Non-spatial sounds regulate eye movements and enhance visual search

Heng Zou
Allgemeine und Experimentelle Psychologie, Ludwig-Maximilians-Universtät München, Munich, Germany
Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universtät München, Martinsried, Germany

Hermann J. Müller
Allgemeine und Experimentelle Psychologie, Ludwig-Maximilians-Universtät München, Munich, Germany
School of Psychological Science, Birkbeck College (University of London), London, United Kingdom

Zhuanghua Shi
Allgemeine und Experimentelle Psychologie, Ludwig-Maximilians-Universtät München, Munich, Germany

Spatially uninformative sounds can enhance visual search when the sounds are synchronized with color changes of the visual target, a phenomenon referred to as "pip-and-pop" effect (van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008). The present study investigated the relationship of this effect to changes in oculomotor scanning behavior induced by the sounds. The results revealed sound events to increase fixation durations upon their occurrence and to decrease the mean number of saccades. More specifically, spatially uninformative sounds facilitated the orientation of ocular scanning away from already scanned display regions not containing a target (Experiment 1) and enhanced search performance even on target-absent trials (Experiment 2). Facilitation was also observed when the sounds were presented 100 ms prior to the target or at random (Experiment 3). These findings suggest that non-spatial sounds cause a general freezing effect on oculomotor scanning behavior, an effect which in turn benefits visual search performance by temporally and spatially extended information sampling.

Keywords: visual search, eye movement, audiovisual integration, intersensory enhancement

Citation: Zou, H., Müller, H. J., & Shi, Z. (2012). Non-spatial sounds regulate eye movements and enhance visual search. *Journal of Vision, 12*(5):2, 1–18, http://www.journalofvision.org/content/12/5/2, doi: 10.1167/12.5.2.

Introduction

Our brain continuously receives sensory input from the external world, including, in particular, visual and auditory signals. When an external object or event elicits multimodal signals simultaneously, such as synchronous audiovisual signals, may be easily picked out by our brain from amongst the other objects or events in the environment. For example, finding your friend in a crowd may become easier when the friend not only waves to you, but also calls your name loudly. Such an enhancement of visual search performance may come about as result of redundant target coding (Krummenacher, Müller, & Heller, 2002) or of an alerting effect exerted by the auditory cue (Posner & Petersen, 1990). Facilitation of visual search by auditory orienting has been demonstrated in various paradigms in which a visual target was accompanied by a sound signal presented at the same location (Bolia, D'Angelo, & McKinley, 1999; Doyle & Snowden, 1998; Perrott, Saberi, Brown, & Strybel, 1990; Perrott, Sadralodabai, Saberi, & Strybel, 1991). For example, Doyle and Snowden (1998) found that simultaneous, spatially congruent sound facilitated covert orienting to non-salient visual targets in a conjunction search paradigm. Recent studies have also reported that audiovisual interaction can enhance visual detection (McDonald, Teder-Salejarvi, & Hillyard, 2000; Shi, Chen, & Müller, 2010; Vroomen & de Gelder, 2000) and visual search performance as a result of enhancing visual salience (van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010; van der Burg et al., 2008; van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011). For instance, using response time (RT) and signal detection measures, McDonald et al. (2000) examined whether involuntary orienting of attention to a sudden sound stimulus would influence the perceptual or post-perceptual processing of a subsequent visual stimulus appearing nearby. They found that the sound improved

doi: 10.1167/12.5.2 Received August 16, 2011; published May 4, 2012 ISSN 1534-7362 © 2012 ARVO

the detectability of a subsequent flash appearing at the same location, they concluded that involuntary orienting of attention to the sound could enhance early perceptual processing of visual stimuli.

Interestingly, intersensory enhancement of visual perception and search performance has been found not only with spatially informative auditory stimuli, but also with spatially uninformative but temporally informative auditory signals (van der Burg et al., 2010; van der Burg et al., 2008; Vroomen & de Gelder, 2000) or tactile signals (van der Burg, Olivers, Bronkhorst, & Theeuwes, 2009). For example, Vroomen and de Gelder (2000) investigated cross-modal influences from the auditory onto the visual modality at an early level of perceptual processing. In their study, a visual target was embedded in a rapidly changing sequence of visual distractors. They found a high tone embedded in a sequence of low tones to improve the detection of a synchronously presented visual target, while this enhancement was reduced or abolished when the high tone was presented asynchronously to the visual target or became part of a melody. In contrast to spatially informative auditory stimuli, spatially uninformative sound cannot influence performance by (automatic) orienting of attention to the location of the visual target; rather, it acts by enhancing the perceptual grouping of temporally close multisensory events (Chen, Shi, & Müller, 2010, 2011; Spence, Sanabria, & Soto-Faraco, 2007). Using a dynamic visual search paradigm, Van der Burg et al. (2008) demonstrated that irrelevant beeps could guide visual attention towards the location of a synchronized visual target, which, if presented without such synchronous beeps, was extremely hard to find. In their experiments, participants had to search for a horizontal or a vertical target bar among oblique distractor bars. Both the target and distractors were either green or red, and changed their color randomly in a pre-determined (1.1-Hz) cycle. Due to the cluttered and heterogeneous search display, finding a target was extremely difficult. However, with the aid of synchronous beeps, search performance was improved substantially (in fact, in the order of seconds). Van der Burg et al. referred to this facilitation as "pip-and-pop" effect. In their follow-up experiments (van der Burg et al., 2008), they ruled out an explanation of this effect in terms of general alerting, and suggested that the facilitation was due to automatic audiovisual integration, generating a relatively salient (visual) bottom-up feature capable of summoning attention automatically.

In a recent study, they found synchronous audiovisual events to enhance ERP amplitudes, compared to unisensory (beep alone plus target color change alone) events, over left parieto-occipital cortex, as early as 50–60 ms post-stimulus onset, and this enhancement was correlated with behavioral benefits in the accuracy of the target discrimination. From this, they argued that the early multisensory interaction played a crucial role for the "pip-and-pop" effect. Note, though, that van der Burg et al. used unspeeded responses in their discrimination task, while the pip-and-pop effect is usually measured in terms of response time (RT) facilitation. Given that the overall RTs in the standard pip-and-pop search paradigm (van der Burg et al., 2008), and the RT facilitation by the beep events, are of the order of seconds, arguably, it remains unclear how much the early audiovisual enhancement found by van der Burg et al. (2011) does actually contribute to the overall RT facilitation effect. In addition, as reported by van der Burg et al. (2008), RT facilitation was evident even under conditions of audiovisual asynchrony (of as much as 100 ms). This finding may not be well explained by the early multisensory modulation (50–60 ms post stimulus) reported by van der Burg et al. (2011).

Given this one alternative, and more direct, way to examine how the pip-and-pop effect is actually brought about in the standard paradigm derives from the examination of saccadic eye movements. Since eye movement analysis can reveal foveal information processing and saccade planning, it has been widely adopted as a tool for examining overt (and linked to them, *covert*) attention shifts in visual search. In fact, studies have demonstrated that the simultaneous presentation of audiovisual stimuli reduces initial (express) saccade latencies (Colonius & Arndt, 2001; Corneil, Van Wanrooij, Munoz, & Van Opstal, 2002), with the reduction even violating the so-called "race model inequality," indicative of early audiovisual integration prior to saccade initiation (Colonius & Arndt, 2001; Hughes, Nelson, & Aronchick, 1998). In one experiment (Experiment 4b), van der Burg et al. (2008) found the pip-and-pop effect to be also evident in the absence of eye movements. In this situation, covert attention triggered by audiovisual integration might well precede any overt ocular movement. Thus, the pip-and-pop effect without eye movements may share similar mechanisms to those elaborated in the eye movement literature (Doyle & Snowden, 1998). Note, however, that the "express saccades" mentioned above are generally observed in a very simple visual display examining the first saccade only, and the latency reduction (for the first saccade) is typically less than 100 ms By contrast, in the typical pip-and-pop visual search task, the facilitation effect by the synchronous beep is of the magnitude of seconds. As proposed by van der Burg and colleagues, the auditory signal may be rapidly relayed to (early) visual cortex, allowing it to interact with a synchronized visual event and generally boost the saliency of visual signals (van der Burg et al., 2008). In more detail, the saliency boost is assumed to be spatially non-specific; but because the auditory signal coincides with a target color change (but not distractor changes), a multiplicative effect (as assumed by van der Burg et al., 2008) on visual salience would boost the saliency signal for the target (which is primarily defined by orientation contrast to the distractors) more than those of any distractors—potentially making the target "pop out" more rapidly. Thus, it is possible that the auditory enhancement of saliency computations reduces the saccade latency at every beep and that the pip-and-pop effect is the product (i.e., the sum) of these time savings.

However, there is an alternative explanation in terms of a "wait-at-beep" strategy, which might be adopted by the participants. It has been found that participants may operate a "sit-and-wait" strategy in a unimodal visual search, in particular with dynamic changes (motion) of the search items (Geyer, Von Mühlenen, & Müller, 2007; von Mühlenen, Müller, & Müller, 2003). Although waiting at beeps may increase saccadic latencies at beep events, with temporally extended fixations, there is more time for effective information intake. Accordingly, the next saccade may become more efficient, for instance, owing to an expanded attentional spotlight (i.e., a wider range over which covert search processes operate) in wait-at-beep fixations. The underlying assumption here is that serial visual search involves successive (fixation) episodes in which "clumps" of items within a certain display region are processed in parallel (Pashler, 1987). There is evidence that in dynamic searches involving eye movements, information is sampled predominantly in the forward direction of the saccadic scanning path; that is, the spotlight (or covert information sampling) is skewed towards the general movement direction. While this sampling pattern is most prominent in reading (McConkie & Rayner, 1976), it also does apply to visual search, where a sequence of saccades may be planned ahead (Baldauf & Deubel, 2008) and inhibition-of-return processes may instantiate a bias against resampling of already scanned (and rejected) stimulus locations (Müller & von Mühlenen, 2000; Peterson, Kramer, Wang, Irwin, & McCarley, 2001). While such scanning processes may drive normal oculomotor behavior (in periods without beep) in the pip-and-pop paradigm, when a beep occurs, the current or the next fixation may be extended to broaden visual sampling, permitting emergent saliency signals to be picked up over a wider region. Note that such "wait-at-beep" behavior may be invoked automatically. As anecdotally reported by Vroomen and de Gelder (2000), some observers felt the visual target was frozen when a synchronous tone occurred together with the target. This "freezing effect" may well be related to prolonged, covert attentional processing during extended fixation periods in the visual identification task.

To further examine the effects of spatially uninformative sound on visual search and the underlying mechanisms, in the present study, we adopted the pip-and-pop paradigm and measured eye movements. Furthermore, we introduced an informative spatial (central-arrow) cue for top-down attentional guidance in Experiment 1 to be able to disentangle different components of the pip-and-pop effect. By presenting observers with advance cues indicating the side of the display likely to contain the target, we expected a top-down cueing effect; that is, valid cues *ought* to improve visual search performance. If the facilitative effect of the non-spatial auditory signal on visual search performance is purely owing to audiovisual target integration, one would expect no interaction of sound presence (vs. absence) with top-down attentional guidance; that is, the facilitation induced by the sound should be the same regardless of the cue validity. However, if the auditory signals can influence cue-induced attentional guidance (e.g., facilitation of attentional orienting away from already scanned display regions in the invalid-cue condition), one would expect an interaction effect between the cuing (valid, invalid) and sound presence (present, absent) conditions. In Experiment 2, we introduced a target-absent condition (in addition to target-present trials) and focused on the relationship between target presence and facilitation by the uninformative sounds. If regular beeps can regulate oculomotor behavior, one would expect a facilitative effect of the auditory signal to be also manifest on target-absent trials, even though there is no (audio-) visual target in this condition. To further disentangle sound-induced regulation of oculomotor scanning from enhancement by audiovisual interaction, we varied the temporal relationship between the auditory and visual events from synchrony to random asynchrony in Experiment 3. If oculomotor regulation is the key factor (rather than an audiovisual interaction dependent on synchrony of the auditory and visual events), one should observe a facilitation effect even under the random-sound condition.

Experiment 1

Method

Participants

Eight right-handed observers (four females, mean age 24.9 years) with normal or corrected-to-normal visual acuity and normal hearing participated in the experiment. They gave written informed consent in accordance with the Declaration of Helsinki (2008) and were paid for their participation. Before the formal experiment, they performed a block of (up to 40) practice trials to become fully familiarized with the task.

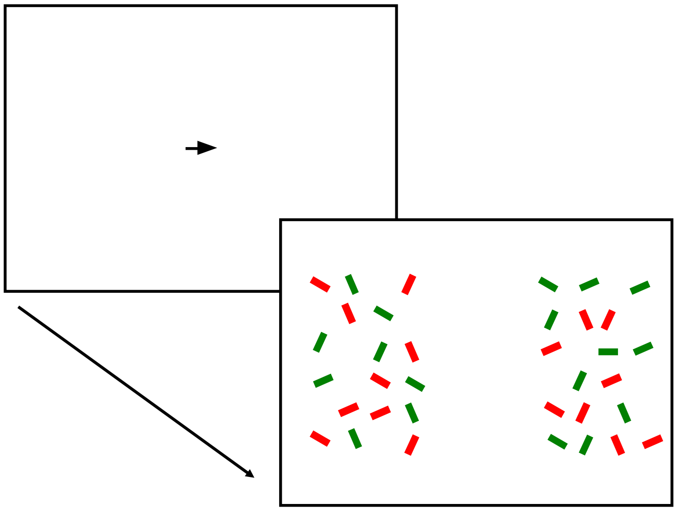


Figure 1. Schematic illustration of a trial in Experiment 1. A central arrow cue was presented prior to the search display, which contained the target, either a horizontal or a vertical bar, among 35 oblique distractors bars. The colors of the items (green or red) were randomly assigned and changed randomly over time. Presentation of the search display was accompanied by repeated mono-tone beeps, in half the trials, which were synchronized with the onset of target color changes (see Method section for further details).

Apparatus and stimuli

The experiment was conducted in a dimly lit cabin. Eye movements were tracked and recorded using an eye tracker device (Eyelink 1000 desktop-mounted system), which communicated with the experimental PC via Matlab using the Psychophysics and Eyelink Toolbox extension (Brainard, 1997; Cornelissen, Peters, & Palmer, 2002; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Visual stimuli were presented on a 21-inch CRT monitor at refresh rate of 100 Hz. We adopted the dynamic search display used in the original pip-and-pop study (van der Burg et al., 2008), but divided the search display into the left and the right regions. In addition, we introduced a pre-cue display. In the pre-cue display (see Figure 1), a black central arrow (CIE $x = 0.32$, $y = 0.34$, subtending 0.73° × 0.64°, 75.8 cd/m^2) was used as a visual spatial cue. The dynamic search display consisted of 36 items, 1 target and 35 distractors. The target was either a horizontal (subtending 0.73° × 0.17°) or a vertical bar (subtending 0.17° × 0.73°). Distractors were oblique bars of the same size as the target, tilted randomly to one of four possible orientations (22.5°, 67.5°, 112.5°, 157.5°). Items were randomly assigned a green color (CIE $x =$ 0.30, $y = 0.60$, 61.3 cd/m^2) or red color (CIE $x = 0.65$, $y =$ 0.34, 15.8 cd/m^2). The left or the right region of search display subtended 5.2° × 8.2° (3.74° horizontally away from the center) and each contained 18 items randomly distributed within a virtual 4 × 6 matrix (each cell size of the matrix subtended 1° × 1°, with random jittering of 0.35°). An example search display is shown in Figure 1. Using similar temporal settings to those adopted in the previous study (van der Burg et al., 2008), a random number of items in the search display dynamically switched color between green and red in randomly generated cycles. Each cycle contained 9 intervals, which varied randomly between 50, 100, and 150 ms, with the following constraints: (i) all intervals occurred equally often within each cycle; (ii) the target changed color only once in each cycle (i.e., 1.1 Hz of target color change); and (iii) the target color change was preceded by an interval of 150 ms and followed by an interval of 100 ms In the remaining intervals, a random number of distractors (between one and three) changed their colors between red and green. Meanwhile, an auditory mono-beep (60 ms, 44.1 kHz, 68.4 dB) was synchronized with each onset of target color change, and it was delivered through earphones in half the trials. The other half of the trials had no auditory stimuli; these trials served as the baseline. Responses and reaction times (RTs) were collected via a parallel-port keypad.

Design and procedure

Two factors, pre-cue and synchronous sound, were examined in a 2 × 2 full-factorial within-subject design. The central arrow presented in the pre-cue display pointed to the correct target side with 80% validity. In half of the trials, the visual search display was accompanied by beeps, which were synchronized with the onset of target color changes. In the other half, there was a visual presentation only. Target orientations (vertical vs. horizontal) were balanced and randomly mixed across trials. There were eight blocks, each consisting of 40 trials, yielding a total of 320 trials.

Participants sat in front of the monitor, at a viewing distance of 80 cm. This distance was maintained with the aid of a chin rest, which also served to stabilize participants' heads. Before the formal experiment, the eye tracker was calibrated for the observer's dominant eye, and a block of practice trials was administered. At the beginning of each trial, a central arrow was presented as a visual cue for 1000 ms. Immediately after the cue offset, the search display was presented (see Figure 1), and it dynamically changed item colors until the observer made a response. Participants were instructed to search for the visual target (horizontal or vertical bar) freely and to make a key press response to indicate the type of target (horizontal or vertical) as soon as they found it, regardless of sound presentation.

Results and discussion

Although the visual search task was difficult, the mean response accuracy was high (98.7%). Mean RTs for correct and incorrect trials were examined in an ANOVA with the single factor correct/incorrect response. This

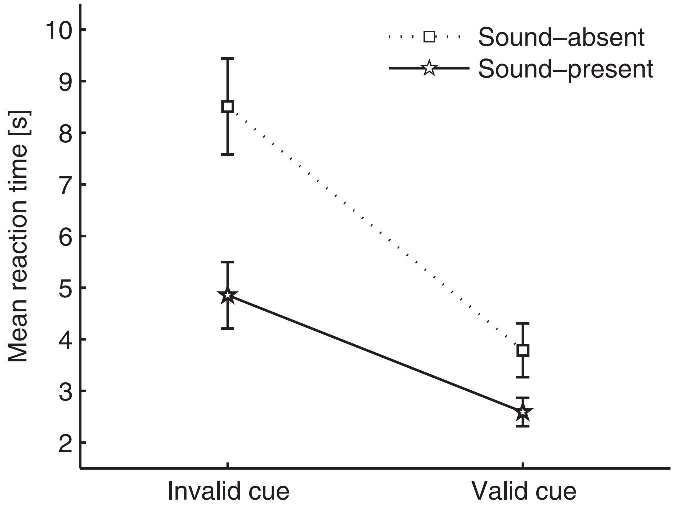


Figure 2. Mean reaction times (±*SE*) in seconds as a function of cue validity and sound presence; stars (solid line) and squares (dotted line) represent the sound-present and sound-absent conditions, respectively.

ANOVA revealed no RT "facilitation" for error versus correct trials, $F(1, 7) = 0.39$, $p = 0.55$, $\eta_p^2 = 0.07$; that is, there was no evidence of a speed versus accuracy trade-off in search task performance. Consequently, in the subsequent analysis, only correct trials were included.

Reaction time effects

Individual mean reaction times (RTs) were estimated for each variable combination, excluding error responses. Figure 2 presents the mean correct RTs averaged across participants. RTs were then submitted to a repeated-measures ANOVA with cue validity and sound presence as factors. The main effect of cue validity was significant, $F(1, 7) = 56.34$, $p < 0.01$, $\eta_p^2 = 0.89$. As expected, RTs on valid-cue trials were faster, by 3.49 s on average, than RTs on invalid-cue trials. The main effect of sound presence was also significant, $F(1, 7) = 19.25$, $p < 0.01$, $\eta_p^2 = 0.73$. Search performance was on average 2.4 s faster on trials with synchronous beeps. This result replicates the pip-and-pop effect, indicating that synchronous beeps facilitate visual search performance. Interestingly, there was a significant interaction between cue validity and sound presence, $F(1, 7) = 11.51$, $p < 0.05$, $\eta_p^2 = 0.62$. The search benefits induced by the accompanying synchronous beeps were larger for invalid trials (mean: 4.72 s) compared to valid trials (mean: 2.26 s).

Oculomotor effects

To further explore dynamic search behavior, we examined all fixations and saccades made during the search. Mean numbers of fixations are shown in Figure 3a. The pattern is similar to the mean RTs (Figure 2) – as also confirmed by a repeated-measures ANOVA, which, similar to the RT ANOVA, yielded a significant main effect of sound presence, $F(1, 7) = 21.42$, $p < 0.01$, $\eta_p^2 = 0.75$; a significant main effect of cue, $F(1, 7) = 56.94$, $p < 0.01$, $\eta_p^2 = 0.89$; as well as a significant interaction between the two factors, $F(1, 7) = 15.54$, $p < 0.01$, $\eta_p^2 = 0.69$. This pattern indicates that the synchronous sounds facilitated visual search in general by permitting participants to plan more effective saccades, and this facilitation was more pronounced when the cue was invalid. To explore the latter effect further, we separated fixations on the target side from those on the non-target side (see Figure 3b). The significant interaction between cue validity and sound presence was largely due to the non-target side, $F(1, 7) = 13.09$, $p < 0.01$, $\eta_p^2 = 0.65$, rather than the target side, $F(1, 7) = 0.29$, $p = 0.61$, $\eta_p^2 = 0.04$. That is, for the non-target side, sound presence reduced the number of saccades (on invalid trials) dramatically (from 17.5 to 8.1 saccades), indicating that the synchronous sound effectively guided saccades to the valid target side.

To examine how participants managed to minimize their number of saccades (and, thus, fixations), we compared the mean fixation durations among each sound and cue condition. A repeated-measures ANOVA of the mean fixation durations (presented in Figure 4a) revealed the main effect of sound to be significant, $F(1, 7) = 6.22$, $p < 0.05$, $\eta_p^2 = 0.47$, but not that of cue validity, $F(1, 7) = 0.23$, $p = 0.64$, $\eta_p^2 = 0.03$. Mean fixation duration was 137 ms longer on trials with sound than on those without sound. However, in contrast to the manual RTs and the number of fixations, there was no interaction between cue validity and sound presence, $F(1, 7) = 2.30$, $p = 0.17$, $\eta_p^2 = 0.25$.

In order to further explore the fixation pattern during the dynamic search, we re-categorized the fixations into three types: fixations accompanied by a beep, fixations without beep but on a trial with sound presence, and fixations on trials without beeps. The mean durations for these three types of fixation are depicted in Figure 4b. A repeated-measures ANOVA with two factors fixation type and cue validity, revealed no significant effect of cue validity, $F(1, 7) = 1.13$, $p = 0.32$, $\eta_p^2 = 0.14$. However, there were significant differences among the three types of fixation, $F(2, 14) = 14.45$, $p < 0.01$, $\eta_p^2 = 0.67$. Bonferroni tests revealed the mean fixation duration to be significantly longer when the fixation was accompanied by a beep than for the other two types of fixation (both $p < 0.05$), while the durations did not differ between the latter two types, $p = 0.11$. The fixation duration was, on average, extended by 440 ms when the fixation was accompanied by a beep relative to the mean of the other two conditions. Furthermore, the interaction between cue

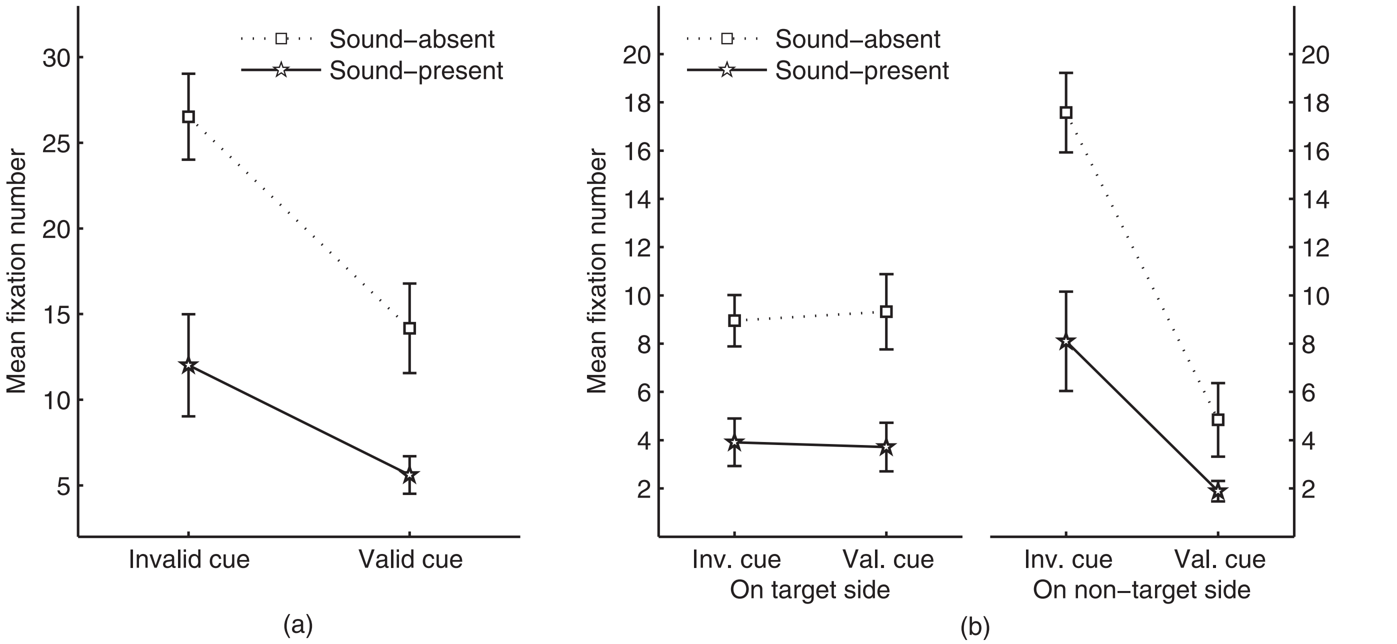
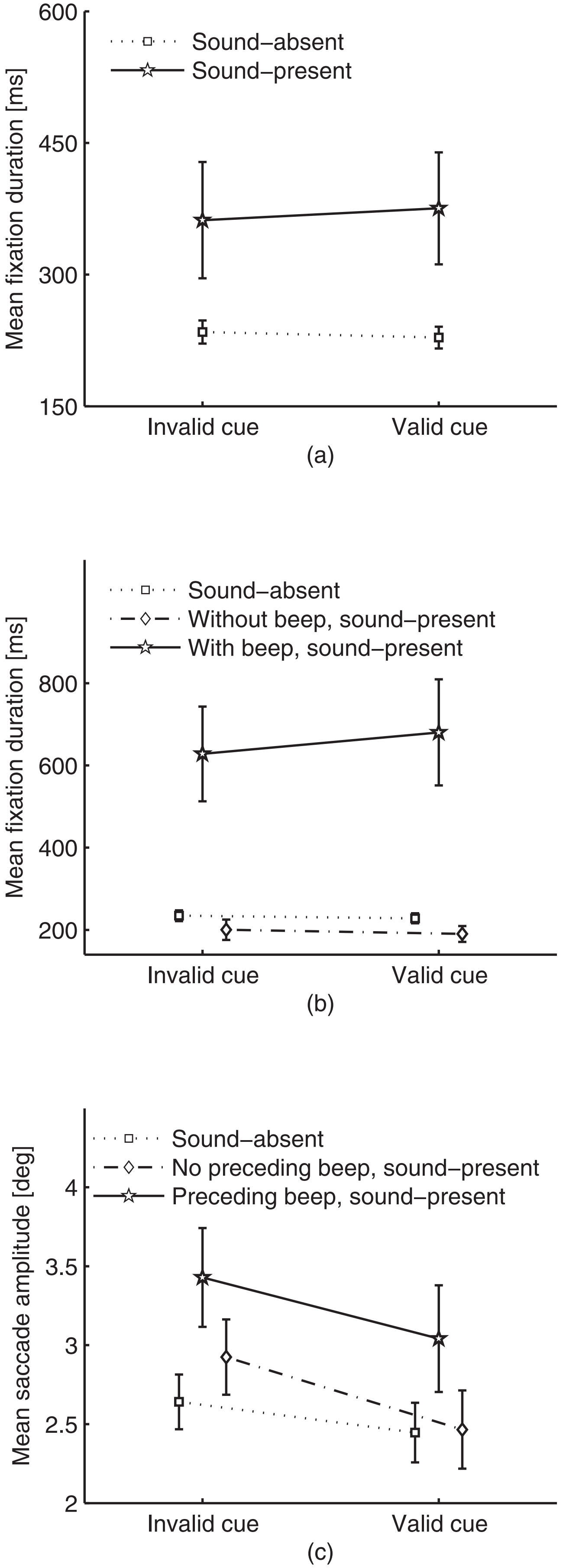


Figure 3. (a) Mean number of fixations (±*SE*) as a function of cue validity and sound presence; stars (solid line) and squares (dotted line) represent the sound-present and sound-absent conditions, respectively. (b) Mean number of fixations (±*SE*) as a function of cue validity and sound presence, separately for fixations on the target side and those on the non-target side.

validity and fixation type was near-significant, $F(2, 14) = 3.64$, $p = 0.05$, $\eta_p^2 = 0.34$. This result was mainly due to the somewhat longer fixations in the valid-cue, compared to the invalid-cue, condition for fixations with beeps (680.2 ms vs. 627.9 ms).

Similarly, we categorized saccades into three types: saccades with a preceding beep; saccades without a preceding beep in the sound-present condition; and saccades in the sound-absent condition. The saccade amplitudes for these three types of saccades are shown in Figure 4c. A repeated-measures ANOVA with two factors, saccade type and cue validity, revealed significant differences among the three types of saccade, $F(2, 14) = 9.83$, $p < 0.01$, $\eta_p^2 = 0.58$. Follow-on Bonferroni tests showed that saccades preceded by a beep had larger amplitudes than the other two types of saccade (both $p < 0.05$), while the mean amplitudes of the latter two types did not differ from each other ($p = 1.00$). The larger amplitudes of saccades following a beep event may be related to the longer duration of fixations accompanied by a beep (see *fixation duration results* above, and discussion below). There was also a significant main effect of cue validity, $F(1, 7) = 8.81$, $p < 0.05$, $\eta_p^2 = 0.56$. Saccades were of greater amplitudes in the invalid-cue condition, due to an increased number of crossing saccades between the left and right sides of the display (Figure 3). Again, there was no interaction effect between saccade type and cue validity, $F(2, 14) = 0.42$, $p = 0.67$, $\eta_p^2 = 0.06$.

In summary, combining spatial cuing with audiovisual dynamic search displays, we replicated a substantial auditory facilitation effect on visual search performance, as described in previous studies (van der Burg et al., 2010; van der Burg et al., 2008). More interestingly, however, we found larger benefits of synchronous sounds in the invalid-cue condition. By examining the oculomotor behavior, we found a similar interaction pattern in the number of fixations: that is, the benefit, in terms of a reduced number of fixations, was greater for the invalid- compared to the valid-cue condition. Surprisingly, further analyses revealed that the interaction between cue validity and sound presence mainly originated from the non-target side; that is, when synchronous beeps were presented, participants were able to “reject” the wrong (non-target) side faster and direct their search across to the target side. In more detail, for invalid trials, participants took on average only about 6.6 s to first reject the wrong side when synchronous beeps were presented (equivalent to, on average, 7.4 target color change events); by contrast, they took about 12.2 s to first switch from the wrong side in the sound-absent condition (i.e., on average 13.6 target color change events). In other words, participants did not change fixation side immediately after the first beep; rather, switching sides took multiple beeps: participants kept scanning the currently searched side of the display for a while (but for a shorter period with, compared to *without*, beep events).



←

Figure 4. (a) Mean fixation duration (±*SE*) in milliseconds as a function of cue validity and sound presence; stars (continuous line) and squares (dotted line) represent the sound-present and sound-absent conditions, respectively. (b) Mean fixation duration (±*SE*) in milliseconds as a function of cue validity (valid, invalid) and fixation type; squares correspond to fixations on trials without sounds (sound-absent condition); stars and diamonds denote fixations with and, respectively, without beep on trials with sounds (sound-present condition). (c) Mean saccade amplitude (±*SE*) in degrees of visual angle as a function of cue validity (valid, invalid) and saccade type; stars and diamonds represent saccades with and, respectively, without preceding beep on sound-present trials, and squares represent saccades on sound-absent trials.

A second interesting finding was that the fixation duration (saccadic latency) became longer when a synchronous beep was presented, and the amplitude of the immediately following saccade was also larger than for the other saccade types. This effect, however, was independent of cue validity. The long duration of fixations at beep could be indicative of participants' search strategy, namely, waiting upon the occurrence of the beep in order to detect a (target) color change. (Changes at beeps would provide an effective attentional pointer to the target.) Also, the long saccade latency may be attributable to a general auditory "freezing effect" (Vroomen & de Gelder, 2000); that is, a sudden sound may inhibit the saccade. Thus, visual information can be sampled for longer and over a larger region of space, allowing covert attention to be deployed more efficiently (Perrott et al., 1990). In more detail, the extended ("wait-at-beep") fixations would permit information sampling over a longer period of time and, probably, across an expanded spatial region—improving both the quality and the spatial range of the sampled information. If no target (saliency) signal is picked up within the currently sampled region, this situation would lead to a larger-amplitude saccade to some other, hitherto non-sampled region. If this were the case, it would predict a pip-and-pop effect to be also evident on target-absent trials. To examine this prediction, we introduced a target-absent condition and beeps synchronizing with distractors in Experiment 2.

Experiment 2

In the predecessor study (van der Burg et al., 2008), as in the present Experiment 1, a target (either a horizontal bar or a vertical bar) was presented on each trial and synchronized (in its color change) with beeps, so that a general effect of the sound (such as "freezing effect" and attendant changes in information sampling) cannot be disassociated from a "pip-and-pop" effect (based on audiovisual integration). In Experiment 1, the auditory facilitation effect was found to derive mainly from the non-target side, suggestive of a general auditory enhancement. Experiment 2 was designed to distinguish a general "pip" from a "pip-and-pop" effect by introducing target-absent trials, in addition to target-present trials. If the non-spatial beeps cause a general enhancement of visual search, for instance, as a result of temporally and spatially extended information sampling during wait-at-beep fixations, the facilitation effect should be observed even on target-absent trials with (in this condition entirely) irrelevant sounds.

Method

The method was same as in Experiment 1, with the exceptions set out below.

Participants

Fifteen right-handed observers (nine females, mean age 25.3 years) with normal or corrected-to-normal visual acuity and normal hearing participated in the experiment. They gave written informed consent and were paid for their participation. They also practiced the task in one block of 40 trials prior to the formal experiment.

Design and procedure

Instead of arrow cues, a white fixation dot in the display center (0.2° × 0.2°, 75.8 cd/m^2) was shown before the start of a given trial. The dynamic search display would be presented only when participants had fixated on the dot for at least 1000 m. In the search display, all items were randomly distributed across an invisible 10 × 10 matrix (10.7° × 10.7°, 0.55° jitter). To avoid immediate detection, targets (if present) never appeared within the four central cells of the matrix (see Figure 5). Overall, the search display contained a target (either a horizontal or a vertical bar) in half of the trials; in the other half, displays contained only (oblique-bar) distractors. Similar to Experiment 1, there were two sound conditions: sound-present and sound-absent. Importantly, in the sound-present condition, the onset of the beeps was synchronized with the target color changes on target-present trials, but with random distractors color changes (1 to 3 items) on target-absent trials. Participants had to make a two-alternative forced-choice (2AFC) response as rapidly as possible to indicate whether or not a target was present. Sound-present and -absent conditions were administered block-wise, with 4 blocks for each condition presented in random order; in contrast, target-present and -absent trials were randomized within each block of 30 trials.

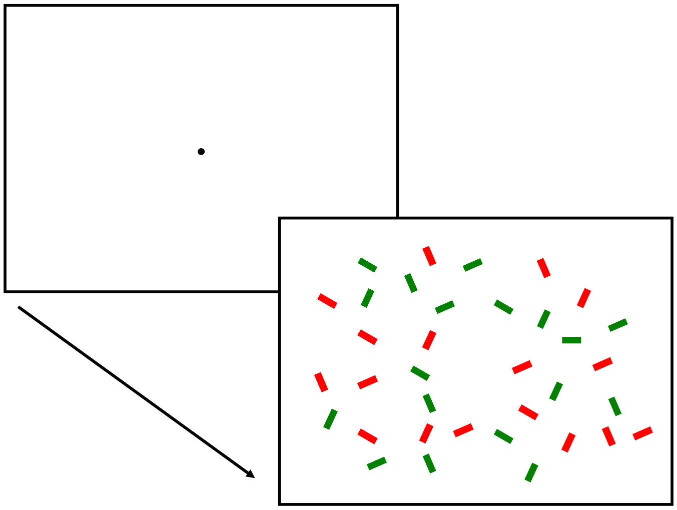


Figure 5. Example search display used in Experiment 2. Displays contained 36 bars of different orientations, and observers had to detect whether or not a target, either a horizontal or a vertical bar, was present. There was a repeating alteration of the display items' colors, occurring at random time intervals. The onset of the color changes were accompanied by mono-tone beeps, which were either synchronized with the changes of the target or of distractors depending on conditions of target presence (see Method section for details).

Results and discussion

Mean accuracy was lower for target-present trials (90.3%) than for target-absent trials (99.8%), $F(1, 14) = 51.85$, $p < 0.01$, $\eta_p^2 = 0.79$. For target-present trials, mean RTs were significantly longer for error (i.e., target miss) responses (12.21 s) than for correct (hit) responses (5.39 s), $F(1, 14) = 62.45$, $p < 0.01$, $\eta_p^2 = 0.82$; by contrast, for target-absent trials, mean RTs did not differ significantly between error (i.e., false-alarm) responses and correct (rejection) responses, $F(1, 14) = 0.15$, $p = 0.71$. This pattern (in particular, the raised error rate for target-present trials) is likely attributable to the difficulty of the search task: participants stopped searching after a certain amount of time had elapsed without a target having been detected. The bias of responding "target absent" in this case yielded an increased error rate on target-present trials. Note that response accuracy was unaffected by sound condition, $F(1, 14) = 2.00$, $p = 0.17$, $\eta_p^2 = 0.12$. Thus, only trials with correct responses were subjected to the subsequent analyses.

Reaction time effects

Figure 6 presents the mean correct RTs as a function of target presence for the conditions with and without sound. A repeated-measures ANOVA with the factors target presence and sound presence revealed target-

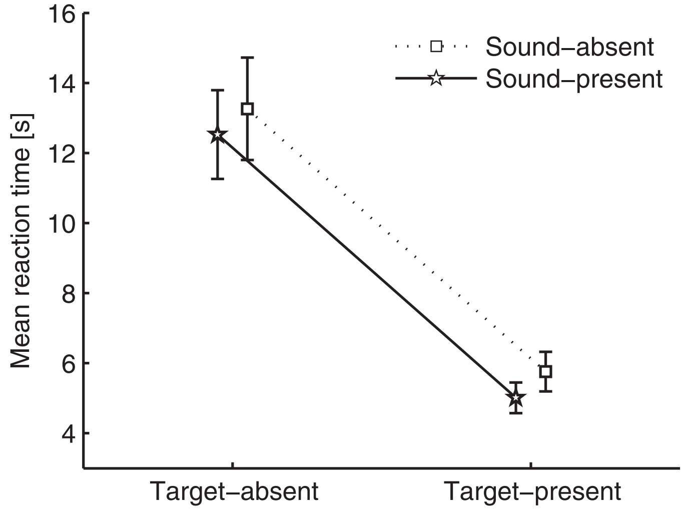


Figure 6. Mean reaction times (±*SE*) in seconds as a function of target presence (present, absent), for sound-present (stars) and sound-absent conditions (squares), respectively.

present responses to be faster than target-absent responses (5.4 s vs. 12.9 s), $F(1, 14) = 62.0$, $p < 0.01$, $\eta_p^2 = 0.82$. The main effect of sound presence was near-significant, $F(1, 14) = 4.48$, $p = 0.05$, $\eta_p^2 = 0.23$: synchronous beeps facilitated search performance by 742 ms, consistent with the results of Experiment 1. The interaction between sound presence and target presence was not significant, $F(1, 14) = 0.001$, $p = 0.97$, indicating that synchronous beeps facilitated responding to essentially the same extent on target-absent as on target-present trials.

Oculomotor effects

The mean fixation durations are depicted in Figure 7a. A repeated-measures ANOVA of the fixation durations failed to reveal a difference between target-present and -absent trials (main effect of target presence: $F(1, 14) = 2.62$, $p = 0.12$, $\eta_p^2 = 0.16$). However, fixation durations were significantly longer on trials with, than on trials without, sound (main effect of sound presence: $F(1, 14) = 6.03$, $p < 0.05$, $\eta_p^2 = 0.3$); thus, they were consistent with the results (in the target-present condition) of Experiment 1. Importantly, this effect of sound presence was also manifest on target-absent trials, indicating that the beeps had a general effect (not confined to target presence) on visual search performance. Similar to Experiment 1, we categorized fixations into three types (fixations in the sound-absent condition, fixations without beeps in the sound-present condition, and fixations with beeps). Further analysis of the fixation durations with the factors fixation type and target presence revealed the mean duration of a fixation to be overall longer when it was accompanied by a beep, compared to the other two fixation types without beeps, $F(2, 28) = 14.90$, $p < 0.01$, $\eta_p^2 = 0.52$, regardless of target presence or absence, $F(1, 14) = 3.10$, $p = 0.10$, $\eta_p^2 = 0.18$ (Figure 7b). Interestingly, the interaction between target presence and fixation type was significant, $F(2, 28) = 3.54$, $p < 0.05$, $\eta_p^2 = 0.2$, mainly due to the slightly longer duration of fixations with beeps in the target-present, compared to the target-absent, condition (461.1 ms vs. 410.9 ms; Figure 7b).

These findings of an increased fixation duration at beeps (with or without a target) and an even slightly longer duration in the sound-and-target-present condition would appear to be at variance with a “pip-and-pop” account assuming a (spatially non-specific) boosting of visual salience by the beeps (van der Burg et al., 2008). Such an account would appear to predict the opposite pattern: if target salience is enhanced by the sound, one would expect the fixation duration (i.e., the latency of next, target-directed saccade) to be shortened. Note that the logic of this argument is similar to that adduced to account for generally slower target-absent compared to target-present decision in visual search (see, e.g., Chun & Wolfe, 1996). By contrast, the finding of increased fixation durations at beeps is more in line with a “wait-at-beep” strategy induced by the sounds. And the slightly longer durations in the sound-and-target-present (vs. the sound-and-target-absent) condition may be explained by assuming that an emerging (target) saliency signal during such a fixation reinforces this strategy to gain confidence (based on further accumulating evidence) in a “target-present” decision.

Analysis of the number of fixations revealed that overall fewer fixations were made in the sound-present than in the sound-absent condition, $F(1, 14) = 10.67$, $p < 0.01$, $\eta_p^2 = 0.43$. And, as typically found in the visual search, establishing target presence required fewer fixations than did establishing target absence, $F(1, 14) = 73.71$, $p < 0.01$, $\eta_p^2 = 0.84$. The interaction between target presence and sound presence was non-significant, $F(1, 14) = 0.67$, $p = 0.43$, $\eta_p^2 = 0.04$ (Figure 7c).

The mean saccade amplitudes are shown in Figure 7d. A repeated-measures ANOVA revealed a significant main effect of the target presence, $F(1, 14) = 6.82$, $p < 0.05$, $\eta_p^2 = 0.33$, while the effect of sound presence was non-significant, $F(1, 14) = 2.00$, $p = 0.18$, $\eta_p^2 = 0.12$. The target presence × sound presence interaction was marginally significant, $F(1, 14) = 4.16$, $p = 0.06$, $\eta_p^2 = 0.23$. On average, saccade amplitude was slightly larger on target-absent than on target-present trials. This result was likely due to the small saccades near the target position at the end of the search (on target-present trials). This premise was confirmed by a comparison of the proportions of saccades smaller than 1° (i.e., approximately the inter-item distance) between target-present (25.1%) and target-absent trials (22.2%): the proportion of such saccades was significantly higher

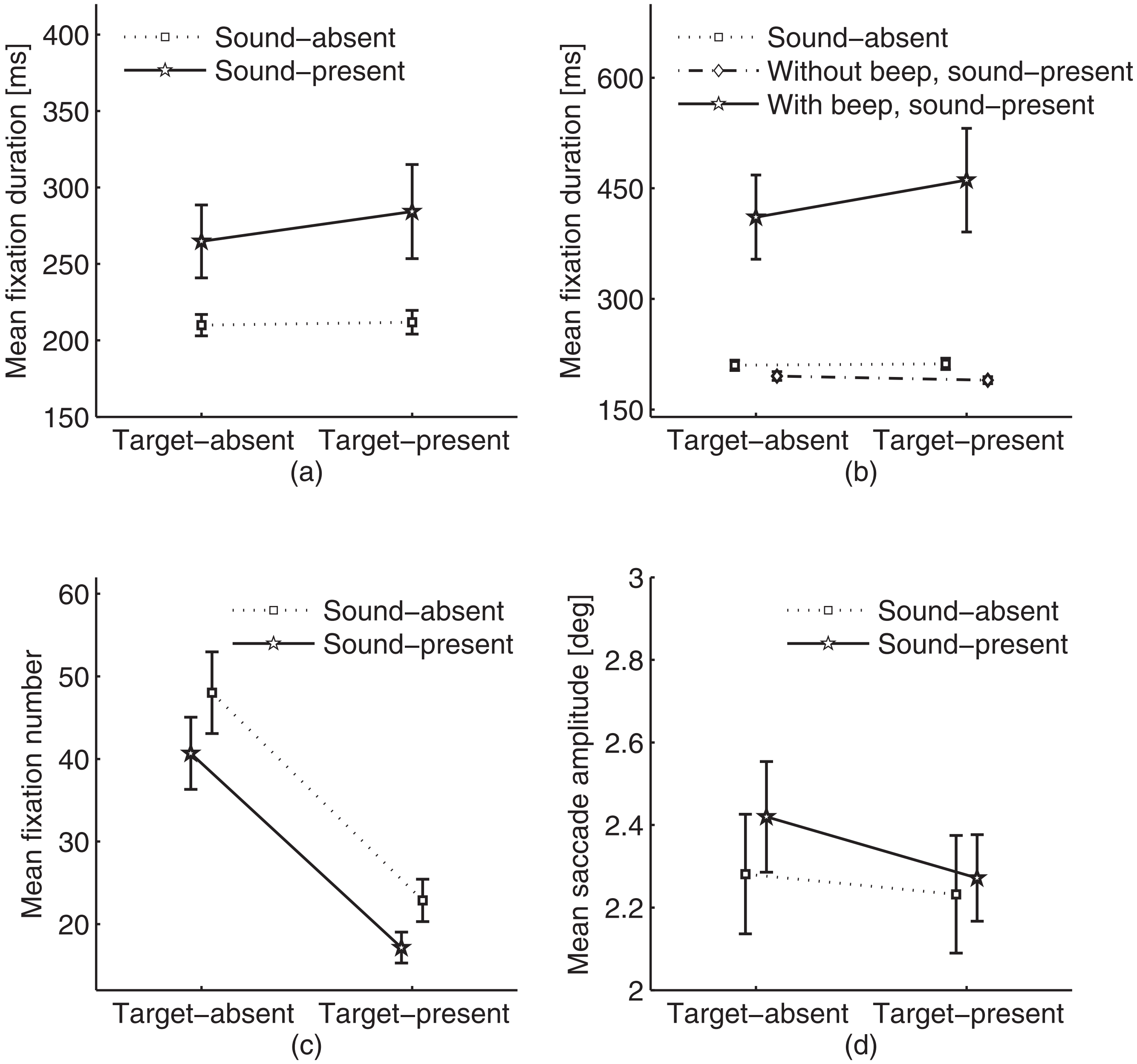


Figure 7. (a) Mean fixation duration (± *SE*) in milliseconds as a function of target presence (present, absent), for sound-present (stars) and -absent conditions (squares), respectively. (b) Mean fixation duration (± *SE*) in milliseconds as a function of target presence (present, absent), separately for fixations on sound-absent trials (squares), and for fixations with (stars) and, respectively, without beep (diamonds) on sound-present trials. (c) Mean number of fixation (± *SE*) as a function of target presence (present, absent), for sound-present (stars) and -absent conditions (squares), respectively. (d) Mean saccade amplitude (± *SE*) in degrees of visual angle as a function of target presence (present, absent), for sound-present (stars) and -absent (squares) conditions, respectively.

in the former condition, $F(1, 14) = 25.85$, $p < 0.01$, $\eta_p^2 = 0.65$. The borderline-significant interaction ($p = 0.06$) was mainly due to (the presence of) beeps increasing saccade amplitudes in the target-absent condition, $t(14) = 2.6$, $p < 0.05$. Recall that the fixation duration analysis (presented above) had revealed the mean fixation duration to be extended in the sound-present conditions. Longer fixation durations may permit the attentional spotlight to be expanded and gain greater confidence that a target is actually absent within the currently attended region. For target-absent trials, this condition would lead to, on average, larger subsequent saccades to outside the currently scanned region.

In summary, consistent with results of Experiment 1 and previous findings (van der Burg et al., 2010; van der Burg et al., 2008), non-spatial beeps synchronized with dynamic color changes of the target can facilitate visual search. The major finding of Experiment 2 was that the beeps exerted a general "pip" facilitation effect on search performance, which is seen even when a target is absent. The fixation and saccade results indicate that the sound actually "froze" the eye movement (prolonged fixation durations), permitting improved saccadic planning (fewer saccades).

However, one could still argue that this 'freezing effect' was actually due to a saliency boost induced by the sound occurring in synchrony with the visual change event (van der Burg et al., 2008), rather than by the presentation of the sound itself—as saliency may be boosted even with sounds occurring synchronized with distractors on target-present trials. Given this argument, Experiment 3 was designed to provide further evidence for a general effect of the auditory beeps on oculomotor regulation, by presenting beeps, with the same frequency and structure over time, in either a random fashion (i.e., not correlated with the onset of a visual target change) or a constant 100 ms prior to the target color change.

Experiment 3

In contrast to Experiments 1 and 2 in which the beeps were always synchronized to color changes of the visual target (or distractor) item(s), in Experiment 3 we introduced two conditions of audiovisual asynchrony (in addition to a synchronous-sound and a sound-absent condition). In one condition, the beep was presented consistently 100 ms prior to the visual target change. In the other condition, the beeps occurred randomly, that is, independently of the visual change, though their temporal frequency and structure were kept the same as in the audiovisual synchrony condition. If the sounds regulate oculomotor scanning behavior, one would expect similar search facilitation and eye movement patterns in all sound-present conditions. In addition, although we made no explicit predictions as to the development of audiovisual search strategies, we were interested in examining possible effects of learning over the course of the experiment. For exploring the learning effect, we implemented a block-wise design.

Method

The method was largely the same as in Experiment 2, except for the following modifications.

Participants

Twelve right-handed observers (nine females, mean age 25.2 years) with normal or corrected-to-normal visual acuity and normal hearing participated in the experiment. They gave written informed consent and were paid for their participation. All participants were naïve with regard to the purpose of the experiment. They practiced the task in one trial block before moving on to the formal experiment.

Design and procedure

The search display always contained a target (either a horizontal or vertical bar) and 35 distractors (oblique bars). Importantly, there were four conditions: sound absent, synchronized sound, preceding sound, and random sound. Sound-absent and synchronized-sound conditions were the same as in Experiment 2. In the preceding-sound condition, the beep would be delivered 100 ms before the target color change. In the random-sound condition, beeps had no temporal relationship with visual (target) changes. Note, however, that the mean temporal frequency of the sound train was kept the same as the target color change frequency (1.1 Hz), in all three sound-present conditions. Furthermore, we kept the range of temporal variation of the sound trains the same (0.65–4 Hz) across all sound-present conditions. Participants had to make a 2AFC response as rapidly as possible to indicate whether the target was a horizontal or a vertical bar. The four conditions were administered block-wise, with two blocks for each condition presented in random order; each block consisted of 40 trials.

Results and discussion

Mean response accuracy was high overall (97.8%). Mean RT for correct-response trials was 5.35 s, as compared to 6.51 s for incorrect trials. There was no evidence of a speed-accuracy trade-off (i.e., a repeated-measures ANOVA revealed no RT "facilitation" for error- vs. correct-response trials: $F(1, 11) = 1.33$, $p = 0.27$, $\eta_p^2 = 0.11$). Error trials were excluded from further analyses, along with "outlier" trials with RTs more than 2.5 times standard deviations from the mean RT in particular conditions. Thus, 5.1% of all trials were left out in total.

Reaction time effects

Figure 8a depicts the mean correct RTs for the four different conditions. A repeated-measures ANOVA revealed a significant main effect of the sound manipulation, $F(3, 33) = 3.24$, $p < .05$, $\eta_p^2 = 0.23$. Post-hoc Bonferroni tests revealed RTs to be significantly slower in the sound-absent condition than in the three sound-present conditions (all $p < 0.05$); there were no significant differences among the three sound-

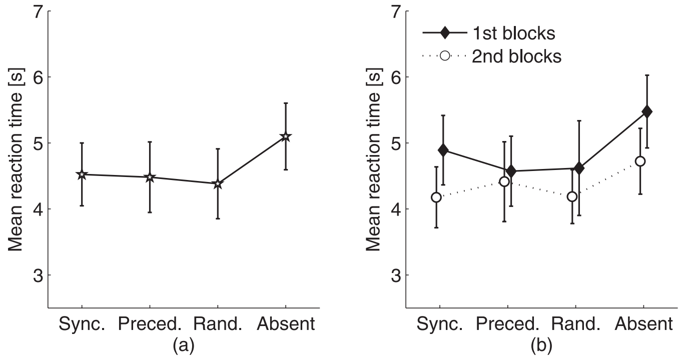


Figure 8. (a) Mean reaction time (±*SE*) in seconds as a function of sound condition. (b) Mean reaction time (±*SE*) in seconds as functions of sound condition and block (first, second). The dotted line with circles illustrates mean reaction times from the first block of each condition, and solid line denotes reaction times from the second blocks.

present conditions (all $p > 0.63$). This pattern is indicative of a general facilitative effect of the beeps. The facilitation observed for the preceding-sound condition is consistent with previous findings (van der Burg et al., 2008). Surprisingly, however, the random sound train, which provided no temporal cues with regard to visual target changes, also facilitated target detection—in fact, to a similar degree to the synchronized-sound condition (see Figure 8a). The only common feature of the three sound-present conditions was that all had the same temporal structure (i.e., the same mean frequency and variance). This commonality suggests that the mean frequency of the sound train (around 1 Hz) alone is an important factor for facilitating visual search through the present type of dynamic displays.

We further separated the "first-block" from the "second-block" data to examine how participants' performance improved during the experiment (see Figure 8b). A repeated-measures ANOVA examining RTs as a function of sound manipulation and block (first, second) revealed both main effects to be significant: sound condition, $F(3, 33) = 3.00$, $p < 0.05$, $\eta_p^2 = 0.22$, and block, $F(1, 11) = 9.43$, $p < 0.05$, $\eta_p^2 = 0.46$. Participants responded significantly faster in the second trial block, indicating a general learning effect. The interaction between two factors was not significant, $F(3, 33) = 0.58$, $p = 0.63$, $\eta_p^2 = 0.05$, suggesting the learning effect was similar in all sound conditions.

Oculomotor effects

Mean fixation durations were subjected to a repeated-measures ANOVA, with sound manipulation as the independent factor. There was no significant difference in mean fixation durations, $F(3, 33) = 1.05$, $p = 0.38$, $\eta_p^2 = 0.09$ (see Figure 9a). [Note though that, numerically, fixation durations were somewhat shorter in the random-sound condition, compared to the two informative-sound conditions (with synchronous and preceding sounds).] However, when individual fixations were categorized with regard to whether they were accompanied by a beep, we found significant differences in their durations $F(2, 22) = 5.44$, $p < 0.05$, $\eta_p^2 = 0.33$: fixations accompanied by a beep were longer, on average, than those not accompanied by a beep in the sound-present conditions (462.8 ms vs. 298.2 ms, $p < 0.05$), and longer compared to those in the sound-absent condition (462.8 ms vs. 318.2 ms, $p = 0.05$; see Figure 9c). This result again confirmed the "freezing effect" observed in previous experiments. In addition, the magnitudes of the freezing effect (i.e., mean duration differences between fixations with and without a beep) did not differ among the three sound-present conditions, $F(2, 22) = 0.15$, $p = 0.86$, $\eta_p^2 = 0.01$, indicating that the freezing effect occurs independently of the tone-target temporal manipulation.

Analysis of the number of fixations indicated significant differences among the four sound conditions, $F(3, 33) = 3{,}75$, $p < 0.05$, $\eta_p^2 = 0.24$ (see Figure 9b). Post-hoc Bonferroni tests revealed that significantly more fixations were made in the sound-absent compared to the three sound-present conditions (all $p < 0.05$), while there were no reliable differences among the latter conditions (all $p > 0.1$). [Note though that, at least numerically, the number of fixations was somewhat larger in the random-sound condition, compared to the two informative-sound conditions (with synchronous and preceding sounds).]

Similar analyses were carried out on saccade amplitudes. Mean saccade amplitude was not influenced by sound condition, $F(3, 33) = 0.16$, $p = 0.92$, $\eta_p^2 = 0.01$. In order to test whether saccade amplitudes were directly influenced by sound events, all saccades were subsequently categorized into three types (as in the analyses in Experiments 1 and 2) according to the preceding fixation: fixation with beep, fixation without beep but from the sound-present conditions, and fixation from the sound-absent condition. Analysis showed that they did not differ significantly, $F(2, 22) = 1.19$, $p = 0.32$, $\eta_p^2 = 0.10$ (see Figure 9d).

In summary, Experiment 3 provided further evidence of a general auditory facilitation effect. Providing trains of beeps (of the same temporal structure: mean frequency 1.1 Hz, range 0.65–4 Hz) facilitated search performance and "froze" the eye movement—regardless of the specific tone-target temporal relationship (synchronous, preceding, random). This result points to a common underlying mechanism, such as automatic freezing of scanning upon sound events or a "wait-at-sound" strategy and attendant changes of visual information sampling. The similarity of the oculomotor

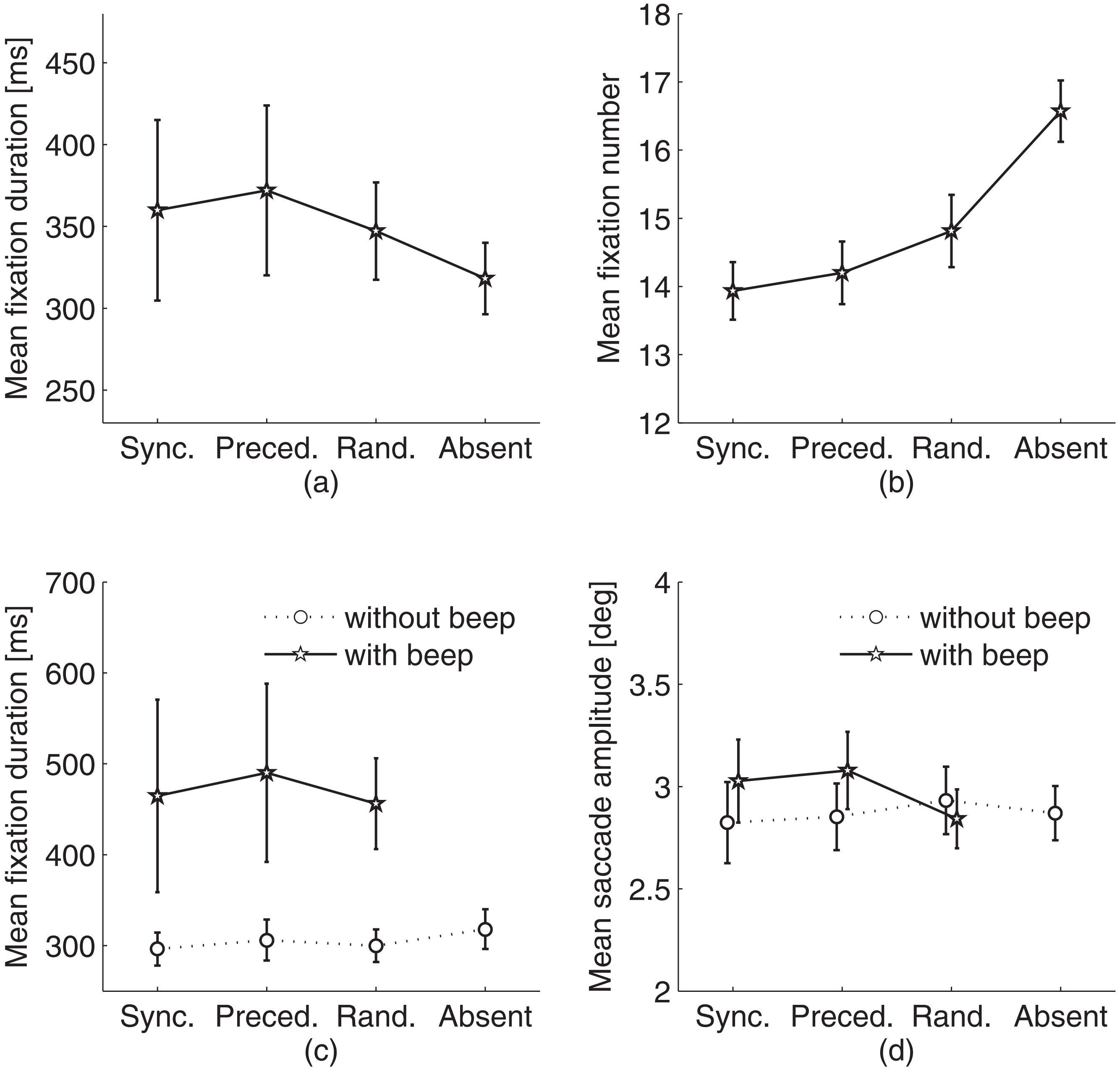


Figure 9. Analysis of oculomotor effects in Experiment 3. X-axis denotes the four sound conditions (i.e., synchronized sound, preceding sound, random sound, and sound absent). (a) Mean fixation durations (±*SE*) in seconds and (b) mean numbers of fixation (±*SE*) are shown as a function of the sound condition. (c) Mean fixation durations (±*SE*) in seconds and (d) mean saccade amplitudes (±*SE*) in degrees as a function of the sound condition, after re-categorizing fixations and saccades based on whether they were accompanied or preceded by a beep. The solid line denotes events with beeps (either the fixation accompanied by a beep or the saccade preceded by a beep), the dashed line events without beeps.

effects in the random-sound conditions to those in the two informative (synchronous and preceding), sound conditions would argue that the underlying mechanism is automatic in nature. However subtle (statistically non-significant) modulations of oculomotor scanning (in particular, the somewhat reduced number of fixations and slightly longer durations in the informative compared to the random-sound conditions) also point to the involvement of a strategic component, namely, to actively take advantage of the informativeness of the sounds about the occurrence of a target color change. [Note that since all participants performed all conditions (in random), the effects in the random-sound condition may be due to carry-over of strategy from the informative condition; however, even participants who performed the random condition before any of the informative conditions showed a similar pattern to that in the (subsequently performed) informative conditions, arguing in favor of a strong automatic component in the generation of these effects.]

General discussion

In three experiments, we replicated the pip-and-pop effect employing a similar paradigm and search displays to those used by van der Burg and colleagues (van der Burg et al., 2008): that is, non-spatial sounds synchronized with periodic color changes of the visual target did substantially facilitate visual search performance. Presenting participants with a visuospatial (central-arrow) cue prior to the onset of the dynamic search display, we found both a general cueing effect and an enhanced facilitation effect by the synchronous sound in the invalid-cue, compared to the valid-cue, condition (Experiment 1). One might argue that the interaction (i.e., the modulation of the facilitation effect by cue validity) might be due to a floor effect: that is, little scope for facilitation in the valid-cue condition. However, a floor effect is unlikely to (fully) account for this pattern given that the mean RTs were generally slow (of the order of 2–3 seconds) and the benefit was of the order of seconds (rather than just fractions of a second). Note that a recent study by Ngo and Spence (2010, Experiment 3) using exogenous spatial cueing in a pip-and-pop paradigm also revealed a "classical" cueing effect for dynamic visual search, which was an additive to the set size effect. Unfortunately, though, there was no sound-absent condition in their study; thus, the study yielded no specific evidence as to how exogenous spatial cues interact with the pip-and-pop effect. Our results thus provide the first evidence of such an interaction between (endogenous) spatial cueing and the pip-and-pop effect.

Examining the oculomotor scanning behavior, we found a similar interaction pattern in the number of fixations: the number was greatly reduced by the synchronous sound in the invalid-cue condition, compared to a smaller benefit in the valid-cue condition. Additional analysis of oculomotor scanning on the (valid) target and the (invalid) non-target sides suggested that the interaction effect between cue validity and sound presence on the number of fixations was attributable mainly to the non-target side: with sound present, even though there was no target on that side, participants would switch over to the correct side more quickly. This finding suggests that the presentation of sounds alters the oculomotor scanning behavior and, thus, information sampling, permitting improved guidance of search to as yet unscanned display regions likely to contain the target.

Furthermore, the results of Experiment 2 showed that the presentation of sounds (synchronous with distractor changes) in the target-absent condition also did facilitate search performance. Experiment 3 further suggested that physical audiovisual synchrony is not a critical factor for observing the search enhancement: sound trains with the same temporal structure as synchronous events (mean 1.1 Hz, range 0.65–4 Hz) facilitated visual search, not only when the beeps were delivered consistently 100 ms prior to the target change, but also when the beep events occurred at random (unrelated to the target changes). Note that in the random sound condition, the audiovisual asynchrony between the beep and the closest visual distractor (or, much less frequently, *target*) color change events never exceeded 75 ms (given the maximum color change interval was 150 ms), which lies within the ranges of perceptual synchrony (Elliott, Shi, & Kelly, 2006) and, respectively, audiovisual integration (Levitin, MacLean, Mathews, Chu, & Jensen, 2000; Stone et al., 2001). Asynchronous tone-distractor change within this temporal range would be perceived similar to synchronized tone-distractor change events. Thus, the findings of Experiment 3 further corroborated and generalized the findings of Experiment 2, in which synchronized tone-distractor change events (on target-absent trials) facilitated performance in a similar manner to synchronized tone-target change events.

Analyses of the eye movements revealed that the duration of the fixation during which a sound event occurred (or the latency of the subsequent saccade) was extended and the amplitude of the immediately following saccade was increased on the non-target side (Experiment 1) or in the target-absent condition (Experiment 2). Similar patterns of fixation were also observed in the sound-preceding and random-sound conditions of Experiment 3. At first glance, this eye movement pattern appears to be at variance with previous studies of (simple) audiovisual "search" (Colonius & Arndt, 2001; Corneil et al., 2002; Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994), in which the latency of the initial saccade was often shortened. However, in these studies, the audiovisual stimuli were presented immediately and only once, and the search task was very simple (e.g., the target was either on the left or on the right), so that the latency of the first saccade could provide a measure for cross-modal integration. By contrast, in the current, complex dynamic search task, the beeps were not presented right at the onset of the search display, but only later once scanning of the display had gotten under way. In this situation, programming of the next saccade must be dynamically adjusted according to what change occurs in the display. During the search process, ongoing saccade programming may be interrupted by sound events, possibly due to automatic freezing or a top-down wait-at-beep strategy (given that participants learnt that the auditory event is potentially informative for target detection). Thus, fixation durations (saccade latencies) became longer following the occurrence of a beep, with extended fixations permitting temporally and spatially expanded information sampling, improving the registration of singleton color changes and thus guiding the next saccade more precisely and efficiently to the target. Similar to the situation in the study of van der Burg et al. (2008), where participants tended to wait for the sound beep, in our experiments, observers tended to fixate longer when additional sounds were presented; this behavior was generally associated with fewer saccades (and larger saccade amplitudes) in sound-present conditions.

However, the oculomotor effects observed in the present study are difficult to explain in terms of a "saliency-boosting" account (van der Burg et al., 2008), which assumes that the sound signals are integrated early on with (e.g., multiplicatively amplify) visual saliency information. This account would predict that when the target becomes more salient (as a result of its saliency signal being enhanced by the synchronous beep), it should be found more easily; that is, the fixation duration (or the latency of the post-beep saccade) should become shorter. However, the durations of fixations at beeps were actually found to be longer compared to those of fixations without beeps, in all experiments.

Instead, the extended fixation duration at beeps may be closely related to the previously described phenomenon of auditory *freezing* (Vroomen & de Gelder, 2000, 2004). In the relevant experiments, participants were asked to search for a target in a rapidly changing display stream, with their eyes fixated on the center (i.e., without making eye movements). When an abrupt sound was synchronized with the rapidly changing display, detectability of the target increased and, perceptually, the display containing the target appeared to last longer or to be brighter. This freezing effect has been attributed to cross-modal enhancement at the level of perceptual organization (Vroomen & de Gelder, 2004). By contrast, for the situation realized in our study, we propose that cross-modal enhancement may directly or indirectly arise from the change of the oculomotor scanning behavior, involving the freezing of eye movements. Longer fixations at beeps permit both temporally extended information sampling and a larger display region to be scanned in parallel (as a result of opening up the attentional spotlight). As a result, if there is a target signal within the attended region (as would be the case in a fraction of fixations on target-present trials), its saliency signal would be more likely to reach the threshold for a detection decision and trigger a direct saccade to the target. Analogously, for fixations on target-absent trials (and fixations on target-present trials in which the target is outside the currently attended region), if a saliency signal fails to emerge within the extended processing time, it becomes more certain that the currently attended region does in fact not contain a target. This situation, in turn, would lead to more target-directed saccades, of shorter amplitudes, on target-present trials, and more efficient scanning, characterized by fewer and larger-amplitude saccades, on target-absent trials. This hypothesis is in line with the present data, which yielded evidence of larger saccade amplitudes immediately after beeps in non-target regions—that is, generally on target-absent trials, and an increased proportion of smaller saccades ($<1°$) on target-present trials. Similarly, longer fixations at beeps in preceding-sound and random-sound conditions would allow for both fast target detection (if the target is located within the currently attended region) and fast rejection of the currently attended region (if the target is not located inside this region), which in turn brings about the overall search enhancement.

Note, however, that in all experiments, the sound trains had the same temporal structure as the target color changes (mean frequency 1.1 Hz, range 0.65–4 Hz), and each experiment contained one synchronous tone-target change condition. Such median presentation rhythms made "freezing" oculomotor scanning possible in the first instance (in pilot experiments, we failed to find pip-and-pop effects with a faster rhythm of 2.2 Hertz!), and the informativeness of the beeps as to the occurrence of a target color change may have additionally encouraged participants to adopt a wait-at-beep strategy to optimize information sampling. In addition, both the visual and auditory stimuli had abrupt onsets, likely facilitating the triggering of the detection threshold by visual events within the focally attended (fixated) region. Audiovisual enhancement has been shown to be reduced or entirely abolished when the temporal square wave modulation of the audiovisual stimuli (i.e., abrupt on- and offsets) was replaced by a sine wave modulation (or gradients of on- and offset ramps) (van der Burg et al., 2010), or when the auditory stimuli were part of a melody (Vroomen & de Gelder, 2000, 2004). Quite possibly, abrupt onsets are a necessary condition for "bottom-up" enhancement of visual saliency by auditory events, along the lines of van der Burg et al. (2008, 2010), as well as the coming into play of a more "top-down" controlled modulation of oculomotor scanning behavior, as demonstrated in the present study. In fact, the top-down changes demonstrated here may actually be a requirement for effectively picking out any weak bottom-up signals, within the dynamically changing displays, even if amplified by an audiovisual interaction.

The oculomotor freezing account, as developed above, would be sufficient to explain the enhanced visual search performance observed in the present experiments. However, the present data do not rule out an alternative, audiovisual integration account (van der Burg et al., 2008). Consistent with such an account, early processing of synchronized audiovisual events (Molholm et al., 2002; Talsma, Doty, & Woldorff, 2007) has been shown to correlate with behavioral visual detectability (van der Burg et al., 2011). However, we argue that any such early enhancement effect in the pip-and-pop paradigm would be too subtle to explain the whole effects found here, particularly the findings in the target-absent condition of Experiment 2 and the random-sound (target-present) condition of Experiment 3. These findings show that oculomotor freezing not only enhances the picking out of a target (change event), but also facilitates the rejection of regions containing only distractors (change events). While not ruling out bottom-up audiovisual enhancement for synchronous tone-target change events (van der Burg et al., 2008; van der Burg et al., 2011), our findings can be taken to highlight the general sound-induced freezing effect in oculomotor scanning for discriminating (audio-) visual events.

In summary, the present study revealed non-spatial beeps not only to enhance detection of target presence in visual search, but also to facilitate establishing target absence. Based on the eye movement data, these enhancements could be attributed to changes in visual scanning behavior induced by the accompanying sounds, in particular, freezing of the eyes at beeps and improved "targeting" of the subsequent saccades. Thus, while the present findings agree with previous studies (van der Burg et al., 2010; van der Burg et al., 2008), the extent to which the effect depends on true audiovisual signal integration, rather than changed information sampling within extended fixations, remains an important question for future research.

Acknowledgments

This research was supported in part by a grant from the German Research Foundation (DFG) within the Collaborative Research Centre (SFB453). The authors express their thanks to Sebastian Urchs for his assistance in carrying out the experimental work.

Commercial relationships: none.
Corresponding author: Zhuanghua Shi.
Email: shi@psy.lmu.de.
Address: Allgemeine und Experimentelle Psychologie, Ludwig-Maximilians-Universität München, Munich, Germany.

References

Baldauf, D., & Deubel, H. (2008). Properties of attentional selection during the preparation of sequential saccades. *Experimental Brain Research,* 184:411–425, doi:10.1007/s00221-007-1114-x. [PubMed].

Bolia, R. S., D'Angelo, W. R., & McKinley, R. L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors,* 41(4):664–669. [PubMed].

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

Chen, L., Shi, Z., & Müller, H. J. (2010). Influences of intra- and crossmodal grouping on visual and tactile Ternus apparent motion. *Brain Research,* 1354:152–162, doi:10.1016/j.brainres.2010.07.064. [PubMed].

Chen, L., Shi, Z., & Müller, H. J. (2011). Interaction of perceptual grouping and crossmodal temporal capture in tactile apparent-motion. *PloS One,* 6(2):e17130, doi:10.1371/journal.pone.0017130. [PubMed].

Chun, M. M., & Wolfe, J. M. (1996). Just say no: how are visual searches terminated when there is no target present? *Cognitive Psychology,* 30:39–78, doi:10.1006/cogp.1996.0002. [PubMed].

Colonius, H., & Arndt, P. (2001). A two-stage model for visual-auditory interaction in saccadic latencies. *Perception & Psychophysics,* 63(1):126–147. [PubMed].

Corneil, B. D., Van Wanrooij, M., Munoz, D. P., & Van Opstal, A. J. (2002). Auditory-visual interactions subserving goal-directed saccades in a complex scene. *Journal of Neurophysiology,* 88(1):438–454. [PubMed].

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers,* 34(4):613–617. [PubMed].

Doyle, M. C., & Snowden, R. J. (1998). Facilitation of visual conjunctive search by auditory spatial information. *Perception, Supplementary,* 27:134.

Elliott, M. A., Shi, Z., & Kelly, S. D. (2006). A moment to reflect upon perceptual synchrony. *Journal of Cognitive Neuroscience,* 18(10):1663–1665, doi:10.1162/jocn.2006.18.10.1663. [PubMed].

Geyer, T., Von Mühlenen, A., & Müller, H. J. (2007). What do eye movements reveal about the role of memory in visual search? *Quarterly Journal Experimental Psychology,* 60(7):924–935. [PubMed].

Hughes, H. C., Nelson, M. D., & Aronchick, D. M. (1998). Spatial characteristics of visual-auditory summation in human saccades. *Vision Research,* 38(24):3955–3963, doi:S0042-6989(98)00036–4 [pii]. [PubMed].

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance,* 20(1):131–153. [PubMed].

Kleiner, M., Brainard, D., & Pelli, D. (2007). *What's new in Psychtoolbox-3?* Paper presented at the Perception ECVP Abstract Supplement.

Krummenacher, J., Müller, H. J., & Heller, D. (2002). Visual search for dimensionally redundant pop-out targets: parallel-coactive processing of dimensions is location specific. *Journal of Experimental Psychology: Human Perception and Performance,* 28(6):1303–1322. [PubMed].

Levitin, D. J., MacLean, K., Mathews, M., Chu, L., & Jensen, E. (2000). The perception of cross-modal simultaneity (or "The Greenwich Observatory Problem revisited)." *Computing Anticipatory Systems,* 517:323–329.

McConkie, G. W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society,* 8:365–368.

McDonald, J. J., Teder-Salejarvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature,* 407(6806):906–908. [PubMed].

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. *Brain research. Cognitive Brain Research,* 14(1):115–128. [PubMed].

Müller, H. J., & von Mühlenen, A. (2000). Probing distractor inhibition in visual search: inhibition of return. *Journal of Experimental Psychology: Human Perception and Performance,* 26(5):1591–1605. [PubMed].

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

Pashler, H. (1987). Detecting conjunctions of color and form: reassessing the serial search hypothesis. *Perception & Psychophysics,* 41(3):191–201. [PubMed].

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision,* 10:437–442, doi:10.1163/156856897X00366. [PubMed].

Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search behavior. *Perception & Psychophysics,* 48:214–226. [PubMed].

Perrott, D. R., Sadralodabai, T., Saberi, K., & Strybel, T. Z. (1991). Aurally aided visual search in the central visual field: effects of visual load and visual enhancement of the target. *Human Factors,* 33(4):389–400. [PubMed].

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science,* 12:287–292. [PubMed].

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience,* 13:25–42, doi:10.1146/annurev.ne.13.030190.000325. [PubMed].

Shi, Z., Chen, L., & Müller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: the influence of time interval. *Experimental Brain Research,* 203(4):723–735, doi:10.1007/s00221-010-2286-3. [PubMed].

Spence, C., Sanabria, D., & Soto-Faraco, S. (2007). Intersensory Gestalten and crossmodal scene perception. In K. Noguchi (Ed.), *Psychology of beauty and Kansei: New horizons of Gestalt perception* (pp. 519–579). Tokyo: Fuzanbo International.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., et al. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society of London Series B-Biological Sciences,* 268(1462):31–38. [PubMed].

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. [PubMed].

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. [PubMed].

van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Per-*

ception and Performance, 34(5):1053–1065. [PubMed].

van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2009). Poke and pop: tactile-visual synchrony increases visual saliency. *Neuroscience Letters,* 450(1):60–64. [PubMed].

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,* 55:1208–1218, doi:10.1016/j.neuroimage.2010.12.068. [PubMed].

von Mühlenen, A., Müller, H. J., & Müller, D. (2003). Sit-and-wait strategies in dynamic visual search. *Psychological Science,* 14(4):309–314. [PubMed].

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance,* 26(5):1583–1590. [PubMed].

Vroomen, J., & de Gelder, B. (2004). Perceptual effects of crossmodal stimulation: Ventriloquism and the freezing phenomenon. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 141–150). London: The MIT Press.