Journal of Experimental Psychology:
Human Perception and Performance
2008, Vol. 34, No. 5, 1053–1065

0096-1523/08/$12.00 DOI: 10.1037/0096-1523.34.5.1053

Pip and Pop: Nonspatial Auditory Signals Improve Spatial Visual Search

Erik Van der Burg and Christian N. L. Olivers
Vrije Universiteit

Adelbert W. Bronkhorst
Vrije Universiteit and TNO Human Factors

Jan Theeuwes
Vrije Universiteit

Searching for an object within a cluttered, continuously changing environment can be a very time-consuming process. The authors show that a simple auditory pip drastically decreases search times for a synchronized visual object that is normally very difficult to find. This effect occurs even though the pip contains no information on the location or identity of the visual object. The experiments also show that the effect is not due to general alerting (because it does not occur with visual cues), nor is it due to top-down cuing of the visual change (because it still occurs when the pip is synchronized with distractors on the majority of trials). Instead, we propose that the temporal information of the auditory signal is integrated with the visual signal, generating a relatively salient emergent feature that automatically draws attention. Phenomenally, the synchronous pip makes the visual object pop out from its complex environment, providing a direct demonstration of spatially nonspecific sounds affecting competition in spatial visual processing.

Keywords: attention, visual search, multisensory integration, audition, vision

Visual attention is readily drawn to visual objects that stand out from the background, such as a unique red object in a field of green objects (Theeuwes, 1992; Treisman & Gelade, 1980). It is thought that strong local differences in the visual signal receive high activation in a saliency map or location map representing the locations of interest (i.e., locations that deserve further inspection). When no such clear bottom-up signals are present, top-down control may play a larger role, such that knowledge on the visual properties relevant to the task determine which object is selected. For example, in more cluttered, heterogeneous displays, search can be limited to red objects only when observers know that the target object is red (Kaptein, Theeuwes, & Van der Heijden, 1995). Within many attention models, the top-down activation further biases the competition between objects within the saliency map by interacting with the bottom-up signals (Bundesen, Habekost, & Kyllingsbæk, 2005; Desimone & Duncan, 1995; Treisman & Sato, 1990; Wolfe, 1994).

In the present study, we show that a signal that is neither low-level visual nor provides any top-down knowledge on the location or identity of the visual target object still affects the selection of that object. We demonstrate that a nonspatial auditory event (a *pip*) can guide attention toward the location of a synchronized visual event that, without such an auditory signal, is very hard to find. In other words, the auditory event makes the target pop out. Previous studies have shown that a sound can guide attention toward a visual target, but in these studies, benefits were found only when the auditory and visual signals came from one and the same location (Bolia, D'Angelo, & McKinley, 1999; Doyle & Snowden, 1998; McDonald, Teder-Sälejärvi, & Hillyard, 2000; Perrott, Saberi, Brown, & Strybel, 1990; Perrott, Sadralodabai, Saberi, & Strybel, 1991; Spence & Driver, 1997). Other studies have demonstrated that synchrony between auditory and visual events can improve visual perception (Dalton & Spence, 2007; Vroomen & De Gelder, 2000). However, in these studies, all objects appeared serially at the same spatial location, and the studies did not address the question of how sound affects the competition between multiple objects concurrently present in a spatial layout.

Experiment 1: Nonspatial Auditory Signals Aid Spatial Visual Search

Figure 1a provides an example of the visual search displays used in our study. A demonstration can be found on http://www.psy.vu.nl/pippop. Participants searched for a horizontal or a vertical line segment, among up to 48 oblique line segments of various orientations. At random intervals, a random number of items changed color between red and green. On average once every 900 ms (1.11 Hz), the target too changed color, and it always did so alone—that is, on such moments it was the only changing item.

Erik Van der Burg, Christian N. L. Olivers, Jan Theeuwes, Department of Cognitive Psychology, Vrije Universiteit, Amsterdam, The Netherlands; Adelbert W. Bronkhorst, Department of Cognitive Psychology, Vrije Universiteit, Amsterdam, The Netherlands, and TNO Human Factors, Soesterberg, The Netherlands.

This research was supported by the Dutch Technology Foundation (STW Grant 07079), a division of the Netherlands Organization for Scientific Research and the Technology Program of the Ministry of Economic Affairs (to Jan Theeuwes and Adelbert W. Bronkhorst), and by a Netherlands Organization for Scientific Research Veni grant.

Correspondence concerning this article should be addressed to Erik Van der Burg, Van der Boechorststraat, Department of Cognitive Psychology, Vrije Universiteit, Amsterdam, The Netherlands. E-mail: e.van.der.burg@psy.vu.nl

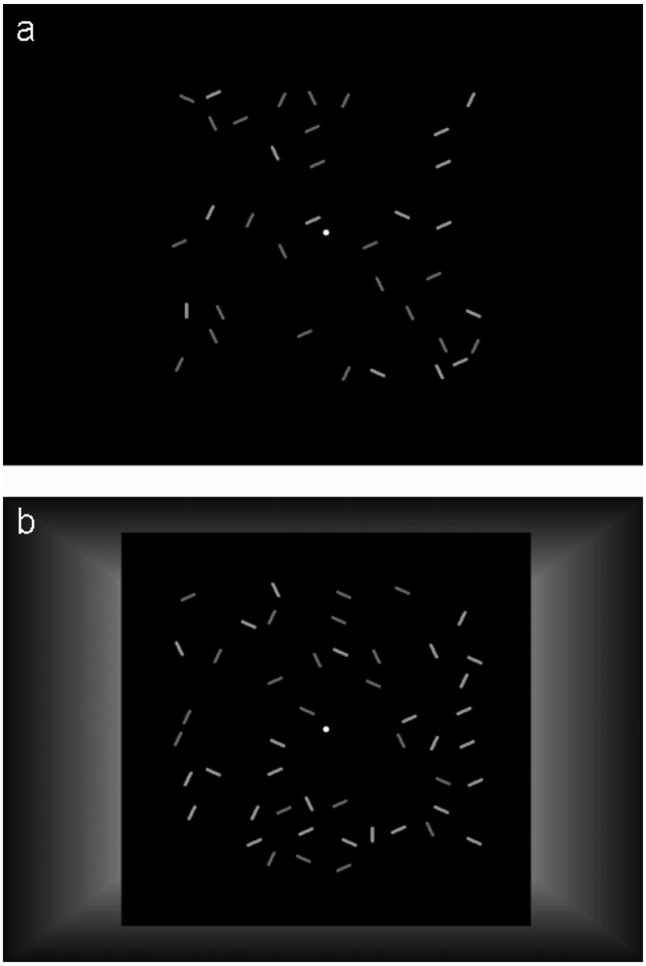

Figure 1. Panel A: Example of the visual search displays used in the present studies. Set size varied among 24, 36, and 48. Participants were instructed to make a speeded response to the orientation of a vertical or horizontal line segment. During the search, the distractors as well as the target continuously changed color between red and green, with a change occurring once every 50, 100, or 150 ms and with each element on average changing once every 900 ms. Panel B: Illustration of the peripheral halo used in Experiment 2b.

However, the target was not unique in this: At other moments, there could be a single distractor that changed. In the tone-absent condition, participants were instructed to search for either the vertical or the horizontal target and to respond as quickly and accurately as possible to its orientation. In the tone-present condition, the task was the same, but the visual target change was accompanied by a short auditory pip. Of importance, this tone provided no information about the location, the color, or the orientation of the visual target, only about the moment of change of the visual target.

Method

Participants. Six participants took part in Experiment 1 (4 female; mean age = 25.5 years, range = 18–35 years). Participants were paid €7 an hour.

Stimuli and apparatus. Experiments were run in a dimly lit, air-conditioned cabin. Participants were seated approximately 80 cm from the monitor and wore Sennheiser HD 202 headphones. The auditory stimulus was a 500-Hz tone (44.1 kHz sample rate, 16 bit, mono) with a duration of 60 ms (including a 5-ms fade-in and fade-out to avoid clicks) presented on the headphones. The visual search displays consisted of 24, 36, or 48 red (13.9 cd/m^2) or green (46.4 cd/m^2) line segments (length 0.57° visual angle) on a black ($<$0.05 cd/m^2) background. Color was randomly determined for each item. All lines were randomly placed in an invisible 10 × 10 grid (9.58° × 9.58°, 0°–0.34° jitter) centered on a white (76.7 cd/m^2) fixation dot, with the constraint that the target was never presented at the four central positions, to avoid immediate detection. The orientation of each line deviated randomly by either plus or minus 22.5° from horizontal or vertical, except for the target, which was horizontal or vertical. The displays changed continuously in randomly generated cycles of nine intervals each. The length of each interval varied randomly among 50, 100, or 150 ms, with the constraint that all intervals occurred equally often within each cycle and that the target change was always preceded by a 150-ms interval and followed by a 100-ms interval. At the start of each interval, a randomly determined number of search items changed color (from red to green or vice versa), within the following constraints: When set size was 24, the number of items that changed was 1, 2, or 3. When set size was 36, 1, 3, or 5 items changed, and when it was 48, 1, 4, or 7 items changed. Furthermore, the target always changed alone and could change only once per cycle, so that the average frequency was 1.11 Hz. The target could not change during the first 500 ms of the very first cycle of each trial. For each trial, 10 different cycles were generated, which were then repeated after the 10th cycle if the participant had not yet responded.

Design and procedure. The set size was 24, 36, or 48. Set sizes were relatively large to avoid immediate target detection before the first auditory signal was presented. The other manipulation involved the presentation of a tone coinciding with the target (tone present and absent). Dependent variables were the reaction time (RT) and accuracy. Note that the RT reflects the time between the search display onset and the response to the target, because the target is present when the search display appeared. Each trial began with a fixation dot presented for 1,000 ms at the center of the screen. The search display was presented until participants responded. Participants were asked to remain fixated on the fixation dot. Participants were instructed to press the *z* or *m* key on the standard keyboard as quickly and accurately as possible when the target orientation was horizontal or vertical, respectively. Target orientation was balanced and randomly mixed within blocks of 48 trials each. Participants received four tone-absent blocks and four tone-present blocks presented in counterbalanced, alternating order and preceded by two practice blocks. Participants received feedback about their overall mean accuracy and overall mean RT after each block.

Results and Discussion

The results of Experiment 1 are presented in Figure 2. RT data from practice blocks and erroneous trials were excluded. All data were subjected to a repeated-measures univariate analysis of variance (ANOVA) with set size (24, 36, 48) and tone presence

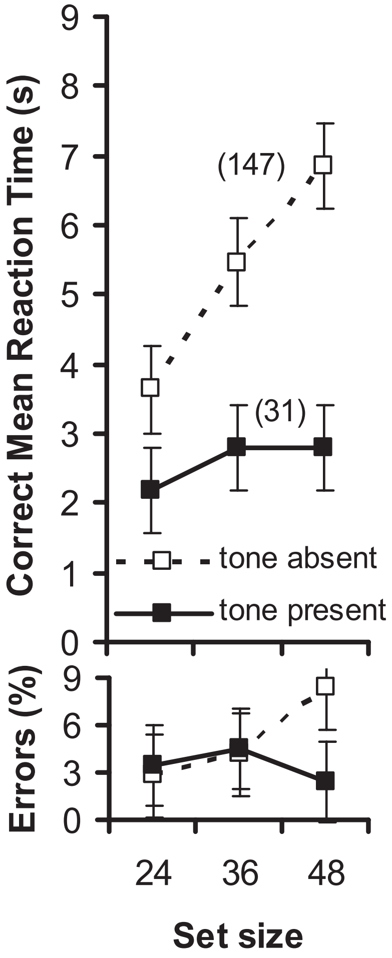

Figure 2. Results of Experiment 1. Mean correct reaction time and mean error percentages as a function of set size and auditory signal presence. Search slopes are printed next to each line (in ms/item). Note that the reaction time reflects the time to respond to the visual target from the search display onset. The first target color change (and tone onset) was between 500 and 900 ms later. The error bars represent the .95 confidence intervals for within-subject designs, following Loftus and Masson (1994). Because we were mainly interested in search slope differences, the confidence intervals are those for the set size interaction effects.

(present vs. absent) as within-subject variables. The reported values for *p* are those after a Huynh–Feldt correction for sphericity violations, with alpha set at .05. The overall mean error rate was 3.7%. There were no significant error effects, and the error pattern followed that of the RTs.

On average, RTs were faster when the tone was present than when the tone was absent, $F(1, 5) = 10.7$, $p < .05$, $\eta_p = .68$. Furthermore, search was more efficient in the tone-present condition than in the tone-absent condition, Tone Presence × Set Size interaction, $F(2, 10) = 12.7$, $p < .005$, $\eta_p = .72$. In the tone-absent condition, the average search slope measured 147 ms/item, and RTs increased significantly with increasing set size, $F(2, 10) = 7.9$, $p = .01$, $\eta_p = .61$. In the tone-present condition, the average search slope measured 31 ms/item, but the set size effect on RTs was not significant, $F(2, 10) = 1.9$, $p = .224$, $\eta_p = .28$.

Thus, despite the target uniquely changing color every now and then, finding it required strong attentional effort when the auditory pip was absent. Apparently, even though abrupt visual changes can usually be quite salient when presented alone, the many temporally neighboring changes in the display effectively camouflaged the target change (cf. Von Mühlenen, Rempel, & Enns, 2005). In other words, the visual system's temporal resolution is apparently insufficient to make it stand out. In the tone-present condition, the concurrent pip caused a dramatic improvement in visual search performance. The auditory system has a better temporal resolution than the visual system (Shipley, 1964; Welch & Warren, 1980), and we suggest that the auditory signal boosts the saliency of the visual change, creating a salient emergent feature, which results in the impression of pop out. We dub this phenomenon the *pip and pop effect*.

However, the substantial search slope and the overall still somewhat long search times (>2 s) may raise doubts about whether the target really pops out in the tone-present condition. We discuss this issue more extensively in the General Discussion section, but here we would like to note two things. First, observers probably waited for the first pip to occur before they started to search (at least they told us so). This occurred on average 750 ms after display onset. The effective RTs may thus be regarded as 750 ms shorter than is plotted in Figure 2. Figure 3 shows the RT distributions for the tone-present and tone-absent conditions, pooled across all set sizes and locked to the first target change (which was also the time of the first tone in the tone-present condition; bin size was 200 ms). Compared to the tone-absent condition, the tone-present condition shows a marked peak around 900 ms, which was on average the time the second tone could occur. On most trials in this condition, this second tone probably occurred too late to affect the response, but it is possible that occasionally, because of eye blinks or other factors, observers waited for the second tone. Thus, on the vast majority of trials, the target popped out after one or two pips. In any case, the tone-present distribution was markedly different from the tone-absent distribution, which spanned an entire range of about 1 to 10 s and more.

Second, with regard to the search slopes, we note that of the 6 observers, 1 demonstrated exceptionally high search slopes: 147 and 375 ms/item in the tone-present and tone-absent conditions, respectively, whereas the group average of the remaining participants was 8 ms/item and 102 ms/item, respectively. Also, in the subsequent experiments, we found a minority of individuals to be overall less efficient in their search.

What we propose here is that the auditory signal is integrated (i.e., directly interacts) with the synchronous visual event, resulting in a pop out of the latter. However, an alternative explanation is that the sound acted as a simple cue or warning signal as to when to expect the target change. Note that the target always changed alone but that this change occurred within a series of other changes. The tone may have simply told participants when to look out for the imperative change. In addition, the tone may have increased general alertness, or arousal, leading to improved performance. These alternative explanations are addressed next.

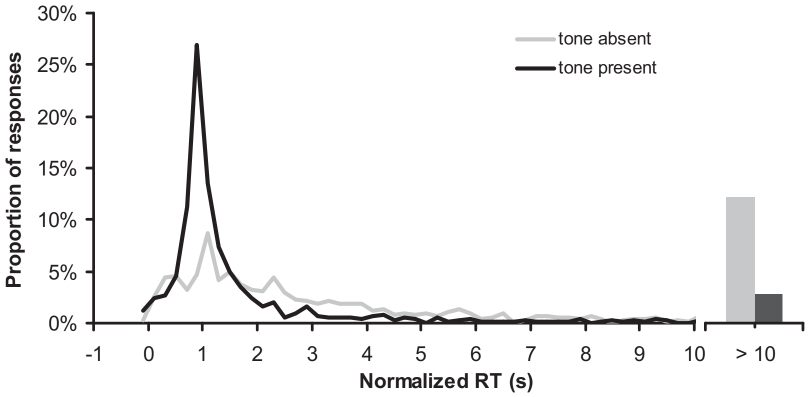

Figure 3. Reaction time (RT) distributions of Experiment 1. Here the proportion of responses is plotted as a function of the normalized RT (bin width is 200 ms). The normalized RT is the time to respond to the visual target from the first target color change.

Experiment 2: Visual Warning Signals Do Not Affect Search

In Experiment 2, we provided a first test of the cuing, arousal, or warning signal hypothesis. We replaced the tone with visual warning signals indicating when the target changed. In Experiment 2a, the warning signal consisted of the fixation dot briefly (but clearly) disappearing, and participants were told that the disappearance always coincided with the target change. Experiment 2b controlled for the possibility that observers would overly focus on the fixation dot and thus narrow their window of attention (Theeuwes, 1991; Yantis & Jonides, 1990). To distribute attention across the screen, the signal was a peripheral halo that gave the impression of a light being briefly switched on behind the visual search display (see Figure 1b for an illustration). If a simple warning signal or cue to start attending is sufficient to increase alertness and make the target change more salient, then we should also have found improvements in these visual cue conditions. If not, then this provides evidence that the pip and pop effect is of a unique multisensory nature. However, the possibility remains that these visual cues were rather ineffective as warning signals. Therefore, in Experiment 2c, we tested how effective these cues were in a typical foreperiod task with the same dynamic stimulus displays. Instead of performing a visual search task, observers now responded to the offset of the displays, as anticipated by either an auditory or visual warning signal.

Method

Six new participants completed Experiment 2a (all female; mean age = 28.0 years, range = 19–39); 6 new participants completed Experiment 2b (4 female; mean age = 18.7 years, range = 18–21), and 6 new participants completed Experiment 2c (2 female; mean age = 22.3 years, range 19–31).

Experiments 2a and 2b were identical to Experiment 1 except that the tone was replaced with a temporary offset (duration 60 ms) of the fixation dot in Experiment 2a or the presentation of a peripheral halo (duration 60 ms) in Experiment 2b.

In Experiment 2c, we used the same dynamic search displays as in the previous experiments, with set size fixed at 48 distractors. However, instead of the visual search task, participants were asked to respond as fast as possible to the offset of the search display by pressing the spacebar or to withhold their response when the display did not disappear (catch trials). The search display disappeared after a random interval from the search display onset, as sampled from an exponential distribution with a mean of 1,600 ms, and an initial constant period of 1,600 (to prevent participants from using the onset of the search display as a temporal reference). Participants could receive a cue about when the search display would disappear. The cue–target interval (CTI; time between the cue and the offset of the search display) was 0, −100, −200, −300, or −600 ms (we use negative values to indicate that the cue occurred before the target event, in accordance with Experiment 3). The cue could also be absent. The cue type was the presentation of the tone from Experiment 1, the disappearing fixation dot from Experiment 2a, or the peripheral halo from Experiment 2b. Of the cue-present trials, 17% were catch trials. All trial types were randomly mixed within blocks, except for cue type, which was blocked and presented in completely counterbalanced order. Participants practiced three blocks (dot, halo, and tone) of 10 trials each. After practice, participants performed nine blocks of 58 trials each. Participants received feedback about their overall mean accuracy and RT after each block.

Results and Discussion

The results of Experiments 2a and 2b are presented in Figure 4. In Experiments 2a and 2b, the data were subjected to a repeated-measures univariate ANOVA with set size (24, 36, and 48) and visual cue presence (present vs. absent) as within-subject variables. Overall mean error rate was 3.2% in Experiment 2a and 4.7% in Experiment 2b. There were no significant effects and no speed–accuracy trade-offs.

Unlike the auditory cue in Experiment 1, neither of the visual cues (the central fixation offset or the peripheral halo onset) resulted in any improvement (or costs) in visual search performance in terms of RTs or search slopes (all F values < 1). There were main effects only of set size, as search slopes differed significantly from zero in both experiments: $F(2, 10) = 94.6$, $p <$

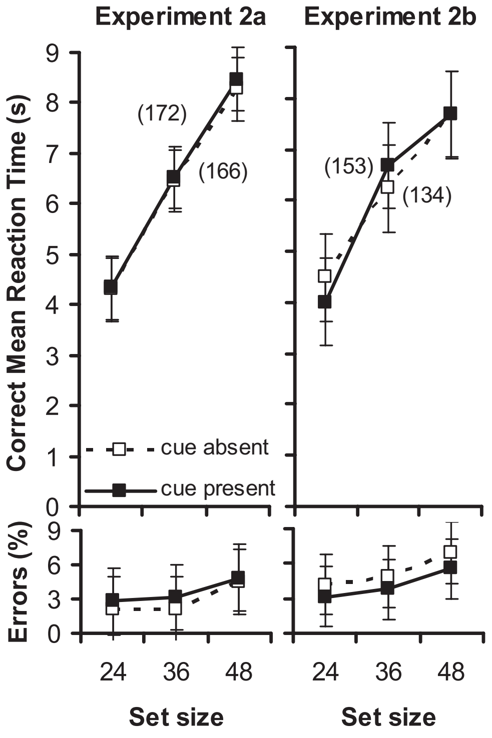

Figure 4. Results of Experiments 2a and 2b. Mean correct reaction time and mean error percentages as a function of set size and visual cue presence. The error bars represent the .95 confidence intervals for within-subject designs, following Loftus and Masson (1994). The confidence intervals are those for the set size interaction effects.

.001, $\eta_p = .95$, for Experiment 2a; $F(2, 10) = 38.4$, $p < .001$, $\eta_p = .89$, for Experiment 2b.

Clearly, the visual signals did not improve visual search, which goes against a warning or cuing explanation of the pip and pop effect found in Experiment 1. However, one might argue that the visual signals used in Experiments 2a and 2b were ineffective as warning signals. Perhaps the clutter, and especially the dynamics of the displays, made the visual signals difficult to perceive. Experiment 2c, therefore, used a foreperiod task to assess the effectiveness of the different cue types under the dynamic display circumstances of Experiments 1, 2a, and 2b. The crucial question was whether the visual cues could in principle be perceived and used by the observers as a warning signal.

The results of Experiment 2c are presented in Figure 5. The data were subjected to a repeated measures univariate ANOVA, with CTI (0, −100, −200, −300, −600 ms or absent) and cue type (dot, halo, and tone) as within-subject variables. Trials in which participants responded faster than 200 ms and slower than 1,000 ms were excluded from further analysis. This led to a loss of 3.5% of the trials. Overall false alarm rate on catch trials was 6.4%. There were no significant error effects and no apparent trade-offs.

The ANOVA on RTs revealed a significant two-way interaction between cue type and CTI, $F(10, 50) = 6.0$, $p < .001$, $\eta_p = .55$. Separate ANOVAs revealed significant effects of CTI for each cue type: for fixation dot, $F(4, 20) = 25.7$, $p < .001$, $\eta_p = .84$; for halo, $F(4, 20) = 18.2$, $p = .001$, $\eta_p = .78$; for tone, $F(4, 20) = 11.9$, $p < .005$, $\eta_p = .70$. Separate two-tailed *t* tests comparing each CTI with the cue-absent condition revealed significant improvements for all CTIs and all cue types (all $ps < .05$ when CTI was −600 ms, all $ps < .005$ when CTI was greater than −600 ms). Furthermore, there were significant improvements for the auditory cue compared with the visual cues (pooling the data of the latter) when the CTI was 0, $t(5) = 4.4$, $p < .01$, when it was −100, $t(5) = 3.3$, $p < .05$, or when it was −200 ms, $t(5) = 2.8$, $p < .05$, but not when the CTI was −300, $t(5) = -1.5$, $p = .183$, or when it was −600 ms, $t(5) < 1$.

The data of Experiment 2c indeed suggest that the tone was a more effective warning signal than either of the visual cues (which were virtually equally effective), at least at the shorter CTIs. However, the more important conclusion here is that the visual cues were far from ineffective. In line with many other findings on preparation or warning effects (Bertelson, 1967; Los & Van den Heuvel, 2001; Niemi & Näätänen, 1981; Posner & Boies, 1971), there was a clear effect of CTI in the visual conditions. Moreover, for all CTIs, performance with a visual cue was better than performance without such a cue. This demonstrates (a) that these cues were clearly visible (under the same dynamic display circumstances as in the preceding experiments) and (b) that observers could make use of them to prepare for the target signal. Yet note that despite the visual cues being effective warning signals, they did not lead to any improvement whatsoever in Experiments 2a and 2b when they accompanied the target change in the visual search task. At the same time, Experiment 1 demonstrated the auditory cue to be highly effective in improving visual search. This suggests that the warning signal or general alertness hypothesis does not provide an adequate explanation of the pip and pop effect, which seems to be due to multisensory integration instead.

Furthermore, it is worth noting that the warning signal hypothesis does not provide the only possible explanation for the in-

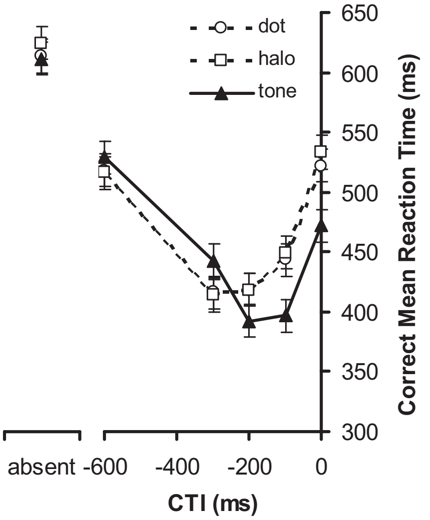

Figure 5. Results of Experiment 2c. Here the correct mean reaction time is plotted as a function of cue type, for each specific cue–target interval (CTI). Note that negative CTIs indicate that the tone was presented before the target. The error bars represent the .95 confidence intervals for within-subject designs, following Loftus and Masson (1994). The confidence intervals are those for the warning signal interaction effects.

creased effectiveness of the auditory cue relative to the visual cues at the shortest CTIs. An equally plausible hypothesis is that, at shorter CTIs, the tone integrated with the visual event (in this case, the offset of the display elements), thus leading to a stronger target signal. It may prove difficult to dissociate these possibilities, because it is not easy to imagine a situation in which a sound and a visual event occur in close synchrony but do not integrate. We leave this issue for the future. For now, we conclude that both visual and auditory cues form effective warning signals, yet only the latter causes improvements in dynamic visual search displays.

Finally, we point to the specific shape of the performance curves as a function of CTI, for the visual as well as for the auditory cues. As has been found many times before (Bertelson, 1967; Los & Van den Heuvel, 2001; Niemi & Näätänen, 1981; Posner & Boies, 1971), these cues are most effective when presented a couple of hundred milliseconds before the target signal (here, 300 ms for visual cues and 200 ms for auditory cues). In Experiment 3, we saw that the time course was quite different for the pip and pop phenomenon, providing further evidence for a dissociation between the warning signal and integration hypotheses.

Experiment 3: Visual Search Is Optimal When Tone and Target Change Are Simultaneous

Experiment 3 was designed to further test the hypothesis that the temporal auditory signal integrates with the visual signal to increase the saliency of the latter. Experiment 3 also provides additional tests of the alternative cuing and alerting accounts. We used the visual search task of Experiment 1 but manipulated the tone–target interval (TTI) so that the tone sounded before (TTI = −150, −100, −50, or −25 ms), simultaneous with (TTI = 0 ms), or after (TTI = 25, 50, or 100 ms) the visual target event. The tone could also be absent.

The literature on alerting effects (Bertelson, 1967; Los & Van den Heuvel, 2001; Niemi & Näätänen, 1981; Posner & Boies, 1971), the cross-modal cuing literature (see, e.g., McDonald et al., 2000; Spence & Driver, 1997), and Experiment 2c all indicate that an auditory cue maximally enhances the response to a visual target when the cue is presented between 100 and 300 ms prior to the visual target. Thus, if the tone merely acts as a warning signal or a cue to start expecting or attending to the visual target event, then performance should benefit the most when the tone precedes this event, so that observers can maximally prepare for the visual change. No benefits would be expected for tones presented after the visual event, because preparation is impossible. In contrast, with a cross-modal integration account, the opposite pattern is expected. That is, greater benefits should be found the closer in time the tones are to the visual event, regardless of whether the tone occurs before or after the event. In fact, a slight asymmetry in performance is expected in favor of tones presented after the visual event, because processing of auditory signals is generally somewhat faster than processing of visual signals (Jaskowski, Jaroszyk, & Hojan-Jezierska, 1990; Lewald & Guski, 2003; Senkowski, Talsma, Grigutsch, Herrmann, & Woldorff, 2007; Wallace, Wilkinson, & Stein, 1996).

Method

Experiment 3 was identical to Experiment 1 except for the following modifications: The tone was presented on most trials but was not necessarily synchronized with the visual target. TTIs varied among −150, −100, −50, −25, 0, 25, 50, and 100 ms. Furthermore, set size was fixed at 48. In Experiment 3, following 2 blocks of practice, participants completed 18 blocks (2 × 8 TTI blocks plus a tone-absent block) of 24 trials each, with order determined by a balanced Latin square. TTI was blocked such that participants could maximally prepare for the upcoming target. Nine new participants participated in Experiment 3 (7 female; mean age = 19.9 years, range = 17–21).

Results and Discussion

The results of Experiment 3 are shown in Figure 6. Overall mean error rate was 7.2%, and the error pattern was consistent with the RT data. A repeated measures ANOVA revealed a significant effect of TTI on RTs, $F(8, 64) = 9.0$, $p < .001$, $\eta_p = .53$. Performance showed a U-shaped function with shorter RTs for shorter intervals between tone and visual target change. Separate two-tailed t tests comparing each TTI condition with the tone-absent condition revealed significant improvements for all TTIs between −100 ms and 100 ms: TTI = −100 ms, $t(8) = 3.0$, $p < .05$; TTI = −50 ms, $t(8) = 4.5$, $p < .005$; TTI = −25 ms, $t(8) = 4.2$, $p < .005$; TTI = 0 ms, $t(8) = 4.5$, $p < .005$; TTI = 25 ms, $t(8) = 4.9$, $p = .001$; TTI = 50 ms, $t(8) = 4.4$, $p < .005$; TTI = 100 ms, $t(8) = 3.6$, $p < .01$. However, no significant improvement was found for −150 ms, $t(8) = 1.5$, $p = .167$. Inconsistent with a warning signal account, search performance was better when the tone was synchronous with the target color change than when it preceded the target color change. Separate two-tailed t tests comparing each negative TTI with a TTI of 0 ms confirms this notion (all $ts > 2.3$, $ps < .05$). On the basis of Experiment 2c, a warning signal account would have predicted optimal performance for the TTI of −150 ms (Bertelson, 1967; Los & Van den Heuvel, 2001;

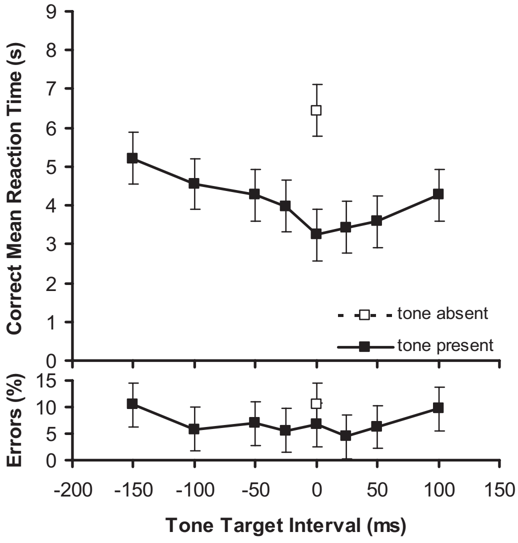

Figure 6. Results of Experiment 3. Mean correct reaction time and mean error percentages as a function of tone–target interval. The error bars represent the .95 confidence intervals for within-subject designs, following Loftus and Masson (1994). The confidence intervals reflect the tone–target interval main effect.

Niemi & Näätänen, 1981; Posner & Boies, 1971). The same time course is predicted on the basis of cross-modal cuing effects (see Experiment 2c; Turatto, Benso, Galfano, & Umilta, 2002). This was clearly not the case here. Note that we do not wish to deny the presence of some warning-related influences on overall RTs. After all, performance in the −150 ms condition was still better than in the tone-absent condition. But these influences just cannot fully explain the pip and pop effect.

Instead, consistent with a cross-modal integration account, search was aided by auditory signals even when these occurred after the visual signal. Furthermore, in accordance with earlier studies, there appeared to be a slight asymmetry in the effect of TTI on performance, in favor of tones lagging behind the visual event. One-tailed t tests confirmed that performance for the TTIs of 25 and 50 ms (tone lagging behind visual event) was better than for the TTIs of −25 and −50 ms (visual event lagging behind tone), ts ≥ 1.9, ps $\leq .05$. Thus, the temporal window of optimal performance that we found is fully consistent with what is regarded as the temporal window of auditory–visual integration and is quite different from that found in the warning signal literature (as well as that found in Experiment 2c). We conclude that the observed benefits in visual search are due to successful binding of the auditory signal with the visual target event and are not due to the auditory signal serving as a mere cue or alerting signal.

Experiment 4: Auditory–Visual Synchrony Automatically Guides Attention

Experiments 1 and 3 indicated that the co-occurrence of auditory and visual signals creates an emergent visual feature that pops out from its background. An important follow-up question is whether this multisensory interaction occurs in an automatic, stimulus-driven fashion or depends on strategic top-down control. To investigate this, we manipulated the validity of the auditory signal. In Experiment 4a, the tone was synchronized with the visual target event on 80% of the trials and was synchronized with a distractor event on the remaining 20%. Thus, strategically, it would make sense to pay attention to the tone and to make it integrate with the visual event if such processes were under top-down control. In other words, we should replicate the search benefits found in Experiment 1. In Experiment 4b, on the other hand, the tone was synchronized with the visual target event on only 20% of the trials and was synchronized with a distractor event on 80%. In this case, it would make sense to ignore the tone and, if possible, prevent integration. If the pip and pop effect is fully subject to strategic control, we should now see it disappear. On the other hand, if the search benefits observed in the previous experiments are mainly due to a stimulus-driven process, then we would expect to find such benefits regardless of the validity of the tone.

We chose two different groups of participants in Experiments 4a and 4b, to minimize possible transfer of search modes between conditions (Leber & Egeth, 2006). Furthermore, in Experiment 4b, eye movements were monitored by recording electro-oculogram (EOG) to make sure that participants remained fixated on the fixation dot. This was done because pilot studies had revealed that observers started search straightaway when they judged the tone to be rather useless (this was in contrast to Experiment 4a, in which participants found the tone useful). Such early eye movements may have adverse effects on the pip and pop effect (e.g., when the target change occurs during a saccade).

Method

Eight new students (5 female; mean age = 19.5 years, range = 18–24 years) participated in Experiment 4a, and 8 new students (7 female; mean age = 21.7 years, range = 18–25 years) participated in Experiment 4b.

The present experiment was identical to Experiment 1, except that we included only set sizes 24 and 48 and included only tone-present blocks. The tone was either synchronized with the color change of the target (80% vs. 20% of the trials in Experiments 4a and 4b, respectively) or synchronized with the color change of a distractor (20% vs. 80% of the trials in Experiments 4a and 4b, respectively). Synchronized item type (distractor or target) was randomly mixed within blocks. There was 1 practice block of 40 trials. After the practice block, participants performed 16 experimental blocks of 40 trials each. Participants were instructed to remain fixated on the fixation dot in both experiments. In Experiment 4b, EOG was measured to make sure that participants adhered to these instructions. Horizontal and vertical EOG were recorded from tin electrodes attached to the outer canthi of each eye and above and below the right eye. The left cheek was used as a ground reference. EOG recordings were amplified, digitized (500 Hz), and processed by NeuroScan (Sterling, VA) hardware and software. Maximum amplitudes were calculated for both channels (vertical EOG and horizontal EOG) for each trial between the onset of the visual search display and the presentation of the first tone. Trials in which either the vertical EOG or the horizontal EOG channel exceeded an 85 μV amplitude were marked as trials in which an eye movement, blink, or other artifact was present.

Results: Experiment 4a

Figure 7 presents the mean RTs for correct responses as well as errors, as a function of set size (24 and 48) and synchronized item type (distractor vs. target). These data were subjected to a repeated measures univariate ANOVA. The overall mean error rate was 3.6%, and the error pattern followed that of the RTs. There were no significant error effects (Fs < 1) and no speed–accuracy trade-offs.

Across conditions, RTs increased significantly with set size, $F(1, 7) = 24.8$, $p < .005$, $\eta_p = 78$. Furthermore, participants responded faster overall when an auditory signal coincided with the color change of the target (2,547 ms) than when the auditory signal coincided with the color change of a distractor (6,001 ms), $F(1, 7) = 22.7$, $p < .005$, $\eta_p = .76$. The interaction between synchronized item type and set size was also significant, $F(1, 7) = 5.8$, $p < .05$, $\eta_p = .46$, confirming that the search slopes were reduced when the auditory signal was synchronized with the target item. Separate analyses revealed a significant effect of set size for synchronized distractor events—120 ms/item, $t(7) = 4.1$, $p < .005$, indicating effortful search—but not for synchronized target events—33 ms/item, $t(7) = 2.0$, $p = .090$, although this approached significance.

Results: Experiment 4b

The data were analyzed in the same way as in Experiment 4a. Figure 7 presents the mean RTs for correct responses as well as

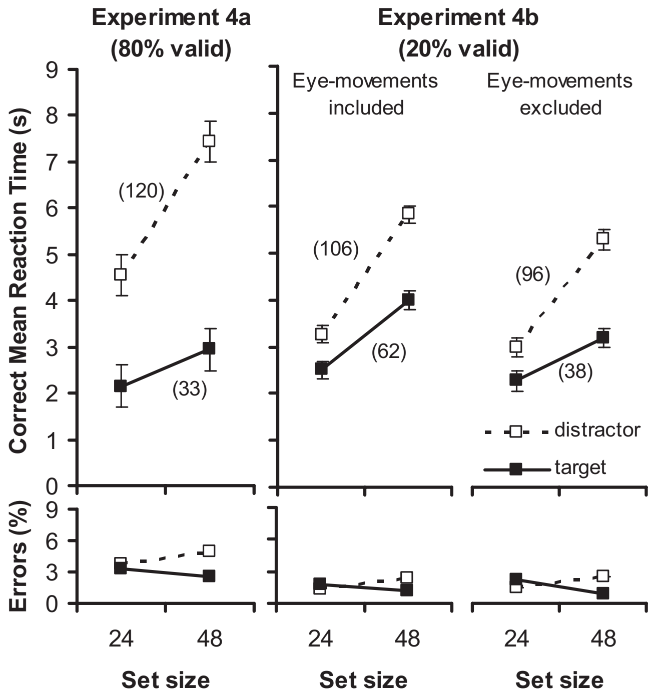

Figure 7. Results of Experiments 4a and 4b. Mean correct reaction time and mean error percentages as a function of set size and synchronized item type. The error bars represent the .95 confidence intervals for within-subject designs, following Loftus and Masson (1994). The confidence intervals are those for the set size interaction effects.

errors, as a function of set size, synchronized item type, and eye movements (included and excluded). When eye-movement trials were included, the overall mean error rate was 1.7%. The ANOVA yielded no significant effects of errors (*F*s < 1), and the error pattern followed that of the RTs. There was no speed–accuracy trade-off. With regard to the RT data, observers were slower with increasing set size, $F(1, 7) = 13.9$, $p < .01$, $\eta_p = .67$. Of importance, participants were faster overall when the tone was synchronized with the visual target (3,251 ms) than when the tone was synchronized with a distractor (4,549 ms), $F(1, 7) = 13.9$, $p < .01$, $\eta_p = .67$. The two-way interaction between synchronized item type and set size was again significant, $F(1, 7) = 8.3$, $p < .05$, $\eta_p = .54$, reflecting the fact that search was more efficient when the visual target color change was accompanied by a tone (62 ms/item) than when a distractor color change was accompanied by a tone (106 ms/item). Sets size effects were significant for both synchronization types: $t(7) = 3.5$, $p < .05$, and $t(7) = 3.7$, $p < .01$, respectively.

Exclusion of eye movement artifacts led to a loss of 32.2% of the trials. However, the overall pattern of results remained the same. The mean error rate was 1.8%, again with no signs of effects or trade-offs (*F*s < 1). With regard to the RT data, responses were overall slower for the larger set size, $F(1, 7) = 14.7$, $p < .01$, $\eta_p = .68$. Of importance, participants were again faster when the tone was synchronized with the visual target (2,732 ms) than when it was synchronized with a distractor (4,141 ms), $F(1, 7) = 17.4$, $p < .005$, $\eta_p = .71$. The two-way interaction between synchronized item type and set size was also significant, $F(1, 7) = 11.7$, $p = .01$, $\eta_p = .63$, indicating that search was more efficient when the visual target color change was accompanied by a tone (38 ms/item) than when a distractor color change was accompanied by a tone (96 ms/item). Set size effects were significant for both conditions, $t(7) = 2.5$, $p < .05$, and $t(7) = 4.2$, $p < .005$, respectively. The improvement in efficiency relative to trials in which eye movements were allowed was significant for synchronized target events—from 62 ms/item to 38 ms/item, $t(7) = 2.7$, $p < .05$—but not for synchronized distractor events—from 106 ms/item to 96 ms/item, $t(7) < 1$.

A between-experiment comparison yielded only an experiment by synchronized item type interaction: $F(1, 14) = 7.2$, $p < .05$, when eye movement trials were included; $F(1, 14) = 6.2$, $p < .05$, and when eye movement trials were excluded (all other *Fs* < 1.2, all *ps* $> .29$). This interaction reflected the fact that the overall RT difference between trials on which the tone was synchronized with a target and those in which it was synchronized with a distractor was greater in Experiment 4a (when the tone was mostly valid) than in Experiment 4b (when it was mostly invalid). More detailed analyses of this interaction revealed no substantial differences other than a trend toward slower RTs on the synchronized distractor trials of Experiment 4a compared to those same trials in Experiment 4b when eye movements were excluded, $F(1, 14) = 2.86$, $p = .113$. A feasible explanation for this slowing is that participants perceived the tone as useful on most trials in Experiment 4a and, as a result, were momentarily more distracted or confused when the tone happened to coincide with a distractor.

Discussion

In Experiments 4a and 4b, we replicated the pip and pop effect as observed in Experiments 1 and 3. The important result was found in Experiment 4b: Search benefited when the target color change was accompanied by a tone, even though this co-occurrence was relatively rare (occurring on only 20% of the trials; on 80% of the trials the tone accompanied a distractor event instead). Thus, making the auditory event rather uninformative about when to expect the target color change did not affect the overall pattern of results. This points toward a substantial contribution of stimulus-driven processes in generating the pip and pop effect. Apparently, the integration of the synchronous auditory and visual signals occurs largely automatically, with the sound guiding attention toward the visual location even when there is little strategic incentive to do so.

Of course, demonstrating a stimulus-driven component does not exclude the possibility of a goal-driven component, and the fact that the tone was overall more effective in Experiment 4a (when it was mostly valid) than in Experiment 4b (when it was mostly invalid) indeed indicates the influence of such a component at least somewhere in the process. Of further interest, Experiment 4b suggests that the pip and pop effect benefits from controlling eye movements. Perhaps observers occasionally miss the visual event, for example, because of saccadic suppression, closed eyes, pushing parts of the display further into the periphery, or other artifacts related to eye movement. Without eye-movement controls, the effect may be underestimated. In any case, one could regard the decision to make an eye movement or not as another strategic component.

One potential caveat of Experiment 4b is that even though the auditory signal was relatively rarely synchronized with the target event (on 20% of the trials), one could argue that it may still have been perceived as useful. That is, the benefits of attending to the sounds on synchronized target trials (the magnitude of which would be in the order of seconds) may have outweighed the costs on synchronized distractor trials (the magnitude of which would be in the order of tens to hundreds of milliseconds). This would explain the benefits found in Experiment 4b even from a strategic perspective. Experiment 5 was therefore designed to provide further evidence for the automatic guidance by synchronized auditory and visual events.

Experiment 5: Pip and Pop Results in Costs When Synchronized With a Distractor

In contrast to the previous experiments, in Experiment 5 the tone was never synchronized with the target event. Instead, the tone was either synchronized with a distractor color change or with no event at all. If the synchronized distractor event automatically captures attention, we should now find a cost in performance relative to the condition in which the tone is not synchronized with any event. Note that such costs would be expected to be relatively small and might therefore drown in the very effortful orientation search we used before (as indicated by the baseline conditions of the preceding experiments). To make search more sensitive to capture effects, we opted for the target to appear by abrupt onset, only after the nontargets had already appeared. This abrupt onset target appeared at various intervals after the synchronized distractor event. If the synchronized distractor draws attention, observers should be less likely to be drawn toward the abrupt onset, resulting in search costs. To control for potential visual effects of the changing distractor on target detection, we also included a condition in which the crucial distractor change was present, but the auditory signal was absent.

Method

The experiment was identical to Experiment 1, except for the following modifications.

Participants. Participants were 16 new students (8 female; mean age = 19.9 years, range = 18–25 years).

Stimuli. As before, the displays consisted of continuously changing distractor items and a target. One of the distractor changes was crucial because it could be synchronized with a tone. The interval between display changes varied randomly among 50, 150, or 200 ms with the constraints that each interval occurred equally often within each cycle and that the synchronized distractor color change (when present) was always preceded by a 150-ms interval and followed by a 200-ms interval (to minimize possible integration with the target; see Experiment 3). The synchronized distractor always changed alone and could change only once per cycle. The synchronized distractor could not change (and hence the tone did not sound) during the first 500 ms of the very first cycle of each trial. The target (a horizontal or vertical line segment) was absent at the onset of the search display. It was presented after a randomly determined interval relative to the synchronized distractor change (when present), as is explained next.

Design and procedure. The tone was present on 80% of the trials (20% sound-absent trials). Of those 80%, the tone was synchronized with a distractor color change on 50% of the trials, at the intervals outlined earlier. On the remaining trials, the tone was present, but there was no synchronized distractor color change. The target appeared on the first, the second, the fourth, or the sixth display change after the crucial distractor change (and the tone), which corresponded to average TTIs of −200, −323, −584, and −860 ms. To control for pure visual effects of the synchronized distractor, we included a condition in which the distractor changed at a TTI of −200 ms, but the tone was absent. The auditory signal, if present, was presented only once on each trial. Synchronized distractor presence (present and absent), tone presence (present and absent), TTI (−200, −323, −584, and −860 ms), and set size (24 and 48) were randomly mixed within blocks. Participants received 1 practice block, followed by 15 experimental blocks of 40 trials each, resulting in 30 trials per cell.

Results and Discussion

Figure 8 presents the correct mean RTs as well as errors. Note that RTs were locked to target onset, which was after the other search items had already appeared (see the *Method* section for Experiment 5). First, the data of the sound-present conditions were subjected to a repeated measures univariate ANOVA, with set size (24 and 48), TTI (−200, −323, −584, and −860 ms), and synchronized distractor presence (present vs. absent) as within-subjects variables. The overall mean error rate in the tone-present condition was 5.5%. Errors increased with set size from 4.5% to 6.5%, $F(1, 15) = 6.8$, $p < .05$, $\eta_p = .31$. All other error effects failed to reach significance (all $F\text{s} \leq 2.2$).

The RTs showed a significant main effect of TTI, $F(3, 45) = 12.1$, $p = .001$, $\eta_p = .45$, as search times decreased with increasing TTI. The same was true for overall search efficiency, resulting in a significant two-way interaction between TTI and set size, $F(3, 45) = 3.0$, $p < .05$, $\eta_p = .16$. This overall pattern suggests that the tone may have had a general alerting effect on the overall RTs. Of importance, however, on top of this effect, there was a highly significant main effect of synchronized distractor presence, $F(1, 15) = 13.9$, $p < .005$, $\eta_p = .48$: Search times were slower when the tone was synchronized with a distractor color change (1,969 ms) than when no such color change was present at the time of the tone (1,715 ms). There was also a tendency for search to become less efficient when a distractor was synchronized with the tone, as indicated by a nearly significant synchronized interaction between distractor presence and set size, $F(1, 15) = 4.0$, $p = .064$, $\eta_p = .21$. No other effects were reliable ($F\text{s} < 1$). Thus, search costs were observed in the conditions in which the auditory signal was synchronized with a visual distractor event compared with the conditions in which the auditory signal was presented without a synchronized event. The results again suggest that auditory–visual synchrony guides attention in an exogenous manner.

An alternative explanation for the observed search costs is that the distractor color change itself captured attention, independent of the tone. Therefore, performance was worse when a distractor color change was present than when a distractor color change was absent. The crucial distractor change may even have masked the target onset. Such effects would be strongest at the shortest TTI (−200 ms). Hence, we performed a second ANOVA, now com-

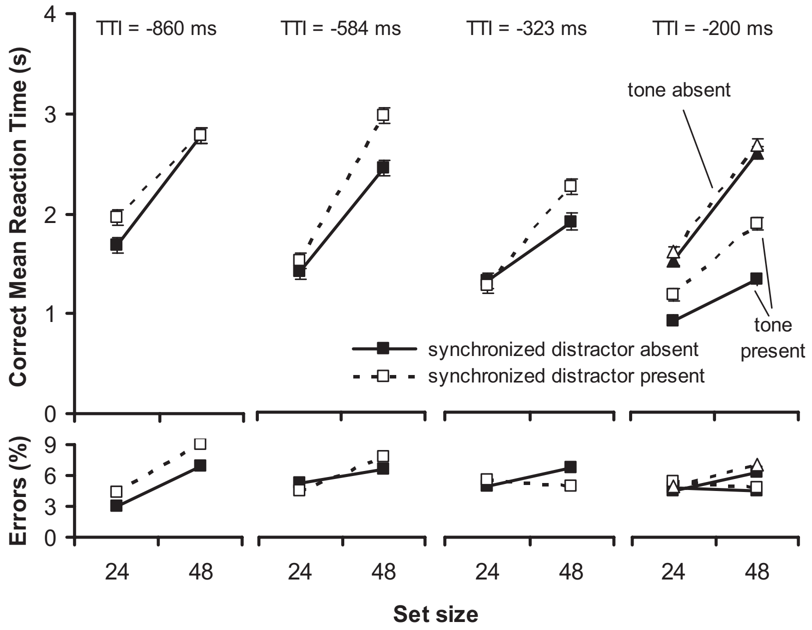

Figure 8. Results of Experiment 5. Mean correct reaction time and mean error percentages as a function of tone–target interval (TTI), set size, synchronized distractor presence, and tone presence. Reaction times were relative to target onset (not display onset, see *Method* section for Experiment 5). The error bars represent the .95 confidence intervals for within-subject designs, following Loftus and Masson (1994). When the TTI was smaller than −200 ms, confidence intervals are those for the synchronized distractor presence main effect. When the TTI was −200 ms, the confidence intervals are those for the interaction between tone presence and synchronized distractor presence.

paring the tone-present condition to the tone-absent condition at a TTI of −200 ms, both with and without an accompanying visual change. The tone-absent conditions are plotted in the last panel of Figure 8. Overall, participants responded faster when an auditory signal was present than when an auditory signal was absent, $F(1, 15) = 11.5$, $p < .005$, $\eta_p = .43$, and responded faster when set size was small, $F(1, 15) = 25.9$, $p < .001$, $\eta_p = .63$. There was a significant interaction between sound presence and synchronized distractor presence, $F(1, 15) = 5.1$, $p < .05$, $\eta_p = .26$. Separate two-tailed t tests comparing each sound presence condition revealed a significant effect of synchronized distractor presence in the sound-present condition, $t(15) = 4.9$, $p < .001$, but not in the sound-absent condition ($t < 1$). In other words, the observed costs are due to the synchronized sound and are not due to the visual change per se. This provides further evidence for the idea that the sound and the visual event interact to create an integrated emergent percept, which attracts attention automatically.

General Discussion

The present study demonstrates that a spatially nonspecific auditory signal can boost the saliency of a concurrent visual signal in a multiobject, dynamic environment (Experiment 1). In other words, a temporal signal affects spatial competition between multiple objects. Furthermore, we show that this attentional guidance by synchronized auditory–visual events is largely automatic. The pip and pop effect, as we have termed it, even occurs when such events involve a distractor on most (Experiment 4) or all (Experiment 5) of the trials.

Is It Alerting?

Can the pip and pop effect be explained in terms of modality-unspecific temporal alerting, rather than in terms of a perceptual integration mechanism? The tone in the present study might have acted as a warning signal, which could, for example, have affected postperceptual response-related stages. We do not deny the possible presence of alerting effects in our experiments. In fact, as mentioned earlier, alerting probably best explains why we still find some RT benefits when the tone is present but not at all synchronized with the target change (e.g., Experiments 3 and 5). However, we believe that the core of the pip and pop effect cannot be explained through alerting effects, especially not when these exert themselves at a postperceptual response level. First, note that the sound carried no information whatsoever on which response should be prepared. Second, if the sound affected only nonspecific response preparation, we would not expect the dramatic effects on search slopes, only on overall RTs. Third, even the overall effect on RTs is unlikely to be explained by alerting alone. In the literature, warning signals have been shown to improve RTs by a fraction of a second at most. Here, we are looking at effects in the order of several seconds for the higher display sizes. This suggests a qualitatively different type of representation is being used for the search process when the sound is present. Fourth, alerting may, of course, also improve perceptual processes (although most theories place it at later stages of the information-processing stream (e.g., Hackley & Valle-Inclán, 2003; Los & Schut, in press; Posner & Boies, 1971). However, Experiment 2 showed that visual cues, although effective warning signals, were not at all effective in improving search efficiency. Furthermore, Experiment 3 showed optimal effects of the tone when it occurred simultaneously or after the visual event, which, assuming that a state of alertness needs time to develop, is inconsistent with a warning signal account. In all, the pip and pop effect follows a time course that is quite different from alerting effects.

Aurally Improved Visual Perception

The present study is not the first to show effects of auditory information on visual search. However, whereas earlier studies demonstrated performance benefits when sound and light were spatially correlated (and thus the sound provided direct knowledge about the target's locations; Bolia et al., 1999; McDonald et al., 2000; Perrott et al., 1990, 1991; Spence & Driver, 1997), the current findings show that this is not always necessary to improve visual search, as long as circumstances allow for successful temporal integration. In the present study, the sound could not act as a top-down signal in the classic sense that it provides goal- or knowledge-driven signals that can raise activity in relevant dimensions in anticipation of the target. This is because the sound did not carry any information on the location, color, or orientation of the target. In Experiments 1 and 3, it provided knowledge only on when a target change occurred.

This study is also not the first to show benefits of uninformative but synchronized sounds with a visual attention task. The findings here are reminiscent of the *freezing effect* reported by Vroomen and De Gelder (2000). They found that presentation durations of targets presented in rapid serial visual presentations (all at a single location) appeared prolonged when accompanied by a sound. However, because all items appeared at the same location, this study did not address the question of how sound may affect the spatial competition between multiple visual events, nor can the freezing effect in itself account for the results here: Simply prolonging the subjective target duration was of no use in our spatial search displays, in which all items were continuously and simultaneously present throughout a trial. In our displays, it was the target change, not the target continuation that was important. Instead, the converse scenario may have been more likely: The increased saliency effects as found here may have contributed to the freezing effect in Vroomen and De Gelder's (2000) study.

The results appear at odds with a recent study by Fujisaki, Koene, Arnold, Johnston, and Nishida (2005), who also looked at the influence of nonspatial auditory signals on visual search. In their study, participants were asked to detect a flashing or rotating visual target among a number of flashing or rotating distractor objects. The target dynamics were synchronized with either amplitude-modulated pips or frequency-modulated sweeps. Unlike our study, however, the presence of these sounds did not result in efficient search, with search slopes reaching as high as 2 s per item. Fujisaki et al. concluded that the integration of auditory and visual events is a serial, attention-demanding process. However, in Fujisaki et al.'s displays, the sound, the distractors, and the target were changing continuously. Combined with the high range of modulation frequencies they used (up to 40 Hz), such circumstances may not have been optimal for spatiotemporal audiovisual integration. For instance, Fujisaki and Nishida (2005) as well as Lewald, Ehrenstein, and Guski (2001) have shown that multisensory integration becomes difficult at temporal frequencies higher than 4 Hz. Furthermore, to find the unique auditory–visual coupling in their displays, the auditory and visual streams needed to be integrated across rather lengthy intervals (up to 2 s), whereas in our paradigm, only single synchronized auditory–visual events occurred, which were temporally isolated.

Early Connections?

By demonstrating powerful and largely automatic integration in multiple object displays, our findings extend earlier work on the spatiotemporal integration of single visual and auditory sources (Dalton & Spence, 2007; Vroomen & De Gelder, 2000). They are also consistent with neurological evidence that such integration occurs relatively early (Falchier, Clavagnier, Barone, & Kennedy, 2002; Giard & Peronnet, 1999; Molholm et al., 2002; Talsma, Doty, & Woldorff, 2007) and effortlessly (Vroomen & De Gelder, 2000). Recent studies demonstrated that an auditory signal can boost the saliency of a concurrently presented visual target, by demonstrating multisensory convergence in low level sensory cortical structures (Schroeder & Foxe, 2004, 2005). For example, auditory activation can be observed in the extrastriate visual cortex (Molholm et al., 2002). Moreover, Giard and Peronet (1999) have shown modulation of visual event-related potentials by concurrently presented auditory stimuli. Here, multisensory interactions were observed extremely early in time (40 ms after stimulus onset), with sources localized at early visual cortex. This further supports the idea that auditory events can affect visual processing in a rapid and exogenous manner. We tentatively propose that in our paradigm, the auditory signal is rapidly relayed to the early visual cortex, allowing it to interact with a synchronized visual event. Thus, the sound would have a rather diffuse, modulating (e.g., multiplicative) function across the visual cortex: It further increases visual signals that must be already present but are by themselves not quite strong enough to demand priority for selection.

How Automatic Is It?

Although we believe the results demonstrate a strong automatic component to the pip and pop effect, some of the results suggest that this is not as strong as other, previously reported automatic attentional capture effects (e.g., for color, Theeuwes, 1992; or abrupt onset, Yantis & Jonides, 1984). As we have already pointed out, even with synchronized sounds, not only were overall RTs quite high (for good reasons), but search slopes never quite reached the values typical for parallel search. Furthermore, Experiment 4 suggested that the effect is susceptible to whether observers make eye movements. These effects may be due to low-level sensory factors, involving, for example, saccadic suppression, increased display density, reduced peripheral vision, or a combination of these. Furthermore, some observers may have been more conservative than others, leading to overall higher search slopes. However, in itself, this kind of explanation already suggests that the bottom-up signal is not always that strong. Therefore, we cannot (nor do we wish to) exclude some top-down influences on the pip and pop effect. For example, the effect may suffer if observers adopt a small, focused attentional window (cf. Theeuwes, 1992), which would suggest that at least some distributed attention is necessary for observers to notice the synchronized event. Such a small attentional window may well correlate with the tendency to make eye movements, thus explaining why filtering out trials in which an early eye movement was made leads to improvements on synchronized target trials. This would be consistent with other evidence that auditory–visual integration requires at least some attention (see, e.g., Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Talsma et al., 2007).

Conclusion

What we tentatively propose here is that in our displays, the binding of synchronized auditory–visual signals occurs rapidly, automatically, and effortlessly, with the auditory signal attaching to the visual signal relatively early in the perceptual process. As a result, the visual target becomes more salient within its dynamic, cluttered environment. However, whether this salient signal is then picked up on by higher order processes and used for further selection may depend on the presence of some distributed mode of attention.

References

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under attention demands. *Current Biology, 15,* 839–843.

Bertelson, P. (1967). The time course of preparation. *Quarterly Journal of Experimental Psychology, 19,* 272–279.

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

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review, 112,* 291–328.

Dalton, P., & Spence, C. (2007). Attentional capture in serial audiovisual search tasks. *Perception & Psychophysics, 69,* 422–438.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

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

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience, 22,* 5749–5759.

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2005). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society B: Biological Sciences, 273*(1588), 865–874.

Fujisaki, W., & Nishida, S. (2005). Temporal frequency characteristics of synchrony–asynchrony discrimination of audio-visual signals. *Experimental Brain Research, 166,* 455–464.

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

Hackley, S. A., & Valle-Inclán, F. (2003). Which stages of processing are speeded by a warning signal? *Biological Psychology, 64,* 27–45.

Jaskowski, P., Jaroszyk, F., & Hojan-Jezierska, D. (1990). Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research, 52,* 35–38.

Kaptein, N. A., Theeuwes, J., & Van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1053–1069.

Leber, A. B., & Egeth, H. E. (2006). It's under control: Top-down search strategies can override attentional capture. *Psychonomic Bulletin & Review, 13,* 132–138.

Lewald, J., Ehrenstein, A., & Guski, R. (2001). Spatio-temporal constraints for auditory–visual integration. *Behavioural Brain Research, 121,* 69–79.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16,* 468–478.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

Los, S. A., & Schut, M. L. J. (in press). The effective time course of preparation. *Cognitive Psychology*.

Los, S. A., & Van den Heuvel, C. E. (2001). Intentional and unintentional contributions to nonspecific preparation during reaction time foreperiods. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 370–386.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407,* 906–908.

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.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin, 89,* 133–162.

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

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,* 389–400.

Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review, 78,* 391–408.

Schroeder, C. E., & Foxe, J. J. (2004). Multisensory convergence in early cortical processing. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 295–309). New York: MIT Press.

Schroeder, C. E., & Foxe, J. J. (2005). Multisensory contributions to low-level, "unisensory" processing. *Current Opinion in Neurobiology, 15,* 454–458.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia, 45,* 561–571.

Shipley, T. (1964, September 18). Auditory flutter-driving of visual flicker. *Science, 145,* 1328–1330.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics, 59,* 1–22.

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,* 691–701.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49,* 83–90.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51,* 599–606.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of*

Experimental Psychology: Human Perception and Performance, 16, 459–478.

Turatto, M., Benso, F., Galfano, G., & Umilta, C. (2002). Nonspatial attentional shifts between audition and vision. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 628–639.

Von Mühlenen, A., Rempel, M. I., & Enns, J. T. (2005). Unique temporal change is the key to attentional capture. *Psychological Science, 16,* 979–986.

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,* 1583–1590.

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology, 76,* 1246–1266.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88,* 638–667.

Wolfe, J. M. (1994). Guided search 2.0. A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 601–621.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 121–134.

Received April 26, 2007
Revision received November 21, 2007
Accepted November 23, 2007 ■

Call for Nominations

The Publications and Communications (P&C) Board of the American Psychological Association has opened nominations for the editorships of **Developmental Psychology**, **Journal of Consulting and Clinical Psychology**, and **Psychological Review** for the years 2011–2016. Cynthia García Coll, PhD, Annette M. La Greca, PhD, and Keith Rayner, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2010 to prepare for issues published in 2011. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations are also encouraged.

Search chairs have been appointed as follows:

- **Developmental Psychology,** Peter A. Ornstein, PhD, and Valerie Reyna, PhD
- **Journal of Consulting and Clinical Psychology**, Norman Abeles, PhD
- **Psychological Review,** David C. Funder, PhD, and Leah L. Light, PhD

Candidates should be nominated by accessing APA's EditorQuest site on the Web. Using your Web browser, go to http://editorquest.apa.org. On the Home menu on the left, find "Guests." Next, click on the link "Submit a Nomination," enter your nominee's information, and click "Submit."

Prepared statements of one page or less in support of a nominee can also be submitted by e-mail to Emnet Tesfaye, P&C Board Search Liaison, at etesfaye@apa.org.

Deadline for accepting nominations is January 10, 2009, when reviews will begin.