simultaneous signals. For example, there is evidence showing that cross-modality mappings are able to modulate the strength of coupling during audiovisual integration (Parise & Spence, 2008, 2009; but also see Keetels & Vroomen, 2011; for review see Parise & Spence, 2013). As the main focus of the present study was the impact of auditory cues on attentional orienting, we adopted and modified the classic Posner attention cuing paradigm (Posner, 1980; Prinzmetal et al.'s designs are derived from this classic paradigm), which has been extensively used to test attention cuing under various circumstances. We found no cuing effect of nonpredictive pitch under greater perceptual and response burden, consistent with the load sensitivity of the volitional mechanisms. However, whether the pitch–location mapping occurs automatically during audiovisual integration constitutes a related but separate issue, and still awaits future research.

Consistent with our proposal that the pitch–location mapping effect on attention reflects top-down volitional processes, recent evidence also suggests that other types of cross-modality associations may involve more high-level rather than stimulus-driven processes. For instance, a recent study demonstrated that the mapping between pitch and brightness (ie high pitch is associated with brighter stimuli whereas low pitch is associated with dimmer stimuli) is not stimulus-driven but instead reflects top-down control (Klapetek, Ngo, & Spence, 2012). This has led some researchers to question whether different types of cross-modality mappings are commonly subserved by volitional operations. In a recent review Spence and Deroy (2013) examined the evidence for and against the automatic and volitional accounts of cross-modal correspondences. They concluded that the answer to this question relies on the researcher's definition of automaticity. If one defines automaticity as the ability to affect performance despite task-irrelevance, the mappings would be classified as automatic processes. However, this liberal definition ignores the fact that cross-modality mappings critically rely on a range of high-level cognitive processes, such as voluntary attention and abstract concepts (Chiou & Rich, 2012; L. Walker, P. Walker, & Francis, 2012; P. Walker, 2012; P. Walker & L. Walker, 2012) and is highly sensitive to perceptual and response demands, as the present study has demonstrated. These properties, taken together, suggest that cross-modal correspondences are underpinned by volitional processes and possibility occur at a late stage of the perceptual processing cascade.

In a broader sense, the present study provides insights as to the cognitive mechanisms that underlie the interaction between the attentional system and stimuli that imply a location in space. A number of stimuli with implicit associations to locations have been shown to affect attention orienting and are often suggested to reflect automatic processes (eg religious terms: Colzato et al., 2010; metaphors: Meier & Robinson, 2004). Similar to auditory pitch, these stimuli are associated with location at a higher cognitive level (eg semantics) rather than at a lower perceptual level (eg the vividness of stimuli). We suggest such claims need to be tested more rigorously with examinations of their sensitivity to pressure and robustness to volitional control before claims of automaticity are made. This is particularly important as we work towards understanding the distinctions and similarities between voluntary and involuntary mechanisms of attention.

**References**


Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433–436.

Carrasco, M., & Yeshurun, Y. (2009). Covert attention effects on spatial resolution. *Progress in Brain Research*, **176**, 65–86.

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

Colzato, L. S., van Beest, I., van den Wildenberg, W. P., Scorolli, C., Dorchin, S., Meiran, N., … Hommel, B. (2010). God: Do I have your attention? *Cognition*, **117**, 87–94.

Dufour, A. (1999). Importance of attentional mechanisms in audiovisual links. *Experimental Brain Research*, **126**, 215–222.

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, **114**, 405–422.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, **10**(1):6, 1–12.

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, **6**, 555–556.

Jonides, J. (1981). Voluntary versus automatic control over mind's eye's movement. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum.

Keetels, M., & Vroomen, J. (2011). No effect of synesthetic congruency on temporal ventriloquism. *Attention, Perception, & Psychophysics*, **73**, 209–218.

Klapetek, A., Ngo, M. K., & Spence, C. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics*, **74**, 1154–1167.

Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, **9**, 75–82.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science*, **15**, 243–247.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, **132**, 297–326.

Mossbridge, J. A., Grabowecky, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual–spatial attention. *Cognition*, **121**, 133–139.

Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters*, **442**, 257–261.

Parise, C., & Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE*, **4**(5): e5664.

Parise, C., & Spence, C. (2013). Audiovisual crossmodal correspondences. *The Oxford handbook of synesthesia.* Oxford: Oxford University Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437–442.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, **32**, 3–25.

Prinzmetal, W., Ha, R., & Khani, A. (2010). The mechanisms of involuntary attention. *Journal of Experimental Psychology: Human Perception and Performance*, **36**, 255–267.

Prinzmetal, W., & Landau, A. N. (2008). Dissecting spatial visual attention. In V. Coltheart (Ed.), *Tutorials in visual cognition* (pp. 43–66). Hove, Sussex: Psychology Press.

Prinzmetal, W., Zvinyatskovskiy, A., Gutierrez, P., & Dilem, L. (2009). Voluntary and involuntary attention have different consequences: The effect of perceptual difficulty. *The Quarterly Journal of Experimental Psychology*, **62**, 352–369.

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review*, **9**, 507–513.

Ristic, J., & Kingstone, A. (2006). Attention to arrows: Pointing to a new direction. *The Quarterly Journal of Experimental Psychology*, **59**, 1921–1930.
Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, **1191**, 182–200.
Spence. C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, **73**, 971–995.
Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*, **22**, 245–260.
Walker, L., Walker, P., & Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception*, **41**, 1186–1192.
Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention, Perception, & Psychophysics*, **74**, 1792–1809.
Walker, P., & Walker, L. (2012). Size–brightness correspondence: Crosstalk and congruity among dimensions of connotative meaning. *Attention, Perception, & Psychophysics*, **74**, 1226–1240.

© 2015 a Pion publication