Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies

Emily Orchard-Mills   School of Psychology, University of Sydney, Sydney, New South Wales, Australia   ✉

Erik Van der Burg   School of Psychology, University of Sydney, Sydney, New South Wales, Australia   ✉

David Alais   School of Psychology, University of Sydney, Sydney, New South Wales, Australia   ✉

A recent study by Guzman-Martinez, Ortega, Grabowecky, Mossbridge, and Suzuki (2012) showed that participants match the frequency of an amplitude-modulated auditory stimulus to visual spatial frequency with a linear relationship and suggested this crossmodal matching guided attention to specific spatial frequencies. Here, we replicated this matching relationship and used the visual search paradigm to investigate whether auditory signals guide attention to matched visual spatial frequencies. Participants were presented with a search display of Gabors, all with different spatial frequencies. When the auditory signal was informative, improved search efficiency occurred for some spatial frequencies. However, when uninformative, a matched auditory signal produced no effect on visual search performance whatsoever. Moreover, search benefits were also observed when the auditory signal was informative, but did not match the spatial frequency. Together, these findings suggest that an amplitude-modulated auditory signal can influence visual selection of a matched spatial frequency, but the effect is due to top-down knowledge rather than resulting from automatic attentional capture derived from low-level mapping.

Introduction

Research into multisensory perception has shown many examples of crossmodal interactions where auditory signals can bias visual perception (Alais, Newell, & Mamassian, 2010). This has been demonstrated in a wide range of tasks including spatial judgments (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003), temporal judgments (Bertelson & Aschersleben, 2003; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004), motion perception (Arrighi, Marini, & Burr, 2009; Hidaka et al., 2009; Teramoto et al., 2012), and perceptual ambiguity (Alais, van Boxtel, Parker, & van Ee, 2010; Holcombe & Seizova-Cajic, 2008; Sekuler, Sekuler, & Lau, 1997; van Ee, van Boxtel, Parker, & Alais, 2009). Detection is faster and more accurate for spatially and temporally congruent audiovisual targets compared to that for unimodal targets (Bolognini, Frassinetti, Serino, & Làdavas, 2004; Driver & Spence, 1998; Frassinetti, Bolognini, & Làdavas, 2002; Stein, Meredith, Huneycutt, & McDade, 1989), and this behavioral benefit has been supported by event-related potential (ERP) data showing evidence of early crossmodal interactions (Giard & Peronnet, 1999; Molholm et al., 2002). These findings illustrate that visual processing is not isolated from signals in other modalities and that behavioral performance can be improved by interactions between vision and audition.

Although many audiovisual interactions involve spatially and temporally coincident stimuli, spatial coincidence is not necessarily required for crossmodal interactions to occur, and temporal coincidence alone can be sufficient. For instance, in a rapid serial visual presentation paradigm, a nonspatial auditory stimulus that is task-irrelevant has been shown to facilitate detection of a synchronized visual target when the auditory stimulus is salient (Olivers & Van der Burg, 2008; Vroomen & de Gelder, 2000). In another example using nonspatial sound, the visual search paradigm was used to show that a transient auditory stimulus synchronized with a change in the visual target can dramatically improve visual search efficiency by creat-

ing a "pop-out" effect (Van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008; Van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011; Zannoli, Cass, Mamassian, & Alais, 2012). The pop-out effect also occurs when the auditory signal is replaced by a tactile signal (Ngo & Spence, 2010; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2009). Similarly Noesselt, Bergmann, Hake, Heinze, & Fendrich (2008) showed that temporally congruent but spatially nonaligned auditory stimuli enhanced detection of a brief blink in a visual stimulus. Furthermore, Murray et al. (2005) demonstrated multisensory interactions in low-level sensory cortices for simultaneously presented auditory and tactile stimuli that were not constrained by spatial alignment. These studies complement the classic demonstrations of crossmodal interactions based on spatio-temporal congruency by showing that purely temporal mechanisms can drive crossmodal interactions between vision, audition, and touch and improve visual performance.

Visual performance is also improved by attention; indeed, attending to one feature of a visual object enhances processing of its other features even when they are task irrelevant (O'Craven, Downing, & Kanwisher, 1999). In an interesting study, Busse, Roberts, Crist, Weissman, and Woldorff (2005) demonstrated that this attentional spread extends across both space and modalities. Using ERP data, they showed that attention voluntarily allocated to a visual event spread to temporally coincident sounds that were spatially separated and task-irrelevant. In another ERP study, Molholm, Ritter, Javitt, and Foxe (2004) found that participants responded more quickly when the image of an animal and its characteristic vocalization were simultaneously presented than for targets presented in one modality. They proposed that this is mediated through object-based attention that is multimodal (Molholm, Martinez, Shpaner, & Foxe, 2007). Fiebelkorn et al. (2009) postulated that the mechanism by which an attended visual stimulus can be modulated by a task-irrelevant sound occurs in two additive stages, firstly by a stimulus-driven spread of attention that is independent of learned associations occurring with synchronously presented targets, and secondly by a representational-driven spread of attention dependent on learned associations. These studies demonstrate that auditory information related to an object can speed detection of the visual object in simple displays.

The previously mentioned experiments demonstrated that auditory stimuli can enhance visual processing and speed visual search through spatial and temporal interactions. In this study we will investigate crossmodal interactions between amplitude-modulated auditory stimuli and visual spatial frequencies. This topic was examined in a recent study by Guzman-Martinez et

al. (2012) that found evidence for a specific mapping between frequency of auditory amplitude modulation and visual spatial frequency using a frequency matching task, such that high frequencies of auditory amplitude modulation corresponded with high visual frequencies, and vice versa. They proposed the association arises from sounds produced by haptic exploration of surfaces, suggesting the frequency of the amplitude modulation of the sound produced is related to the spatial scale of the texture of the surface (in effect, its visual spatial frequency). To confirm the crossmodal frequency mapping, participants were presented with two Gabor patches of very different spatial frequencies (0.5 and 4.0 cycles per cm) and an auditory stimulus matching one of them (which participants were instructed to ignore). After 250 ms, one of the Gabors (i.e., the target) phase-shifted either to the left or right and participants responded as quickly as possible as to the direction of shift. They reported that participants responded faster when the auditory stimulus matched the Gabor that shifted and suggested the uninformative auditory stimulus automatically guided attention to the matching visual-spatial frequency.

In the experiments reported here, we replicate the matching results of Guzman-Martinez et al. (2012), finding that participants can match auditory amplitude-modulation frequencies and visual spatial frequencies with a linear relationship. In the following experiments these matches were used to test the proposal that an amplitude-modulated auditory stimulus can guide attention to a target of matched visual-spatial frequency. First, using the visual search paradigm, the specificity of these matches was probed in a cluttered visual display. Next, the effect of the difference in spatial frequency between target and distractor was assessed in a simplified task with only two elements. Then, in a search task we assessed the effect of an uninformative auditory stimulus on search times to a matched visual spatial frequency. Following this, the influence of the range of spatial frequencies in the display was investigated. We also examined the effect of unmatched auditory frequencies in guiding visual search by offsetting the range of auditory and spatial frequencies from their nominal "match." Our findings show that although amplitude-modulated auditory signals can be used to guide attention to visual spatial frequencies, the mapping is limited and flexible, as an uninformative matched auditory signal resulted in no search benefit, whereas an informative auditory signal shifted from the reported matching produced search benefits. This argues against an automatic, stimulus-driven attentional capture and instead implies a more arbitrary mapping that is broadly tuned and categorical.

Experiment 1

This experiment was very similar to the matching experiments of Guzman-Martinez et al. (2012; experiments 1 and 4). Using the method of adjustment, we asked participants to adjust the frequency of the amplitude-modulated noise until it matched the spatial frequency of a Gabor patch. The aim was to replicate the approximately linear relationship reported by Guzman-Martinez et al. (2012) between auditory and visual frequency matches, and to do so for each participant so that subsequent experiments can use each participant's own subjective matching of audio-visual frequency.

Method

Ethical statement

Written consent was obtained from each participant prior to the experiments. The experiments were approved by the local ethics committee of the University of Sydney and adhere to the tenets of the Declaration of Helsinki.

Participants

Thirteen individuals (four female, mean age 27 years, range 19 to 46 years) participated, all reported normal hearing and normal or corrected-to-normal vision. Ten participants were naive as to the purpose of the experiment.

Stimuli and apparatus

Visual stimuli were displayed on a linearized 16-in. CRT monitor (100 Hz, 1024 × 768, mean luminance 51.5 cd m^{-2} and maximum luminance 116 cd m^{-2}) situated 50 cm from the participants in a dark quiet room. Visual stimuli were Gabor patches, 0.73 peak-to-trough contrast, standard deviation (SD) of 0.42 cm (0.37° visual angle) presented in the center of a gray screen set to mean luminance. The Gabors were presented between cos and sine phase, in order to show both a peak and trough. The spatial frequencies used were 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, and 4.5 cycles per centimeter (corresponding to 0.43, 0.87, 1.3, 1.74, 2.17, 2.61, 3.04, and 3.48 cycles per degree). For consistency with Guzman-Martinez et al. (2012), we have reported visual spatial frequency in cycles per cm, as they demonstrated the matching relationships were associated with physical rather than retinal spatial frequency. Experiments were conducted using a Mac-Pro (Apple, Cupertino, CA) running Matlab (The MathWorks, Natick, MA) with the Psychophysics

Toolbox. The auditory stimulus was amplitude-modulated white noise. The amplitude modulation was 100% and the noise was high-pass filtered at 3500 Hz (the starting value used by Guzman-Martinez et al., 2012 in experiment 1). It was presented at a comfortable listening level (65 dB SPL at listening position) through speakers located on either side of the monitor.

Design and procedure

Each trial began with the presentation of a gray screen; after a key-press a Gabor patch of a randomly selected spatial frequency appeared, and participants were asked to adjust the frequency of amplitude modulation of the auditory noise to match the visual-spatial frequency. Participants could increase or decrease the frequency of the amplitude modulation in 0.5 Hz steps by pressing the up and down arrows on the keyboard (the minimum frequency was 0.5 Hz and the maximum 12 Hz). The task was unspeeded and participants pressed the enter key when they were satisfied with the match made. The starting frequency of the auditory signal was randomized for each trial. Participants completed five practice trials to familiarize themselves with the task, followed by a single block in which each of the nine Gabors was presented 10 times (90 trials in total) and the matches for each frequency averaged.

Results

An analysis of variance (ANOVA) was performed on the auditory matches, with visual spatial frequency as a within-subject variable. Figure 1 plots the mean auditory amplitude-modulation frequency match (Hz) against each visual spatial frequency (cycles per cm) in the left panel and individual responses in the right panel.

There was a highly significant effect of spatial frequency, $F(8, 96) = 133.4$, $p < 0.001$, indicating that participants' auditory matching varied with visual spatial frequency. To investigate whether participants map neighboring visual spatial frequencies with distinct auditory frequencies, we compared the auditory match for each spatial frequency with the auditory match for the neighboring frequency, resulting in eight comparisons. All comparisons were significant, $t(12)$, $ps < 0.002$, indicating that participants reported distinct matches for each spatial frequency. To test for linearity across participants we used the approach described in experiment 4 of Guzman-Martinez et al. (2012). Individual matching relationships were first normalized, then each relationship was fitted to a quadratic function $y = ax^2 + bx + c$. Across participants the coefficient a was not significantly different from zero,
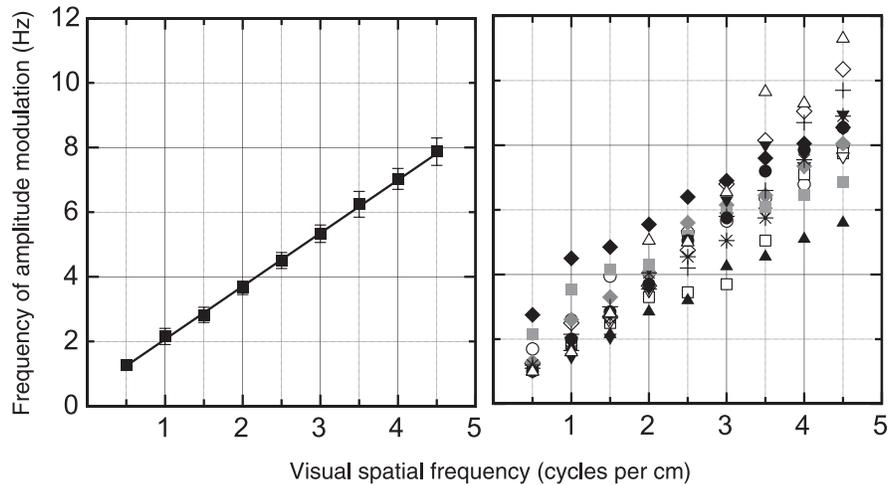
Figure 1. Data from Experiment 1 showing mean ($n = 13$, left panel) and individual (right panel) frequency matching data. Observers adjusted the frequency of amplitude-modulation of auditory noise (Hz) until it matched the visual spatial frequency (cycles per cm). Error bars represent ± 1 *SEM*.

$t(12) = 0.42$, $p = 0.7$, indicating a linear relationship. Furthermore, the chi-squared statistic for the difference between the linear and quadratic fit was not significant for any of the individual participants matching relationships (the largest difference was 0.85, well below the critical value of 3.84), demonstrating a linear function adequately predicted each participant's responses and was not improved by adding a second parameter.

A post-hoc analysis was also performed comparing the matching responses as a function of the initial amplitude-modulation frequency of the auditory signal. More precisely, we compared the overall mean match when the initial amplitude modulation rate started at a low rate (i.e., <6 Hz) to when the initial amplitude modulation rate started at a high rate (i.e., >6 Hz). The mean matches are shown in Figure 2, demonstrating that when participants were initially presented with a low rate, their final match was lower than when participants were initially presented with a high rate. This was statistically confirmed by a two-tailed *t* test, $t(12) = 5.8$, $p < 0.001$. Figure 2 also shows the overall matching curve for each spatial frequency as a function of the initial frequency rate of the auditory stimulus. In this figure, responses from all participants were pooled for each visual spatial frequency and the means compared based on the starting frequency of amplitude modulation.

Discussion

The results of Experiment 1 replicate the matching relationship demonstrated by Guzman-Martinez et al. (2012) in showing that participants can systematically match the frequency of auditory amplitude modulation to a given visual spatial frequency and can do so in a consistent and linear manner. This is an interesting observation, but raises questions about what might underlie this systematic matching performance. For example, it may be that this matching relationship is a high-level process done at a cognitive level, rather than at a lower perceptual level where the stimulus attributes of auditory amplitude modulation and visual spatial frequency are initially encoded. The post-hoc analysis suggests that matching depends on the starting frequency of the auditory signal, which is more consistent with higher-level processing. The following experiments are designed to test whether or not the ability to match auditory amplitude modulation to visual frequencies reflects a low-level sensory process in which the two stimulus dimensions are bound together, as proposed by Guzman-Martinez et al. (2012). If it is due to an early sensory combination, presenting matched auditory and visual stimuli should facilitate visual selection in a stimulus-driven way. Our next experiment will test this using the visual search paradigm.

Experiment 2

This experiment used the visual search paradigm to investigate whether search for a target Gabor among distractor Gabors (with different spatial frequencies) could be facilitated by an amplitude-modulated auditory signal matching the target frequency. Participants were instructed that the auditory signal was informative. This is a point of difference with experiment 5 reported by Guzman-Martinez et al. (2012), who
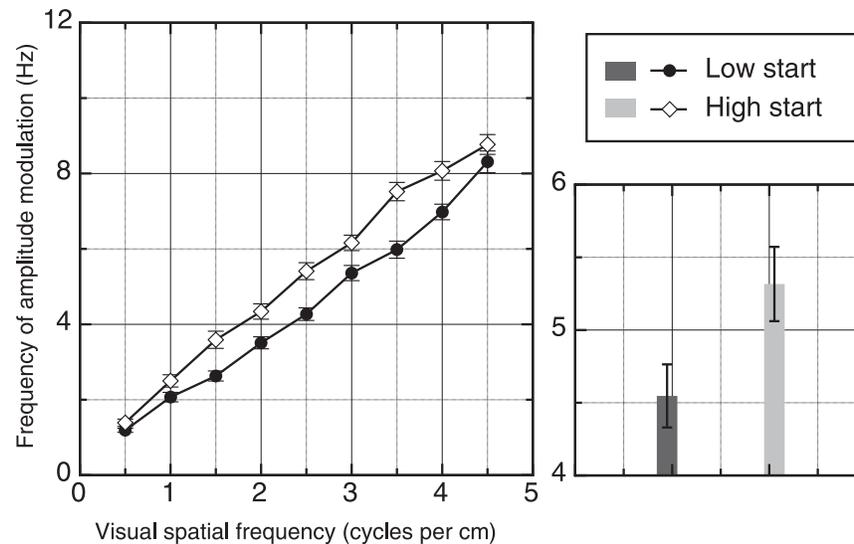
Figure 2. Results from the post-hoc analysis performed on the data from Experiment 1. Mean matches for pooled responses of amplitude-modulated auditory noise (Hz) as a function of visual spatial frequency (cycles per cm) presented shown on left. Filled circles are low starting rates and open diamonds are high rates. Overall mean for all participants ($n = 13$) shown on right; light gray is low and dark gray is high starting amplitude-modulation rates. Error bars are ± 1 *SEM*.

instructed participants to ignore the auditory signal. This was done in order to maximize the chances of finding search benefits. Instructing participants to use the auditory signal will likely enhance any existing bottom-up attentional capture, shown in previous crossmodal interactions (Mishra, Martinez, & Hillyard, 2010). Furthermore, in line with this, other studies have shown that attention is a prerequisite for binding information from different modalities (see Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010, for a recent review about attention and multisensory integration). In the present experiment and all subsequent ones, we used the same participants from Experiment 1 and their own frequency matching data. The right panel of Figure 1 shows the frequency matching data from Experiment 1 separately for each observer.

Using a cluttered visual display (see Figure 3 for an example) comprised of a target Gabor plus multiple distractor Gabors with unique spatial frequencies, we manipulated the set size to assess search efficiency for both sound-present and sound-absent trials. Participants were asked to find the target Gabor among the distractors. To verify that participants detected the target Gabor, we presented a ring around each Gabor with a notch on the left or right side, and participants were required to indicate as quickly and accurately as possible the location of the notch on the ring surrounding the target. The distractors also had surrounding rings with a notch on either the left or right side, but the notch deviated $\pm 15°$ from horizontal. The dependent variable was mean response time (RT) for correct search. In sound-present trials each participant was presented with the auditory signal that

they had previously indicated matched the visual-spatial frequency of the target Gabor. If amplitude-modulated noise of matched frequency guides attention towards the target spatial frequency in an automatic, stimulus-driven fashion, then we expect efficient visual search for the sound-present trials. This is characterized by shallow set-size functions as search time is little affected by additional distractors. In contrast, we expect inefficient search (steep set-size functions) for sound-absent trials as the participants have no information about the frequency of the target Gabor and detection of the notch requires serial inspection of the items in the display.

Method

Participants

Nine of the individuals from Experiment 1 participated (four female, mean age 29 years, range 19 to 46 years; two authors).

Stimuli and apparatus

The stimuli and apparatus were identical to Experiment 1. The same visual spatial frequencies were used and each participant's own auditory matches were used (see right panel of Figure 1).

Design and procedure

Each trial began with the presentation of a white fixation cross for 1750 ms in the center of a gray screen;
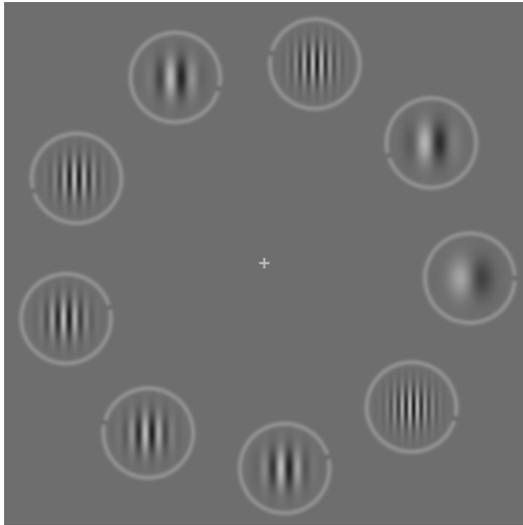
Figure 3. Example visual display, displaying all the visual spatial frequencies tested. The ring surrounding the target spatial frequency has a notch that is horizontally aligned, while the rings' surrounding distractors have notches at ±15° from horizontal. For illustrative purposes, the rings and notches are wider and brighter than in the display used.

following this five or nine Gabor patches appeared on the screen. The Gabor patches were equally spaced around fixation on an imaginary circle (radius 6°). The location of each element was randomly determined on each trial. The target spatial frequency was randomly chosen for each trial, and the distractor spatial frequencies were then randomly chosen from the remaining spatial frequencies without replacement. Each Gabor was surrounded by a gray ring (radius 3°) with a small notch gap randomly on the left or right side. The rings (and the notch gap) were drawn with a smooth Gaussian profile. All stimuli were surrounded by a notched ring, but the target's notch was horizontally aligned, while for the distractors it deviated ±15° from horizontal (see Figure 3, which is an example of the display screen). Participants were asked to respond as quickly and accurately as possible to the location of the notch by pressing the left or right arrow key, respectively.

In half of the trials the auditory signal was present (sound-present condition), and in the remaining trials the display was presented in silence (sound-absent condition). In the sound-present condition, the auditory signal matched the spatial frequency of the target Gabor. Participants were instructed that the auditory signal was matched with the target and to use this information to guide their search. Consistent with experiment 5 of Guzman-Martinez et al. (2012), in sound-present trials the auditory signal began 250 ms prior to the presentation of the search display. A block of 15 practice trials was given to familiarize partici-

pants with the task, followed by three blocks of 360 trials (1,080 trials in total). The notch location (left or right) for the target ring, the target visual spatial frequency (0.5–4.5 cycles per cm), set size (five or nine elements) and sound condition (present or absent) were randomly chosen for each trial and balanced within blocks. No feedback was given to participants in this and all subsequent experiments.

Results

The overall error rate was low (1.7%), and therefore no further analysis of errors was conducted. All erroneous trials were excluded from the RT analysis. RTs greater than 10 s and those that differed more than 3 SDs from the mean for each participant for each condition were excluded from analysis (1.2%). RTs were subjected to a repeated-measures univariate ANOVA with spatial frequency, set size, and the sound condition as within-subject variables. The Huynh-Feldt correction was used when appropriate. The top left panel of Figure 4 shows corrected mean RTs (collapsed over spatial frequency) as a function of set size and sound condition.

The ANOVA yielded a significant effect for the sound condition, $F(1, 8) = 26.3$, $p = 0.001$, indicating that participants responded faster when the auditory signal was present (2124 ms) than when the auditory signal was absent (2740 ms). This effect was not consistent across all spatial frequencies tested (see lower panel of Figure 4).

The dependence on spatial frequency was confirmed by a significant three-way interaction among sound condition, set size, and spatial frequency in the ANOVA, $F(8, 64) = 6.3$, $p < 0.01$. Separate ANOVAs for each spatial frequency showed reliable sound condition × set-size interaction for the lower and higher spatial frequencies of 0.5, 1.0, and 1.5, and 3.5, 4.0, and 4.5 cycles per cm (Fs > 8.4; ps < 0.02) but not for the intermediate spatial frequencies of 2.0, 2.5, and 3.0 cycles per cm (Fs < 0.4; ps > 0.56). Separate ANOVAs for each spatial frequency for each sound condition, showed no reliable set-size effect, $F(1, 8) = 3.6$, $p = 0.1$, when the visual spatial frequency was 0.5 cycles per cm. In contrast, the set-size effect was significant for all other spatial frequencies (Fs > 45.4; ps < 0.001).

One might argue that comparing the sound-present condition with the sound-absent condition is not a valid comparison. For instance, in the sound-present condition, the sound may have given some temporal information about when to expect the search display, which was absent in the sound-absent condition, explaining the search benefits in the sound-present condition versus the sound-absent condition (see e.g.,
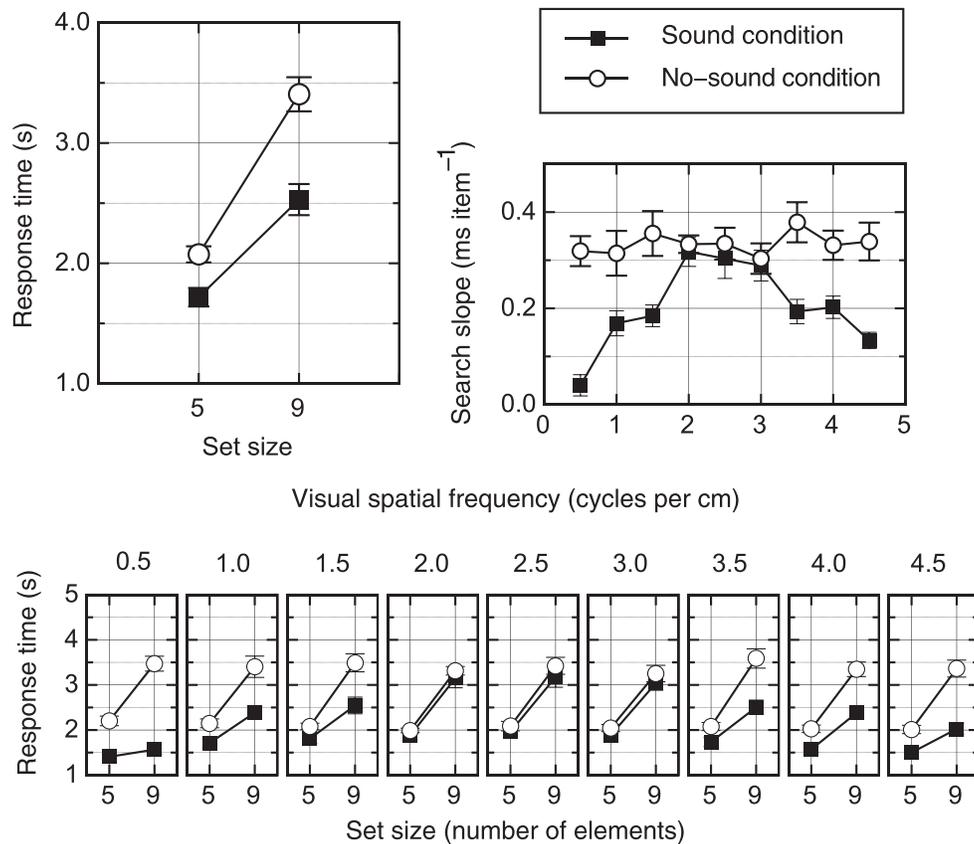
Figure 4. Data from Experiment 2 showing corrected mean RTs ($n = 9$) for each set size, for sound-present (filled squares) and sound-absent (open circles) in top left panel and set-size slope, a measure of search efficiency, is shown for the sound-present and sound-absent conditions in the top right. The RTs for both set sizes and sound conditions for each spatial frequency are plotted in the lower half of the Figure. Error bars are ± 1 *SEM*.

Los & Van der Burg, in press). Therefore, an ANOVA was also performed on the search slopes (i.e., search efficiency), with spatial frequency and sound condition as within-subject variables. The search slopes are shown in the top right panel of Figure 4. The benefit of this analysis is that any beneficial effects of the sound-present condition (such as preparation effects) are eliminated after subtraction of the RTs in the two set-size conditions. The effects seen in RTs were replicated, with a significant effect of the sound condition, $F(1, 8) = 24.8$, $p = 0.001$, and a significant interaction between spatial frequency and sound condition, $F(8, 64) = 6.3$, $p < 0.001$. Pair-wise comparisons (Bonferroni-corrected) between the sound-present and sound-absent condition showed significant differences for the lowest and highest spatial frequencies of 0.5, 1.0, 1.5, 3.5, 4.0, and 4.5 cycles per cm ($ps < 0.02$) but not for the intermediate spatial frequencies of 2.0, 2.5, and 3.0 cycles per cm ($ps > 0.6$). A separate ANOVA on the slopes of the sound condition showed a significant effect of spatial frequency, $F(8, 64) = 14.8$, $p < 0.001$.

Discussion

Experiment 2 showed that for some spatial frequencies participants were able to use the auditory signal to speed RTs when searching a cluttered visual display for a matched spatial frequency. To an extent, this supports the findings of Guzman-Martinez et al. (2012) that crossmodal interactions between amplitude-modulated auditory signals and visual spatial frequency can facilitate visual selection. However, although there was a reliable effect of sound for spatial frequencies of 0.5, 1.0, and 1.5, and for 3.5, 4.0, and 4.5 cycles per cm, such effects were absent for the intermediate spatial frequencies (2.0, 2.5, and 3.0 cycles per cm). It may be that the mapping between auditory and visual-spatial frequencies is broadly tuned, allowing attention to be drawn to spatial frequencies in the display around the nominally "matching" frequency. This may have restricted the effective set size by reducing the possible target locations to just a few options. Another possibility is that the high display density forced participants to adopt a focused attentional window, considering only a few elements at a time. It has been

shown that objects that would normally stand out from the environment (such as a red object among green objects) capture less attention when participants adopt a small, focused attentional window compared to a large, more distributed attentional window (Belopolsky, Zwaan, Theeuwes, & Kramer, 2007; Van der Burg, Olivers, & Theeuwes, 2012). To address these possibilities, Experiment 3 will use a simplified display (with two elements only) to assess if participants can consistently select the matched visual-spatial frequency when only a single distractor is present.

Experiment 3

In order to assess the precision of audiovisual frequency matching more accurately, Experiment 3 used a simplified display with only two Gabor patches: one target frequency and one distractor frequency. As in Experiment 2, participants were asked to locate the target Gabor and indicate if the notch was on the left or right side. The auditory signal was always present (unlike Experiment 2) and its validity was manipulated so that in 80% of trials it matched the target Gabor, while in 20% of trials it matched the distractor Gabor. If participants can use the auditory signal to guide attention to a matched spatial frequency, there should be search benefits when the Gabor spatial frequency is consistent with the auditory signal, and search costs when it is not (i.e., a validity effect). The spatial frequency difference between the target and distractor was manipulated to examine the resolution with which participants can use an auditory signal to find a matching spatial frequency. If the auditory signal guides attention specifically to a matching visual spatial frequency, we should see a validity effect for all the combinations presented. However, if mapping between the auditory signal and visual spatial frequency is broad and imprecise, the search advantage would be present only for combinations containing substantial spatial frequency differences.

Method

Participants

Eight of the individuals from Experiment 1 participated (four female, mean age 27 years, range 19 to 38 years; one author).

Stimuli and apparatus

The experimental set-up was the same as Experiment 1 and used the same visual spatial frequencies and specific matches for each participant for the frequency of amplitude modulation of the auditory stimulus.

Design and procedure

Experiment 3 was similar to Experiment 2, with the Gabor patches and their surrounding rings drawn from the same set; however, there were always just two Gabors (one target and one distractor). The auditory signal was always present, but in 80% of trials it matched the visual spatial frequency of the target Gabor (valid), and in the remaining trials it matched the distractor Gabor (invalid). Six combinations of visual spatial frequency were tested in separate blocks: 0.5/4.0, 1.0/3.5, 1.5/3.0, 2.0/2.5, 1.0/1.5, and 3.0/3.5 cycles per cm. The combination of 0.5/4.0 was completed first to ensure participants were able to use the auditory signals effectively, and thereafter the combinations were randomized for each participant.

Results

The overall error rate was low at 1.9% and erroneous trials were not examined further. RTs greater than 5 s and those more than 3 SDs from the mean for each participant for each condition were excluded from analysis (0.8%). Figure 5 shows corrected mean RTs for valid and invalid trials for the largest spatial frequency difference of 0.5/4.0 cycles per cm. There is a clear validity effect from pairing the target Gabor with a matched auditory signal, with the valid condition being faster by 226 ms.

The RTs were subjected to an ANOVA with spatial frequency difference and the auditory condition (valid or invalid) as within-subject variables. The ANOVA yielded a significant main effect for the auditory condition, $F(1, 6) = 23.5$, $p < 0.01$, with participants showing faster RTs when the auditory signal matched the target Gabor than when the auditory signal matched the distractor Gabor (1100 vs. 1230 ms). The two-way interaction between spatial frequency combination and auditory condition was also significant, $F(5, 30) = 6.7$, $p < 0.01$. Separate ANOVAs were conducted for each spatial frequency combination. These showed a significant search advantage for the valid auditory signal for spatial frequency combinations of 0.5/4.0, 1.0/3.5, and 1.5/3.0 cycles per cm (Fs > 14.3; ps < 0.01). Frequency combinations of 1.0/1.5, 2.0/2.5, and 3.0/3.5 cycles per cm did not produce significant search advantages for valid auditory signals (Fs < 4.6; ps > 0.08).

Figure 6 shows the differences in search times between valid and invalid conditions for all combinations of spatial frequency difference expressed in octaves (that is, the base 2 logarithm of the frequency
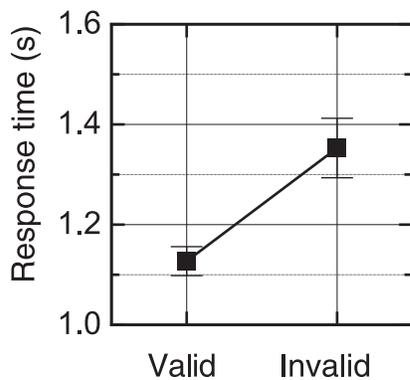
Figure 5. Data from Experiment 3 showing corrected mean RTs (*n* = 8) as a function of sound condition (Valid/Invalid) for the 0.5/4.0 cycles per cm combination. Error bars are ±1 *SEM*.

Figure 6. Data from Experiment 3 showing the difference in corrected mean RTs (*n* = 8) between valid and invalid conditions as a function of the octave spatial frequency difference of the visual elements in the display. Error bars are ±1 *SEM*.

ratio). Expressing spatial frequency differences in octaves is standard because the spatial frequency dimension is linear on a $\log_2$ scale (for example, cortical spatial frequency filters are equally wide when expressed in octaves). When plotted this way, there is a clear monotonic increase in search advantage as spatial frequency difference increases, becoming significant only when there is a difference of at least one octave. This suggests that large frequency differences between the target and distractor Gabors (many times greater than the just-noticeable difference) are required to obtain a validity effect.

Discussion

Experiment 3 demonstrated that participants could direct their attention to the visual spatial frequency that matched the auditory signal; however, in accordance with the findings of Experiment 2, the search benefit afforded by the matching auditory signal depended on the combination of visual spatial frequencies used for the target and distractor. As the spatial frequency difference between target and distractor reduced, participants were less able to use the matched auditory signal to direct their attention to the target. This pattern of results is consistent with those of Experiment 2, showing that participants are able to use the auditory signal for a broad categorical match (low or high spatial frequency) but are unable to make specific matches unless the Gabors differ substantially in spatial frequency. Indeed, a frequency difference of one octave (i.e., a frequency doubling) is needed to reliably obtain a spatial frequency search advantage from a matching auditory signal. To put the large size of this difference in context, the Weber fraction for detecting changes in spatial frequency is in the range of 0.02–0.05 (Hirsh & Hylton, 1982; Regan, Bartol, Murray, & Beverly, 1982). Therefore, a one-octave frequency difference is
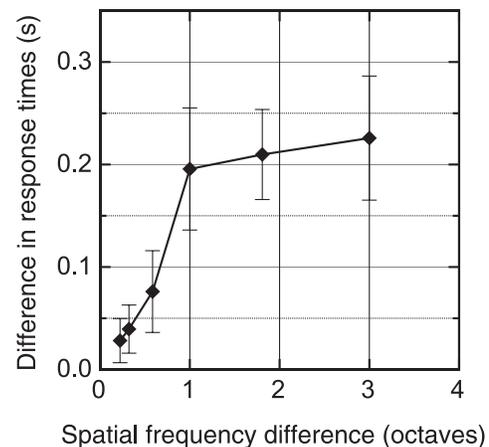
20–50 times the threshold for noticing a spatial frequency difference. These results do not fit with a bottom-up, stimulus-driven mapping of auditory to visual frequency but suggest instead a broadly tuned, categorical relationship.

Experiment 4

In the previous experiments, participants were instructed to use the auditory information. Therefore, it is unclear whether the observed search benefits for the lower and higher frequencies were due to a top-down or bottom-up effect. In this experiment we examine whether the observed search benefits were due to the auditory stimulus capturing attention to the matched Gabor (as proposed by Guzman-Martinez et al., 2012). The search task used was similar to Experiment 2; however, the set size was always fixed (five elements) and the auditory stimulus used was the mean matching response for the lowest visual spatial frequency Gabor from Experiment 1 (amplitude modulation of 1.25 Hz). This was chosen as it produced the largest search benefits in Experiment 2. As the auditory stimulus was the same for all sound-present trials, it matched the target in only one out of five trials. In the remaining trials the auditory signal would be matched to a distractor Gabor. The participants' task was the same as in Experiment 2; however, participants were instructed to ignore the auditory stimulus. As a matching experiment prior to the search task may have alerted participants to a relationship between the auditory signal and the spatial frequency of the Gabors, the search task was completed first and followed by the matching task.

The matching task was used to examine whether the frequency for the amplitude-modulated auditory stimulus used in the search task differed from the responses in the matching task. If automatic capture occurs, then we expect firstly, search benefits when the sound matched the target compared to when the sound matched a distractor, and secondly, search costs when the auditory stimulus matched a distractor Gabor compared to the sound-absent condition (see Theeuwes, 2010, for a review regarding attentional capture).

Method

Participants

Ten new participants (two female, mean age 26 years, range 21 to 31 years) participated in Experiment 4. All participants were naive as to the purpose of the experiment. All reported normal hearing and normal or corrected-to-normal vision.

Stimuli and apparatus

The experimental set-up was the same as Experiment 1. However the visual spatial frequencies used were 0.5, 1.5, 2.5, 3.5, and 4.5 cycles per cm. The frequency of amplitude modulation of the noise was 1.25 Hz.

Design and procedure

The search task was identical to Experiment 2, except that the set size was always five, and in the sound-present condition participants were always presented with 1.25 Hz amplitude-modulated noise. Participants were told to ignore the sound. In half of the trials the sound was present, and in the remaining trials the display was presented in silence. Sound presence was randomly determined. Since the auditory stimulus always matched the lowest spatial frequency, the sound matched the target in only 20% of the trials. In the remaining trials the sound matched a distractor instead. A block of 15 practice trials was given to allow participants to become comfortable with the task, followed by a single block of 280 trials. Following this, participants performed a matching task identical to that of Experiment 1, except that the visual spatial frequencies used for the Gabors were 0.5, 1.5, 2.5, 3.5, and 4.5 cycles per cm.

Results

The overall error rate was low (1.3%), and therefore, no further analysis of errors was conducted. All erroneous trials were excluded from the analysis.

RTs greater than 5 s and those that differed by more than 3 *SD*s from the mean for each participant for each condition were excluded from analysis (0.4%). Figure 7 (upper panel) illustrates the corrected mean RTs as a function of spatial frequency and sound condition.

RTs were subjected to an ANOVA with spatial frequency and the sound condition as within-subject variables. The ANOVA showed significant effect of spatial frequency, $F(4, 36) = 10.4$, $p < 0.01$, as participants responded slower to low frequency Gabors (2297 ms for 0.5 cycles per cm) than higher frequency Gabors (2047, 1990, 1904, and 1925 ms for 1.5, 2.5, 3.5, and 4.5 cycles per cm, respectively). Most importantly, both the main effect of sound condition and the two-way interaction between sound condition and spatial frequency were not significant, $F(1, 9) = 2.9$, $p = 0.12$, and $F(4, 36) = 1.1$, $p = 0.38$, respectively. As such no specific benefit is seen in the sound-present condition for the matching visual spatial frequency; if anything, search times for the matched visual spatial frequency were slower than all the other spatial frequencies in the display. Furthermore, no costs are seen for the sound matching a distractor compared to the sound-absent condition.

The mean matching responses for spatial frequency are shown in the lower panel of Figure 7. There was a highly significant effect of spatial frequency, $F(4, 36) = 71.2$, $p < 0.001$, indicating that participants' auditory matching varied with visual spatial frequency (see also Experiment 1, and Guzman-Martinez et al., 2012). Critically, the mean match for participants for the lowest spatial frequency target was 1.25 Hz, which was equal to the amplitude-modulation rate used in the search task (1.25 Hz, $p < 0.001$), indicating that the absence of search benefits was not due to an incorrect amplitude-modulation rate in the search task.

Discussion

If the amplitude-modulated auditory stimulus produced automatic attentional capture to the matched visual spatial frequency, we would expect participants to respond faster to this matched visual spatial frequency compared to the other unmatched spatial frequencies in the display. The results clearly show that this does not occur. Furthermore, we also expect search costs when the auditory stimulus matched a distractor compared to the no-sound condition, this effect is also absent. These results suggest that the search benefits in Experiment 2 were due to top-down information rather than stimulus-driven effects. Next, we test whether the auditory stimulus must be exact to produce search benefits or if a relative match is sufficient.
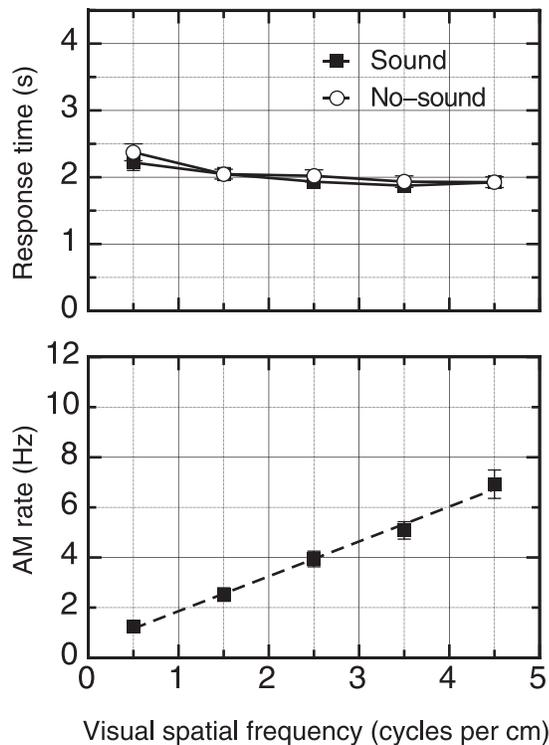
Figure 7. Data from Experiment 4 showing corrected mean RTs ($n = 10$) for sound-present (filled squares) and sound-absent (open circles) in the top panel and the mean matching responses of frequency of auditory amplitude-modulation (Hz) to visual spatial frequency (cycles per cm) in the lower panel. Error bars represent ± 1 SEM.

Experiment 5

Experiment 5 tests the specificity of the audio-visual frequency interaction using a visual search task with auditory signals that are *not* matched to the visual-spatial frequencies. In this task participants are presented with a cluttered display of the same form as Experiment 2. However, the visual spatial frequencies used were limited to the upper half of the range previously tested, 2 to 4.5 cycles per cm. Additionally, the frequencies of auditory signals used extended well above and below the range of perceptual matches. The mean perceptual matches from Experiment 1 for the spatial frequencies 2 to 4.5 cycles per cm ranged from 3.7 to 7.9 Hz, whereas the auditory signals used here ranged from 1 to 16 Hz. Expanding the set of auditory frequencies from a range of roughly 4 Hz to a range of 15 Hz means they were different from the matching data obtained in Experiment 1 and should therefore violate a precise and automatic mapping of auditory to visual frequency.

Despite violating the specific match between auditory and visual frequencies obtained in Experiment 1, Experiment 5 will use a systematic mapping of another kind. Both the visual and auditory signals were ordered from lowest to highest, and then a mapping based on ordinal position was used in all trials. That is, the lowest auditory signal was always paired with the lowest visual signal, and the highest auditory signal was always paired with the highest visual signal. If the effects of the auditory signals are truly specific to the matched visual spatial frequency, as stimulus-level feature mapping would predict, we expect no search benefits will arise from mismatched frequencies. In contrast, if the relative relationship is sufficient to guide attention (i.e., low corresponds with low; high corresponds with high), then we would expect the boundary visual spatial frequencies to show improved search times, just as we observed in Experiment 2 (see Figure 4).

Method

Participants

Nine of the individuals from Experiment 1 participated (three female, mean age 29 years, range 19 to 38 years; two authors).

Stimuli and apparatus

The experimental set-up was the same as Experiment 2. However the visual spatial frequencies used were 2.0, 2.5, 3.0, 3.5, 4.0, and 4.5 cycles per cm and frequencies of amplitude modulation used were 1, 4, 7, 10, 13, and 16 Hz. Visual and auditory stimuli were paired in order of frequency, lowest to highest (i.e., the 2.0 and 4.5 cycles per cm Gabor patches were paired with the 1 and 16 Hz amplitude-modulated auditory signals, respectively).

Design and procedure

The task was identical to Experiment 2, except that the set size was either four or six. In the sound-present condition participants were presented with the auditory signal that was paired (but not perceptually matched) with the target visual spatial frequency. Participants were told the auditory signal was informative and to use it to guide their search. A block of 15 practice trials was given to allow participants to become comfortable with the task, followed by three blocks of 240 trials (720 trials in total).

Results

The overall error rate was low (1.6%) and therefore no further analysis of errors was conducted. All erroneous trials were excluded from the analysis. RTs

greater than 5 s and those that differed by more than 3 *SD*s from the mean for each participant for each condition were excluded from analysis (0.7%). RTs were subjected to an ANOVA with spatial frequency, set size, and the sound condition as within-subject variables. The top left panel of Figure 8 shows corrected mean RTs (collapsed over spatial frequency) as a function of set size and sound condition.

The ANOVA yielded a significant effect for the sound condition, $F(1, 8) = 38.5$, $p < 0.001$, indicating that participants responded faster when the auditory signal was present (1670 ms) than when the auditory signal was absent (2047 ms). In line with Experiment 2, this effect was not consistent across all spatial frequencies tested. The lower panel of Figure 8 illustrates the change in effect over the spatial frequencies tested.

The dependence on spatial frequency was confirmed by a significant two-way interaction between sound condition and spatial frequency in the ANOVA, $F(5, 40) = 14.4$, $p < 0.01$. Separate ANOVAs for each spatial frequency showed reliable sound condition × set-size interactions for the two lowest spatial frequencies of 2.0 and 2.5 cycles per cm ($Fs > 15.1$; $ps < 0.005$) and the highest frequency 4.5 cycles per cm, $F(1, 8) = 5.4$; $p < 0.05$. However, this interaction was not significant for the other spatial frequencies of 3.0, 3.5, and 4.0 cycles per cm ($Fs < 2.01$; $ps > 0.19$).

Discussion

The findings of Experiment 5 demonstrate two important effects. Firstly, the results show that search benefits are not based on the specific matches between vision and audition, but rather that the frequency of the amplitude modulation of the auditory signal allows attention to be guided to visual spatial frequencies that are matched in position in the relative scale in the display (low or high, relative to the range presented). Search benefits are present for the lowest (2 cycles per cm) and highest (4.5 cycles per cm) visual spatial frequencies even though the frequency of amplitude modulation of the auditory signals was well outside of the matched range from Experiment 1. More specifically, the spatial frequency showing the greatest search benefit when it was presented with a 1 Hz amplitude-modulated auditory signal (2 cycles per cm), showed no search benefits in Experiment 2 when it was presented with the individual's perceptually matched amplitude-modulated auditory signal.

The second point revealed by Experiment 5 is that the search benefit arising from pairing auditory and visual frequencies is largest for the boundaries of the display. In Experiment 2, the 2.0 and 2.5 cycles per cm Gabors were displayed amongst Gabors of lower and higher spatial frequency, and in this case neither showed significant search benefits. However in Experiment 5, the 2.0 and 2.5 cycles per cm Gabors were the lowest spatial frequencies in the display, and both showed significant search benefits. The 2.0 cycles per cm Gabor was presented here with an auditory signal of 1 Hz (far from the mean match of 3.7 Hz obtained in Experiment 1), while the 2.5 cycles per cm Gabor was presented here with an auditory signal of 4 Hz, close to the mean perceptual match from Experiment 1 of 4.5 Hz. Experiment 5 supports the findings of Experiments 2 and 3 that interactions between auditory amplitude-modulation rate and visual spatial frequency are broad.

General discussion

The experiments reported here addressed crossmodal interactions between amplitude-modulated auditory stimuli and visual spatial frequency. Experiment 1 supported a recent finding that participants are able to match the frequency of amplitude modulation of auditory noise to visual spatial frequency with a linear relationship (Guzman-Martinez et al., 2012). However, results from Experiments 2, 3, 4, and 5 cast doubt over the claim by Guzman-Martinez, et al. (2012) that this relationship arises from an automatic and specific mapping of auditory and visual frequencies. In these experiments, we used a visual search paradigm to investigate whether the auditory signal can guide attention to the "matched" visual spatial frequency. We found that visual search efficiency for the target spatial frequency with a matching and informative auditory signal was improved for some spatial frequencies, but not for all (Experiment 2). In a simplified two-element search display (Experiment 3), we found the informative auditory signal reduced search times only when there was substantial difference (one octave or more) between the target and distractor spatial frequencies. We then showed that although participants can actively use a match between auditory and visual frequencies, there are no search benefits when the auditory stimulus is uninformative (Experiment 4). Lastly, Experiment 5 suggests that there is no specific mapping of auditory to visual frequencies: It is a relative and variable mapping. Whether a particular auditory frequency guides search to a given visual frequency or not depends on what other frequencies are present in the stimulus set. Taken together, these experiments show that the conditions under which visual selection of spatial frequency can be enhanced by "matched" amplitude-modulated auditory stimuli are both limited and variable. This is incompatible with a specific mapping of auditory and visual frequencies that occurs early and automatically and instead points to a late, flexible, and strategic mapping.
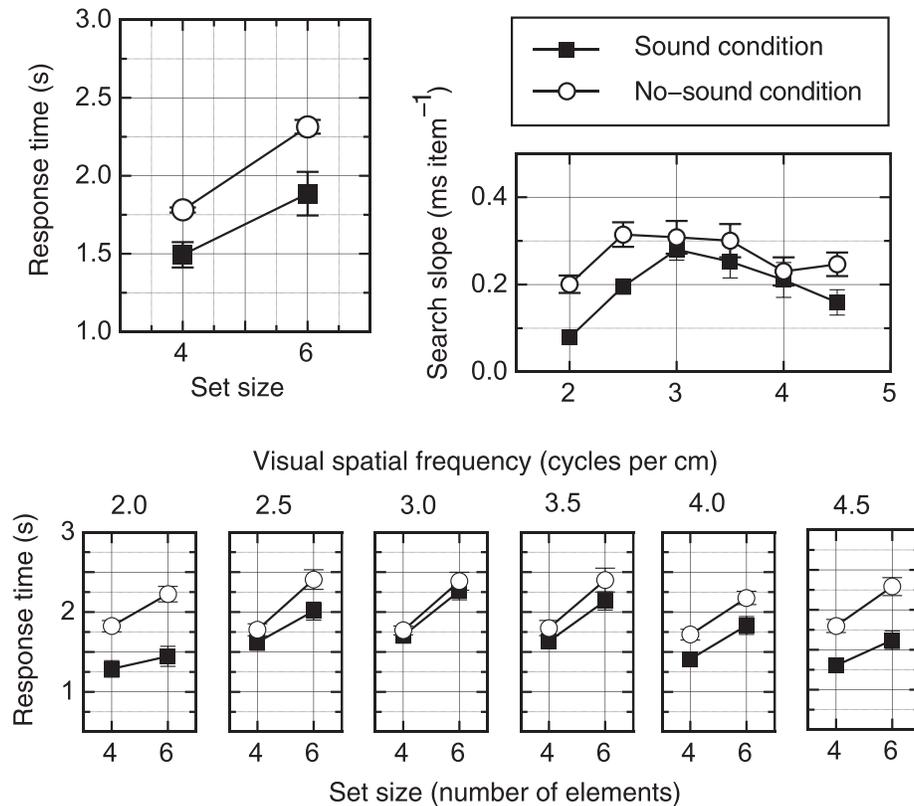
Figure 8. Data from Experiment 5 showing corrected mean RTs ($n = 9$) for each set size, for sound-present (filled squares) and sound-absent (open circles) in the top left panel, and the set-size slope, a measure of search efficiency, is shown for the sound-present and sound-absent conditions in the top right panel. The RTs for both set sizes and sound conditions for each spatial frequency are plotted in the lower half of the figure. Error bars are ± 1 *SEM*.

This is also reflected in the dependency on the starting amplitude-modulation rate demonstrated in Experiment 1, suggesting the crossmodal mapping of auditory to visual frequencies is performed at a broad categorical level rather than reflecting specific stimulus matches.

Although we do find a search advantage for a target spatial frequency from a "matching" auditory frequency, it only occurs for spatial frequencies at the higher and lower ends of the range in the stimulus set, with a gradual decline from the boundary frequencies. It is possible that the fine sampling of spatial frequency in Experiment 2 exacerbated the poor performance seen for intermediate spatial frequencies. It may be that the auditory stimulus captured attention to neighboring spatial frequencies, as the matches were overlapping. However, even if attention were captured to the correct spatial frequency in only one-third of the trials, assuming overlap with two neighboring frequencies, we would still expect search benefits compared to the sound-absent condition. Regardless, the display used in Experiment 4 ensures the auditory stimulus was exclusively matched to one visual item, as there is a 1.5 octave difference, greater than the difference suggested in Experiment 3, between the two lowest spatial

frequencies. Therefore, the absence of an effect in Experiment 4 is strong evidence that automatic capture does not occur.

We suggest that the pattern of results in Experiments 2 and 5 reflects a high-level and strategic approach that exploits the range limits of the auditory stimulus set. Participants can quickly categorize the auditory stimuli at the boundary as being "very low" or "very high" in rate and could use this to guide their search within a smaller subset of items, restricting it to the higher spatial frequency items for a "high" auditory signal (or to the lower frequency items for a "low" auditory signal). By ignoring the opposite end of the range in the display, participants can effectively exclude some of the distractors and find the target more quickly in these boundary conditions, relative to intermediate conditions. Consistent with this, intermediate spatial frequencies did not show a search benefit from a matched auditory signal in Experiment 2, and yet an intermediate frequency was found to support efficient search in Experiment 5, provided it was the lowest spatial frequency in the display.

One key difference between our experiments and those of Guzman-Martinez et al. (2012) concerns the possible influence of attention in their experiment. In

their experiment, both the auditory signal and Gabor patches were presented prior to the phase change of the target by 250 ms, which would allow participants to attend to, or even make eye movements to, the matched Gabor before the target's positional shift occurred. This would result in better performance when the target was presented at the attended location compared to the unattended location (Posner, 1980). Exacerbating this, the two Gabor patches had spatial frequencies that differed enormously (by three octaves: 0.5 and 4.0 cycles per cm). Such a large frequency difference facilitates categorical decision making, as one stimulus is obviously very fine and the other very coarse. Correspondingly, the auditory amplitude-modulation frequencies matched to these widely differing spatial frequencies are also substantially different, one clearly modulating much faster than the other. As a consequence, it would have been easy to categorize the auditory signals as "low" or "high" during the 250 ms prior to the visual change, and then shift attention quickly to the matching visual stimuli, also very evidently "low" or "high" in spatial frequency. This approach may produce data resembling efficient visual search, but it would arise as a simple consequence of categorical perception, rather than a direct, specific mapping of auditory and visual frequencies.

Cross-modal interactions induced by stimulus factors such as spatial and temporal congruence are typically considered automatic and pre-attentive. In a purely visual task, a salient visual stimulus will capture attention even when participants try to ignore it (Theeuwes, 1992). Similarly, in the "pip and pop" effect, a salient auditory cue can capture attention to a synchronously changing visual object even when it is spatially uninformative and participants attempt to ignore it (Van der Burg et al., 2008). Additionally, the ventriloquism effect, or visual bias of auditory location, has been shown to be unaffected by attentional manipulation (Bertelson, Vroomen, de Gelder, & Driver, 2000). However, in a recent review Talsma, Senkowski, Soto-Faraco, and Woldorff (2010) have suggested interplay between attention and crossmodal interactions that is influenced by competition for attention. There is now evidence of attentional modulation of crossmodal interactions thought to be automatic. Selective spatial attention has been shown to enhance the sound-induced flash illusion (Mishra et al., 2010) and the McGurk illusion can be reduced when attentional demands are high using tasks that are visual, auditory (Alsius, Navarra, Campbell, & Soto-Faraco, 2005), or tactile (Alsius, Navarra, & Soto-Faraco, 2007). These findings are also supported by results from an attentional blink paradigm (Van der Burg, Brederoo, Nieuwenstein, Theeuwes, & Olivers, 2010) and a rapid serial visual presentation paradigm (Talsma, Doty, & Woldorff, 2007) showing that

attention to at least one modality was required in order to trigger crossmodal interactions. Considering this, the failure to find search benefits for a matched auditory stimulus when uninformative (in Experiment 4), and a matched auditory stimulus for intermediate spatial frequencies when informative (Experiment 2), is inconsistent with low-level stimulus-driven crossmodal interactions.

In conclusion, the results reported here provide some support for previous findings that participants are capable of matching frequencies of amplitude-modulated noise to specific visual spatial frequencies. Our results demonstrate that visual selection of one spatial frequency from another can be aided by an amplitude-modulated auditory stimulus; however, the effects are limited. Experiment 2 demonstrated that informative auditory stimuli improve performance for some spatial frequencies. Experiment 3 revealed that a substantial difference between the target and distractor spatial frequency is required, and Experiment 4 showed that an uninformative auditory stimulus has no influence on response times to a matched visual spatial frequency. Finally, Experiment 5 showed the matches are flexible and based on relative not specific frequency matches. These findings are unlikely to reflect stimulus-driven interactions and instead reflect top-down attentional guidance based on a broad framework of matched categories.

Acknowledgments

Commercial relationships: none.
Corresponding author: David Alais.
Email: david.alais@sydney.edu.au.
Address: Brennan MacCallum (A18), School of Psychology, University of Sydney, Sydney, New South Wales, Australia.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257–262, doi:10.1016/j.cub.2004.01.029.

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing & Perceiving*, *23*(1), 3–38, doi:10.1163/187847510X488603.

Alais, D., van Boxtel, J. J., Parker, A., & van Ee, R. (2010). Attending to auditory signals slows visual alternations in binocular rivalry. *Vision Research*, *50*(10), 929–935, doi:10.1016/j.visres.2010.03.010.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839–843, doi:10.1016/j.cub.2005.03.046.

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, *183*(3), 399–404, doi:10.1007/s00221-007-1110-1.

Arrighi, R., Marini, F., & Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of Vision*, *9*(4):25, 1–7, http://www.journalofvision.org/content/9/4/25, doi:10.1167/9.4.25. [PubMed] [Article]

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A: Optics, Image Science, & Vision*, *20*(7), 1391–1397, doi:10.1364/JOSAA.20.001391.

Belopolsky, A. V., Zwaan, L., Theeuwes, J., & Kramer, A. F. (2007). The size of an attentional window modulates attentional capture by color singletons. *Psychonomic Bulletin & Review*, *14*(5), 934–938, doi:10.3758/BF03194124.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, *50*(1–2), 147–155, doi:10.1016/S0167-8760(03)00130-2.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*(2), 321–332, doi:10.3758/BF03205552.

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2004). 'Acoustical vision' of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, *160*(3), 273–282, doi:10.1007/s00221-004-2005-z.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences USA*, *102*(51), 18 751–18 756, doi:10.1073/pnas.0507704102.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, *2*(7), 254–262.

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2009). Dual mechanisms for the cross-sensory spread of attention: How much do learned associations matter? *Cerebral Cortex*, *20*(1), 109–120, doi:10.1093/cercor/bhp083.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), 332–343, doi:10.1007/s00221-002-1262-y.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778, doi:10.1038/nn1268.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*(5), 473–490, doi:10.1162/089892999563544.

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. *Current Biology*, *22*(5), 383–388, doi:10.1016/j.cub.2012.01.004.

Hidaka, S., Manaka, Y., Teramoto, W., Sugita, Y., Miyauchi, R., Gyoba, J., et al. (2009). Alternation of sound location induces visual motion perception of a static object. *PLoS One*, *4*(12), e8188, doi:10.1371/journal.pone.0008188.

Hirsch, J., & Hylton, R. (1982). Limits of spatial-frequency discrimination as evidence of neural interpolation. *Journal of the Optical Society of America*, *72*(10), 1367–1374, doi:10.1364/JOSA.72.001367.

Holcombe, A. O., & Seizova-Cajic, T. (2008). Illusory motion reversals from unambiguous motion with visual, proprioceptive, and tactile stimuli. *Vision Research*, *48*(17), 1743–1757, doi:10.1016/j.visres.2007.05.019.

Los, S. A., & Van der Burg, E. (in press). Sound speeds vision through preparation, not integration. *Journal of Experimental Psychology: Human Perception & Performance*, doi:10.1037/a0032183.

Mishra, J., Martinez, A., & Hillyard, S. A. (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash

illusion. *Journal of Cognitive Neuroscience*, *22*(8), 1714–1729, doi:10.1162/jocn.2009.21295.

Molholm, S., Martinez, A., Shpaner, M., & Foxe, J. J. (2007). Object-based attention is multisensory: Co-activation of an object's representations in ignored sensory modalities. *European Journal of Neuroscience*, *26*(2), 499–509, doi:10.1111/j.1460-9568.2007.05668.x.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, *14*(4), 452–465, doi:10.1093/cercor/bhh007.

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. *Cognitive Brain Research*, *14*(1), 115–128, doi:10.1016/S0926-6410(02)00066-6.

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., et al. (2005). Grabbing your ear: Rapid auditory–somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex*, *15*(7), 963–974, doi:10.1093/cercor/bhh197.

Ngo, M. K., & Spence, C. (2010). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics*, *72*(6), 1654–1665, doi:10.3758/APP.72.6.1654.

Noesselt, T., Bergmann, D., Hake, M., Heinze, H.-J., & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Research*, *1220*, 157–163. doi:10.1016/j.brainres.2007.12.060.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*(6753), 584–587, doi:10.1038/44134.

Olivers, C. N. L., & Van der Burg, E. (2008). Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Research*, *1242*, 191–199, doi:10.1016/j.brainres.2008.01.070.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25, doi:10.1080/00335558008248231.

Regan, D., Bartol, S., Murray, T. J., & Beverley, K. I. (1982). Spatial frequency discrimination in normal vision and in patients with multiple sclerosis. *Brain*, *105*(Pt 4), 735–754, doi:10.1093/brain/105.4.735.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound

alters visual motion perception. *Nature*, *385*(6614), 308–308, doi:10.1038/385308a0.

Stein, B. E., Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioral indices of multi-sensory integration: Orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience*, *1*(1), 12–24, doi:10.1162/jocn.1989.1.1.12.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, *17*(3), 679–690, doi:10.1093/cercor/bhk016.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410, doi:10.1016/j.tics.2010.06.008.

Teramoto, W., Hidaka, S., Sugita, Y., Sakamoto, S., Gyoba, J., Iwaya, Y., et al. (2012). Sounds can alter the perceived direction of a moving visual object. *Journal of Vision*, *12*(3):11, 1–12, http://www.journalofvision.org/content/12/3/11, doi:10.1167/12.3.11. [PubMed] [Article]

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606. doi:10.3758/BF03211656

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*(2), 77–99, doi:10.1016/j.actpsy.2010.02.006.

Van der Burg, E., Brederoo, S. G., Nieuwenstein, M. R., Theeuwes, J., & Olivers, C. N. L. (2010). Audiovisual semantic interference and attention: Evidence from the attentional blink paradigm. *Acta Psychologica*, *134*(2), 198–205, doi:10.1016/j.actpsy.2010.01.010.

Van der Burg, E., Cass, J., Olivers, C. N., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS One*, *5*(5), e10664, doi:10.1371/journal.pone.0010664.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception & Performance*, *34*(5), 1053–1065, doi:10.1037/0096-1523.34.5.1053.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2009). Poke and pop: Tactile-visual synchrony increases visual saliency. *Neuroscience Letters*, *450*(1), 60–64, doi:10.1016/j.neulet.2008.11.002.

Van der Burg, E., Olivers, C. N. L., & Theeuwes, J. (2012). The attentional window modulates capture by audiovisual events. *PLoS One*, *7*(7), e39137, doi:10.1371/journal.pone.0039137.

Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *Neuroimage*, *55*(3), 1208–1218, doi:10.1016/j.neuroimage.2010.12.068.

van Ee, R., van Boxtel, J. J. A., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *Journal of Neuroscience*, *29*(37), 11 641–11 649, doi:10.1523/JNEUROSCI.0873-09.2009.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception & Performance*, *26*(5), 1583–1590, doi:10.1037//0096-1523.26.5.1583.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Research: Cognitive Brain Research*, *22*(1), 32–35, doi:10.1016/j.cogbrainres.2004.07.003.

Zannoli, M., Cass, J., Mamassian, P., & Alais, D. (2012). Synchronized audio-visual transients drive efficient visual search for motion-in-depth. *PLoS ONE*, *7*(5), e37190, doi:10.1371/journal.pone.0037190.