Attention, Perception, & Psychophysics
2010, 72 (6), 1654-1665
doi:10.3758/APP.72.6.1654

Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli

MARY KIM NGO AND CHARLES SPENCE
University of Oxford, Oxford, England

Presenting an auditory or tactile cue in temporal synchrony with a change in the color of a visual target can facilitate participants' visual search performance. In the present study, we compared the magnitude of unimodal auditory, vibrotactile, and bimodal (i.e., multisensory) cuing benefits when the nonvisual cues were presented in temporal synchrony with the changing of the target's color (Experiments 1 and 2). The target (a horizontal or vertical line segment) was presented among a number of distractors (tilted line segments) that also changed color at various times. In Experiments 3 and 4, the cues were also made spatially informative with regard to the location of the visual target. The unimodal and bimodal cues gave rise to an equivalent (significant) facilitation of participants' visual search performance relative to a no-cue baseline condition. Making the unimodal auditory and vibrotactile cues spatially informative produced further performance improvements (on validly cued trials), as compared with cues that were spatially uninformative or otherwise spatially invalid. A final experiment was conducted in order to determine whether cue location (close to versus far from the visual display) would influence participants' visual search performance. Auditory cues presented close to the visual search display were found to produce significantly better performance than cues presented over headphones. Taken together, these results have implications for the design of nonvisual and multisensory warning signals used in complex visual displays.

Searching for a target defined by a conjunction of features in a complex and dynamically changing visual display often requires slow and exhaustive search. That is, each item in the search display has to be examined individually in order to determine whether it is a target (e.g., Treisman & Gelade, 1980; Treisman & Sato, 1990; for reviews, see Quinlan, 2003; Treisman, 1996). For interface operators, such as pilots and air traffic controllers, this can pose a major problem because successful conflict resolution often requires not only the rapid detection of potential threats but also the accurate interpretation of those threats (see Pawlak & Vicente, 1996; Previc, 2000; Vicente & Rasmussen, 1992). Under such conditions, the presentation of spatially informative cues may offer an effective means not only of reducing the time needed to detect potential threats but also of improving the subsequent discrimination of those threats. The presentation of spatially informative nonvisual cues—specifically, auditory cues that are colocalized with visual targets—has been shown to reduce visual search latencies by several thousand milliseconds for peripherally located visual targets (i.e., for targets presented at eccentricities exceeding ±90º from central fixation; see, e.g., Perrott, Cisneros, McKinley, & D'Angelo, 1996; Perrott, Saberi, Brown, & Strybel, 1990; Perrot, Sadralodabai, Saberi, & Strybel, 1991). The presentation of spatially uninformative auditory cues has also been shown to reduce visual search latencies for visual targets presented in the central field by more than 200 msec (e.g., Dufour, 1999; Perrott et al., 1996; Perrott et al., 1990; Perrott et al., 1991).

For visual displays, such as computer monitors, one would expect the benefit of having auditory stimuli spatially colocalized with visual targets to be minimal, given that the average minimum audible angle threshold is approximately 1º (Perrott & Saberi, 1990), and computer monitors tend to be small and cluttered. However, the limitations of screen size, clutter, and human auditory localization ability do not appear to hinder the potential advantages of spatially colocalized auditory cues under such conditions. Rudmann and Strybel (1999) investigated whether the presentation of auditory cues that were spatially coincident, displaced by 6º, or else spatially uninformative with regard to the location of the visual target would facilitate participants' visual search performance. Although spatially coincident auditory cues were found to be most effective in terms of reducing participants' visual search latencies, the displaced auditory cues were still effective in terms of enhancing visual search when compared with performance in the uninformative-cue condition.

Meanwhile, Spence and Driver (1997) reported that the auditory precuing of target side can provide an effective means of improving visual target discrimination performance, even when the side of the cue is not predictive of

M. K. Ngo, thuan.ngo@psy.ox.ac.uk

the side on which the visual target is likely to occur (see also Dufour, 1999; for a review, see Spence, McDonald, & Driver, 2004). These crossmodal exogenous spatial attentional cuing benefits were found to be maximal when the auditory cue preceded the visual target by approximately 100 to 300 msec.

Spatially uninformative auditory and vibrotactile cues have also been shown to facilitate participants' visual search performance when they are temporally synchronized with a change in the target stimulus (e.g., Bolognini, Frassinetti, Serino, & Làdavas, 2005; Chan & Chan, 2006; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008, 2009; Vroomen & de Gelder, 2000). For example, Van der Burg et al. (2008, 2009) measured search latencies for visual targets presented in a frequently changing central search field. The participants in these studies had to search for a horizontal or vertical line segment presented among distractor line segments oriented at ±22.5º. On a given trial, the color of the target and distractor stimuli changed regularly from red to green or vice versa. The participants were presented with either an auditory tone cue over headphones (Van der Burg et al., 2008) or a vibrotactile cue to their left hand (Van der Burg et al., 2009). Either the onset of the cue was synchronized with the color change of the target stimulus or no cue was presented. It is important that the cue gave participants absolutely no information concerning the orientation of the target, hence ruling out a response-priming account of any facilitatory effects observed (cf. Spence & Driver, 1997). Nevertheless, Van der Burg et al. (2008, 2009) reported that the presentation of a temporally synchronous cue still resulted in a significant facilitation of participants' visual search performance relative to their performance in those trials in which no cue was presented. In fact, the average reduction in search latencies for temporally synchronous cues was in excess of 1,000 msec (for displays consisting of 24–48 items). Van der Burg et al. (2008) also showed that the search slopes were significantly shallower (indicating more efficient visual search performance) for the auditory cue trials (31 msec/item) than for the no-cue trials (147 msec/item). They concluded that the temporally synchronous auditory or vibrotactile cue and the synchronously color-changing visual target were likely being integrated, thus creating a more salient visual target that somehow "popped out" from among the distractors.[1]

The present study was designed to replicate and extend Van der Burg et al.'s (2008, 2009) findings using the auditory–visual pip-and-pop and vibrotactile–visual poke-and-pop paradigms to explore how visual search performance is modulated by cues that were not only temporally synchronous (presented in synchrony with the color change of the visual target) but also spatially informative with regard to the likely location of the target (in the left or right hemifield). The benefits of having auditory or tactile cues that are either temporally synchronous (e.g., Chan & Chan, 2006; Dalton & Spence, 2007; Van der Burg et al., 2008, 2009; Vroomen & de Gelder, 2000) or spatially informative (Ho, Santangelo, & Spence, 2009; Ho, Tan, & Spence, 2006; Perrott et al., 1996; Perrott et al., 1990; Perrott et al., 1991) have been repeatedly demonstrated in previous research, but the potential benefit of having cues that are both spatially informative and temporally synchronous has not been studied before in the context of a dynamic visual search paradigm.[2]

Experiment 1 was designed to replicate Van der Burg et al.'s 2008 and 2009 studies, in which they compared visual search performance between a cue-absent condition and a temporally synchronized cue-present condition. In Experiment 2, in order to check for any benefits of multisensory over unimodal cuing, we compared the consequences of presenting unimodal auditory and vibrotactile cues with the consequences of presenting multisensory audiotactile cues. Having successfully replicated Van der Burg et al.'s (2008, 2009) basic effects in Experiments 1 and 2, we then went on in Experiment 3 to explore the combination of temporally synchronous and spatially informative cues in order to ascertain whether or not participants' visual search performance could be improved still further by making the nonvisual cue spatially informative with regard to the likely target side. For the first three experiments, we compared the changes in visual search performance following auditory and tactile cuing in order to determine whether one cue was more effective than the other in facilitating visual search. In our final experiment, we examined the influence of cue location (headphones vs. external loudspeaker cones) on the efficiency of participants' visual search performance.

EXPERIMENT 1

Van der Burg et al. demonstrated, in separate studies, the effectiveness of providing participants with temporally synchronous auditory (2008) or vibrotactile (2009) cues in order to facilitate their visual search performance. They reported that the performance benefits observed following vibrotactile cuing were comparable to those observed following auditory cuing, but they did not perform any direct comparisons between the results of these two cue conditions (experiments). The goal of our first experiment was, therefore, to replicate the findings of Van der Burg et al.'s 2008 and 2009 studies and to extend their findings by directly comparing the performance of participants who were presented with auditory cues with that of those who were presented with vibrotactile cues.

Method

Participants. Twenty-two participants from the University of Oxford (12 female), ranging in age from 20 to 38 years (*M* = 27 years), took part in Experiment 1. All of the participants reported normal or corrected-to-normal vision, hearing, and tactile sensitivity. The experiment took approximately 45 min to complete. The participants received a £5 gift voucher for taking part in the study.

Apparatus and Stimuli. The experiment was conducted in a dimly lit, soundproof chamber. The experiment was conducted on a PC using E-Prime (Schneider, Eschman, & Zuccolotto, 2002). The participants sat in a chair approximately 80 cm from a 17-in. CTX PR711F visually flat CRT computer monitor (screen refresh rate = 75 Hz). The auditory stimulus consisted of a 500-Hz tone (sampling rate = 44.1 kHz, 16-bit) with a 60-msec duration (including a 5-msec fade-in and 5-msec fade-out to avoid clicks) that was presented from two Dell A215 PC loudspeaker cones, one of each placed 16º to either side of the center of the visual search display.

The tactile stimulus consisted of a 50-msec, 200-Hz vibration presented through two AEC VBW32 tactors fastened to a waistbelt, with one tactor placed on either side of the participant's waist. In pretesting, 4 participants (who did not take part in the main experiment) subjectively matched the intensity of the vibrotactile stimulus to that of the auditory stimulus. The visual search displays consisted of 24, 36, or 48 red (RGB = 255, 0, 0) or green (RGB = 0, 255, 0) line segments (length = 0.57º visual angle) presented against a black background. The color of each display item was determined randomly. All of the line segments were randomly placed in an invisible 10 × 10 grid (9.58º × 9.58º, 0º–0.34º jitter) centered on a white fixation point, with the sole constraint being that the target was never presented in one of the four central positions. The orientation of each line deviated randomly by exactly ±22.5º from the horizontal or vertical, except for the target, which was presented in either the horizontal or vertical orientation.

Each of nine sequentially presented display screens (making up one complete display cycle) were presented for 50, 100, or 150 msec, with each display duration being repeated randomly three times within a sequence, during which time a certain number of items (of either the target or distractors) changed color. When the set size was 24, one, two, or three distractors changed color. When the set size was 36, one, three, or five distractors changed color. When the set size was 48, one, four, or seven distractors changed color.

When the color of the target changed, it was the only item in the display to do so. When the target changed color, the preceding display duration was always 150 msec, and the following display duration was always 100 msec. The target changed color only once per display cycle, so the average frequency of the target's color changing was 1.11 Hz (i.e., once every 900 msec). The target did not change color during the first 500 msec of the very first display cycle of each trial. Ten different display cycles were generated and presented one after the other (without any gap) until the participant responded or the 10th display cycle had been presented, at which time the whole sequence (of 10 display cycles) was repeated.

Design and Procedure. The modality of the cue was varied on a between-participants basis. The presence versus absence of the cue was varied on an alternating block-by-block basis. Set size was varied on a trial-by-trial basis. During the cue-present blocks, the participants (n = 10) either heard a tone or felt a vibration (n = 12) that was synchronized with the color change of the target. The reaction time (RT) and accuracy of the participants' responses were measured. RTs were measured from the onset of the search display until the initiation of the participant's response. Each trial began with the presentation of a fixation point for 1,000 msec. The search display was presented until the participant pressed a response key. The participants were instructed to press the "z" or "m" key as rapidly and accurately as possible in order to indicate the orientation of the target. The assignment of the targets to the response keys was counterbalanced across participants. Each target orientation was randomly determined and occurred equally often per block of 36 trials. There were four cue-absent and four cue-present blocks of trials presented in a counterbalanced, alternating order. These test blocks were preceded by two 36-trial practice blocks. After each block of trials, the participants received feedback concerning their overall mean accuracy and RT. They were given the opportunity to take a break before starting the next block.

Results and Discussion

All of the participants exceeded 85% correct for the practice blocks at the start of the experiment. (The same criterion was met by all of the participants in the subsequent experiments.) The data from the two practice blocks and from those trials in which the participants made an erroneous response (M = 6.7% of the trials) were excluded from the analysis. Inverse efficiency (IE) scores were computed to correct for any potential speed–accuracy trade-offs (see Spence, Kingstone, Shore, & Gazzaniga, 2001; Townsend & Ashby, 1983). IE scores were calculated by dividing each participant's mean RT for each condition by their proportion of correct responses for that condition. Due to the potential skew in the distribution of the RT data (and, hence, in the IE scores), we converted the IE scores to efficiency scores (−1,000 / IE) to normalize the distributions (see Box & Cox, 1964) and reported the analyses of the E data instead (see Figure 1).

The participants in Experiment 1 reported that the task became easier over time, so we also decided to compare participants' performance on the first half of the experimental session (Blocks 1–4) with that on the second half (Blocks 5–8) to look for any potential practice effects in the data. A mixed univariate ANOVA was conducted on the E data with practice (first vs. second half of the experiment), set size (24, 36, 48) and cue presence (absent, present) as the within-participants factors and cue modality (auditory, tactile) as the between-participants factor. The Huynh–Feldt correction was applied whenever sphericity was violated (alpha = .05).

Analysis of the E scores revealed significant main effects of practice [$F(1,20) = 17.88, p < .001, \eta^2 = .47$], cue presence [$F(1,20) = 7.84, p = .01, \eta^2 = .28$], and set size [$F(2,40) = 49.02, p < .001, \eta^2 = .71$]. Participants' E scores were significantly lower in the second ($M = -.41$) than in the first ($M = -.36$) half of the experiment. Note that lower E scores signify better performance, hence supporting the subjective reports of our participants. The presence of a cue ($M = -.43$) resulted in a significant improvement in participants' performance, as compared

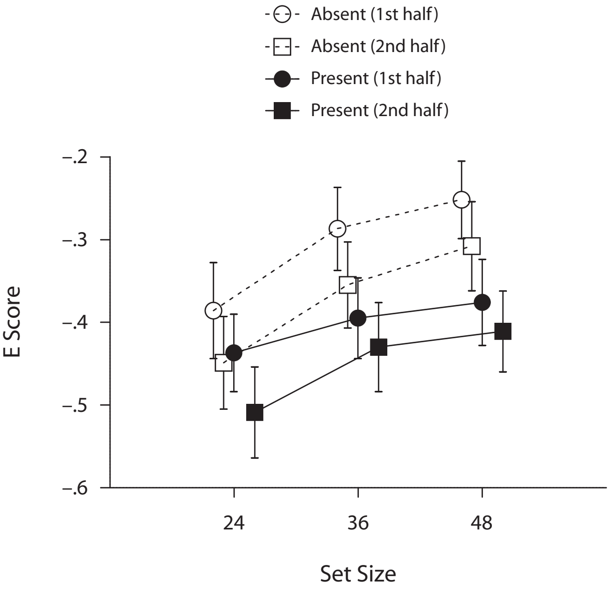

Figure 1. E scores, as a function of practice (first or second half of experiment), set size, and cue presence (absent, present), collapsed across auditory and tactile cue modalities in Experiment 1. Error bars represent the standard errors of the means for each combination of set size and cue presence, collapsed across cue modalities (Bakeman & McArthur, 1996).

with the no-cue condition ($M = -.34$). As the set size increased from 24 to 36, participants' mean E scores increased as well ($-.45$ at set size 24; $-.37$ at set size 36; $-.34$ at set size 48).

There was no main effect of cue modality [$F(1,20) < 1$] and no interaction between cue modality and cue presence [$F(1,20) = 2.13$, n.s.], thus showing that participants' visual search performance was facilitated just as much by the auditory as by the vibrotactile cues in Experiment 1. Mean E scores for the auditory ($-.44$) and vibrotactile ($-.42$) cue-absent conditions were also indistinguishable [$t(20) < 1$, n.s.].

The results of Experiment 1 revealed that the efficiency of participants' visual search performance was improved significantly simply by presenting a temporally synchronous nonvisual (either auditory or vibrotactile) cue at the same time as the color change of the visual target. This was evident in the fact that participants' E scores were significantly reduced in the presence of either of the nonvisual cues. Note that, as the set size increased from 24 to 48, there was a .14-point increase in mean E scores in the no-cue condition (from $-.42$ to $-.28$) and a .08-point increase in those in the cued condition (from $-.47$ to $-.39$). It is important that the nonvisual cues used in Experiment 1 were completely uninformative with regard to the identity (i.e., horizontal or vertical) of the visual target.

The comparison of participants' performance in the two halves of the experiment confirmed their subjective reports that the task became easier over time. Given that the entire experiment took less than 45 min to complete, this result suggests that participants did not require much practice in order to learn how to use the auditory and vibrotactile cues. Thus, the temporally synchronous nonvisual cues appear to be quite intuitive (Ho, Reed, & Spence, 2006) and effective in producing significant improvements in visual search performance that are immediately measurable.

The magnitude of the performance benefit observed in Experiment 1 was numerically very similar to that reported by Van der Burg et al. (2008, 2009). Thus, the results of Experiment 1 successfully replicated Van der Burg et al.'s (2008, Experiment 1; 2009) recent findings using a nearly identical experimental setup. However, in contrast to these recent studies, the vibrotactile warning signals in the present study were presented from participants' waists, rather than from their wrists (as in Van der Burg et al., 2009), and from external loudspeakers, rather than over headphones (as in Van der Burg et al., 2008).

EXPERIMENT 2

Whereas a number of previous studies have shown that bimodal cuing can produce performance benefits that are significantly larger than those seen following unimodal cuing (e.g., Ho et al., 2009, Experiment 2; Santangelo, Ho, & Spence, 2008; Spence & Santangelo, 2009), others have shown bimodal cuing to be no better than unimodal cuing or, even on occasion, worse than unimodal cuing (e.g., Fitch, Kiefer, Hankey, & Kleiner, 2007; Ho et al., 2009, Experiment 1; Lee & Spence, 2009; Lindeman, Yanagida, Sibert, & Lavine, 2003; Santangelo, Van der Lubbe, Olivetti Belardinelli, & Postma, 2006; see also Spence & Ho, 2008). Thus, the answer to the question of whether bimodal (or multisensory) cuing leads to better performance than that seen following the presentation of the best of the unimodal cues seems to depend on the particular task and experimental setting under investigation.

Given this uncertainty, we therefore decided, in Experiment 2, to compare bimodal audiotactile cuing to unimodal auditory and tactile cuing using the same experimental setup as had been used in Experiment 1. In particular, we were interested in determining whether bimodal cues would further improve participants' visual search performance as compared with either unimodal auditory or unimodal vibrotactile cues. If the nonvisual cues facilitate participants' performance in a bottom-up manner, perhaps by increasing the saliency of the visual target (see, e.g., Stein, London, Wilkinson, & Price, 1996; Van der Burg et al., 2008, 2009) relative to the visual distractors, one might expect the bimodal cues to facilitate visual search more than the unimodal cues do. If, however, the nonvisual cues facilitate participants' performance in more of a top-down manner, perhaps by providing the participants with some sort of temporal marker about when the target color change will occur (see Van der Burg et al., 2008; Vroomen & de Gelder, 2000), one might expect a bimodal cue not to be any more effective than the best of the unimodal cues, since both types of cues would most likely provide equivalent temporal information to the participant.

Method

Ten participants (8 female; mean age = 27 years; age range = 21–33 years) took part in Experiment 2. The experiment took approximately 60 min to complete. The experimental setup was identical to that used in Experiment 1, with the following exceptions. (1) The various cue conditions were now interleaved within each block of experimental trials. (2) Within each block of trials, four cue types (no cue, auditory, vibrotactile, audiotactile) were presented equiprobably and in a random order. (3) There was only one practice block at the beginning of the experiment, followed by seven experimental blocks.

Results and Discussion

The E data from the practice block and from those trials in which the participants made an erroneous response ($M = 3.0\%$ of the trials) were excluded from the data analysis. The remaining data were subjected to a repeated measures ANOVA with set size (24, 36, 48) and cue type (no cue, auditory, vibrotactile, audiotactile) as the within-participants factors. The E data are shown in Figure 2.

The analysis of the E data revealed a significant main effect of set size [$F(2,18) = 69.14$, $p < .001$, $\eta^2 = .89$], with E scores increasing as the set size increased. There was also a significant main effect of cue type [$F(3,27) = 6.68$, $p = .015$, $\eta^2 = .43$]. All three of the nonvisual cue conditions (auditory, vibrotactile, audiotactile) resulted in performance that was significantly better than that in the no-cue condition ($p = .013$, $p = .045$, and $p = .026$, respectively), as shown by paired-samples t test post hoc comparisons. However, there were no significant differ-

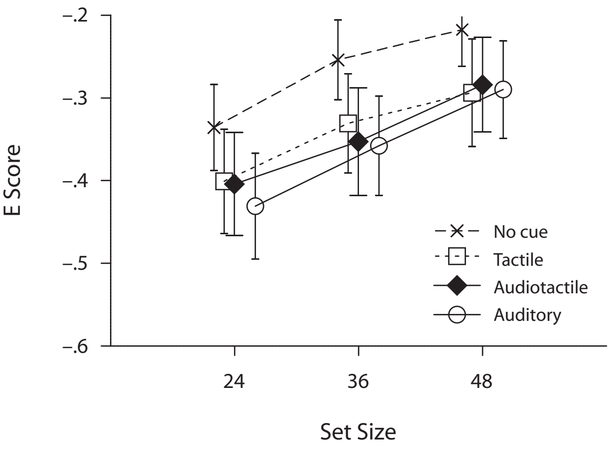

Figure 2. E scores, as a function of set size and cue type (no cue, vibrotactile, audiotactile, auditory) in Experiment 2. Error bars represent the standard errors of the means for each combination of set size and cue type (Bakeman & McArthur, 1996).

ences among the three cued conditions. Mean E scores were higher in the no-cue condition (−.27) than in the auditory- (−.36), vibrotactile- (−.34), or audiotactile-cue conditions (−.35). There was no interaction between set size and cue type [$F(6,54) < 1$, n.s.].

Experiment 2 was designed to test whether bimodal audiotactile cues would facilitate participants' visual search performance more than unimodal auditory or vibrotactile cues would. The results showed no added performance benefits associated with the presentation of the bimodal audiotactile cues, whose onsets were synchronized with the color change of the visual target, as compared with the unimodal cues. No further improvements in participants' visual search performance were observed with the multisensory (audiotactile) cues in the present study, but it is important to note that overall performance was still significantly better than in the no-cue condition and comparable to that seen in the two other unimodal-cue conditions. Thus, rather than increasing the saliency of the visual target in a bottom-up manner, it appears that the nonvisual cues used in the present study facilitated participants' visual search performance in more of a top-down manner, possibly by providing the participants with a temporal marker indicating when the target color change was likely to occur (see Vroomen & de Gelder, 2000).

It is important to note, however, that some question has been raised over the temporal marker account of the facilitation of visual search performance by the results of another of Van der Burg et al.'s (2008) control experiments. In Experiment 2B of that study, Van der Burg and colleagues replaced the synchronous auditory cue with a peripheral visual halo surrounding the entire visual search display. This visual cue provided the same temporal information about the color change of the visual target as the auditory cue, but it did not lead to any facilitation of participants' visual search performance. That is, there was no benefit (relative to the no-cue condition) from this form of visual cuing. However, the temporal marker hypothesis cannot easily account for the lack of facilitation of participants' visual search performance by the presentation of the visual cue. We return to this issue in the General Discussion.

An important difference between the present Experiment 2 and the experiments reported by Van der Burg et al. (2008, 2009) is that we interleaved the four cue conditions (no-cue, auditory, vibrotactile, audiotactile) within each block of experimental trials. Thus, the presentation of each cue type varied on a trial-by-trial basis, whereas, in Van der Burg et al.'s experiments, each of the cue conditions was presented to the participants in separate blocks of trials. This raises the possibility that participants adopted somewhat different "attentional control settings" (Folk, Remington, & Johnston, 1992) in response to the somewhat different task demands required of the participants in Experiment 2 of the present study and of those in Van der Burg et al.'s experiments.

EXPERIMENT 3

Perrott et al. (1990) reported that the presence of a spatially informative auditory cue reduced participants' visual target detection and identification latencies by nearly 200 msec, even when the visual targets appeared within 10º of central fixation. The auditory cue in their study consisted of a 10-Hz click train, whose onset was simultaneous with that of the visual display and was presented until the participants made a response. The auditory cue was presented either from a stationary loudspeaker positioned directly in front of the participant or from a boom-mounted loudspeaker that was positioned directly behind the visual target on each trial. Thus, the onset of the auditory cue could be both temporally synchronous with that of the visual target and spatially informative regarding its precise location. Perrott et al. (1990) reported that target identification latencies were significantly faster following the presentation of the spatially informative auditory cue than following the presentation of the spatially uninformative cue. They went on to conclude that the spatial coincidence of the auditory cue and visual target was necessary for guiding participants' spatial attention efficiently in the direction of the visual target (see also Dufour, 1999).

Experiment 3 was designed to investigate whether the combination of temporally synchronous and spatially informative cuing would further improve participants' visual search performance relative to cues that were temporally synchronous but spatially uninformative, with respect to the location of the target in the left or right hemifield (as in the present Experiments 1 and 2). We also wanted to examine whether the presentation of spatially invalid cues would result in significant visual search costs relative to the spatially uninformative or spatially valid cue conditions (cf. Tan, Gray, Spence, Jones, & Roslizawaty, 2009).

Method

Thirty-eight participants (23 female; mean age = 26 years; age range = 19–44 years) took part in Experiment 3. Eighteen participants received auditory cues, and the remainder received vibrotactile cues. The experimental setup was identical to that used in Experiment 1, with the exception that a temporally synchronous cue was now presented on every trial—that is, the no-cue blocks from Experiment 1 were no longer presented. In half of the blocks, the cues were spatially uninformative. In the remainder of the blocks, they were spatially informative, with 80% of these trials being spatially valid and the remaining 20% being spatially invalid, with regard to the likely location of the visual target in either the left or right hemifield. For spatially valid cue trials, the auditory or vibrotactile cue was presented from the loudspeaker or tactor corresponding to the side on which the target appeared (left or right hemifield of the search screen); the opposite was true for spatially invalid cue trials. On spatially uninformative cue trials, the auditory or vibrotactile cues were presented from both loudspeakers or tactors.

The spatially uninformative and informative cues were presented in separate blocks of experimental trials in order to avoid the possibility that the spatial location of the cues might have an influence on the participants' responses, even in the spatially uninformative blocks (cf. Zampini, Guest, Shore, & Spence, 2005). The order of presentation of the blocks (spatially uninformative, spatially informative) was counterbalanced, and the blocks were presented in alternating order. The experiment took approximately 45 min to complete.

Results and Discussion

The E data from the two practice blocks and from the trials in which the participants made an erroneous response ($M = 2.1\%$ of the trials) were excluded from the data analyses. The remaining data were subjected to a mixed ANOVA with set size (24, 36, 48) and cue type (spatially uninformative, spatially valid, spatially invalid) as the within-participants factors and cue modality (auditory, tactile) as the between-participants factor. Once again, the Huynh–Feldt correction was used whenever sphericity was violated (alpha = .05). The E data are shown in Figure 3.

The analysis of the E data revealed a significant main effect of set size [$F(2,72) = 69.00$, $p < .001$, $\eta^2 = .66$], with E scores increasing as the set size increased. There was also a significant main effect of cue type [$F(2,72) = 5.27$, $p = .019$, $\eta^2 = .13$], with significantly better performance being observed following the spatially valid ($M = -.41$) than following either the spatially uninformative ($M = -.36$; $p < .001$) or spatially invalid ($M = -.35$; $p = .016$) cues. Neither the main effect of cue modality nor any other interactions were significant.

Experiment 3 was designed to investigate whether making the temporally synchronous auditory or vibrotactile cue spatially informative with regard to the likely location (side) of the visual target would further improve participants' visual search performance. To our knowledge, this is the first study to combine temporally synchronous and spatially informative nonvisual cues in order to measure their effectiveness directly in a visual search task. The results showed that the presentation of both spatially informative auditory and vibrotactile cues did indeed enhance the efficiency of participants' visual search performance relative to that seen following either spatially uninformative or otherwise spatially invalid cues (cf. Tan et al., 2009).

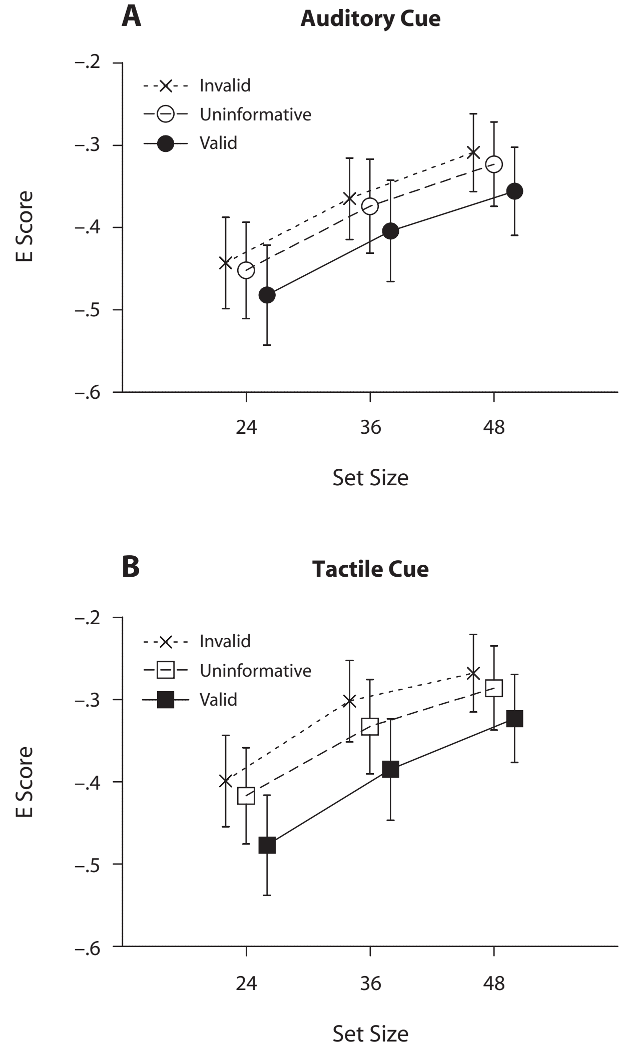

Figure 3. E scores, as a function of set size, cue type (invalid, uninformative, valid), and cue modality (auditory, tactile; panels A and B, respectively) in Experiment 3. Error bars represent the standard errors of the means for each combination of set size and cue type, collapsed across cue modalities (Bakeman & McArthur, 1996).

EXPERIMENT 4

In the present study, the auditory cues came from the same functional region of space as the target (in extrapersonal space), whereas the vibrotactile cues were delivered to the body (in peripersonal space; cf. Previc, 1998, 2000). Ho, Tan, and Spence (2006) suggested that presenting a cue that is directionally congruent (i.e., in the left–right dimension) with respect to a visual target might not be sufficient to produce significant spatial attentional effects (i.e., performance benefits). Rather, they argued that the approximate spatial colocalization of the cue and target (within the same functional region of space) is also necessary to elicit attentional facilitation (cf. Perrott et al., 1990).

If the spatial colocalization of the cue and target is as important as Ho, Tan, and Spence (2006) asserted, then one might expect to see a significant difference in the magnitude of any crossmodal cuing effect as a function of whether the auditory cue was presented close to versus far from the participants' body. On the basis of this line of reasoning, the facilitation of participants' visual search performance should be larger in the latter case (i.e., when the cue and target originate from the same distal region of space). However, if spatial colocalization is not vital in terms of facilitating participants' visual search performance with spatially informative cues, then one would expect to find no significant differences between conditions in which the cue originated close to versus far from the participants' body.

In order to test these two possibilities, we conducted our fourth and final experiment, in which spatially informative auditory cues were now presented either close to the participant's body (i.e., over headphones) or close to the visual display (but far from the participant's body, just as in Experiment 3). We tested whether a difference in the efficiency of participants' visual search performance would be observed between a condition in which the auditory cues were presented via headphones (i.e., far from the visual display) and a condition in which they were played over external loudspeaker cones situated close to the visual display.

Method

Twenty-eight participants (13 female; mean age = 27 years; age range = 20–40 years) took part in Experiment 4. The experimental setup was identical to that used in Experiment 3, but with two exceptions. (1) The spatially informative cues were now always valid with regard to the location of the visual target in the left or right hemifield—that is, no invalid trials were included, as had been the case in Experiment 3. (2) For half of the participants ($n = 14$), auditory cues were presented over closed ear headphones (Beyer Dynamic DT 531); for the other participants, the auditory cues were presented by means of external loudspeakers. The intensity of the auditory cues (based on the combined output of the speaker pair) was subjectively matched for both modes of stimulus presentation. In half of the blocks, the cues were spatially uninformative with regard to the likely side of the visual target; in the remainder of the blocks, they were informative (100% valid) with regard to the target side. The order of presentation of the blocks (spatially uninformative, spatially informative) was counterbalanced, and the blocks were presented in alternating order. The experiment took approximately 45 min to complete.

Results and Discussion

The E data from the two practice blocks and from those trials in which the participants made an erroneous response ($M = 2.7\%$ of the trials) were excluded from the data analyses. The remaining data were subjected to a mixed ANOVA with set size (24, 36, 48) and cue type (spatially uninformative, spatially informative) as the within-participants factors and mode of presentation (headphone, loudspeaker) as the between-participants factor. The E data are shown in Figure 4.

Analysis of the E data revealed a significant main effect of set size [$F(2,52) = 39.36$, $p < .001$, $\eta^2 = .60$], with E scores increasing as the set size increased. Of particular interest, there was a significant main effect of mode of presentation on the E data [$F(1,26) = 6.54$, $p = .017$, $\eta^2 = .20$], with lower mean E scores (i.e., improved performance) being reported for the loudspeaker condition (−.53), as compared with that for the headphone condition (−.38). There was no significant effect of cue type [$F(1,26) = 1.31$, n.s.].

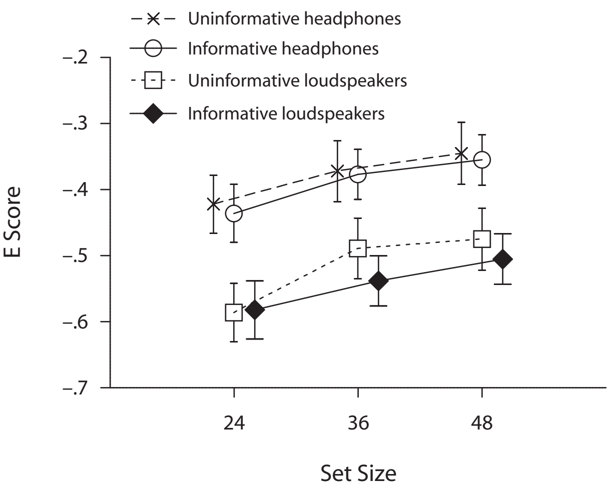

Figure 4. E scores, as a function of set size, cue type (uninformative, informative), and mode of presentation (headphones, loudspeakers) in Experiment 4. Error bars represent the standard errors of the means for each combination of set size and cue type, collapsed across modes of presentation (Bakeman & McArthur, 1996).

Experiment 4 was designed to investigate whether the mode of presentation of the auditory cues would affect the efficiency of participants' visual search performance. Previous research suggests that the spatial colocalization of a cue and target event can yield larger performance benefits than when the cue and target are not colocalized (see, e.g., Bolia, D'Angelo, & McKinley, 1999; Ho, Tan, & Spence, 2006; Perrott et al., 1990). This is indeed what we found in our final experiment. Here, we demonstrated that participants' visual search performance was significantly better (i.e., E scores were significantly lower) when the auditory cues were presented close to the visual display (via external loudspeakers) than when they were presented close to the participant's body (via headphones). Thus, a greater facilitation of visual search performance was observed when the cue and target stimuli originated from the same distal region of space. In fact, presenting the auditory cues via loudspeakers led to a 34% improvement in participants' performance relative to when the same auditory information was presented via headphones.

Interestingly, in Experiment 4, there was no significant difference between participants' visual search performance following either the spatially informative or spatially uninformative auditory cuing. Note that, for the spatially uninformative cue condition, the auditory cue was perceived to originate from the center of the visual display. For the spatially informative cue condition, on the other hand, the sound source was 16º from the center of the display. Thus, one might argue that the uninformative cues were actually more closely aligned with the spatial location of the visual targets than were the informative cues.

Rudmann and Strybel (1999) found that the search latencies for visual targets were significantly higher when auditory cues were displaced from the target by 6º than when they were spatially coincident. The lack of a significant difference between the spatially uninformative and spatially informative cue conditions reported in Experiment 4 is consistent with Rudmann and Strybel's findings, given the fact that the spatially uninformative cues in the present study were, in actuality, more spatially coincident with the target than the informative cues were. The close spatial alignment of the cue and target, in this case, may have been more important than the spatial information (regarding the hemifield in which the visual target appeared) carried by the cues.

GENERAL DISCUSSION

The goal of the four experiments reported in the present article was to extend Van der Burg et al.'s (2008, 2009) recent studies using the auditory–visual pip-and-pop and tactile–visual poke-and-pop paradigms. In those studies, Van der Burg and colleagues investigated the facilitation of participants' visual search performance when nonvisual cues were temporally synchronized with the color change of a visual target whose orientation (horizontal vs. vertical) their participants had to discriminate. Van der Burg and colleagues reported that the presence of either a temporally synchronous auditory or vibrotactile (but, interestingly, not visual) cue resulted in significantly faster search latencies than when no cue was presented.

It is important to note here that, rather than presenting the auditory cues via headphones (as in Van der Burg et al., 2008) and presenting the vibrotactile cues to the participants' hand (as in Van der Burg et al., 2009), the auditory cues in the present study were presented via loudspeakers, and the vibrotactile cues were presented to the participants' waists. Even though the position of the cues differed between the present study and those reported by Van der Burg et al. (2008, 2009), the results of Experiment 1 nevertheless converge, showing that the presence of auditory or vibrotactile cues can give rise to a substantial facilitation of participants' visual search performance, as compared with the performance seen when no cue is present.

We find it important, however, that the results of Experiments 1, 2, and 3 also showed that the auditory and vibrotactile cues were equally effective in enhancing participants' visual search performance. In Experiment 2, we compared the effectiveness of unimodal auditory and vibrotactile cues with that of bimodal audiotactile cues. The results showed that the presentation of the combined audiotactile cues resulted in a performance enhancement that was similar to but no bigger than that observed following unimodal (either auditory or vibrotactile) cuing.

The results of Experiment 3 provide evidence that the presentation of auditory and vibrotactile cues that are not only temporally synchronous but also spatially informative with respect to the likely location of a visual target hidden in a cluttered, dynamic visual search display can lead to an even larger performance benefit. Experiment 3 constitutes the only study to date to have combined both temporally synchronous and spatially informative cues and to have shown that the combination of these two cue features can further improve participants' visual search performance, as compared with cues that are only temporally synchronized (but spatially uninformative), with respect to the dynamic color change of the visual target. The spatially informative nonvisual cues improved participants' visual search performance still further (on the valid trials) relative to the no-cue baseline trials; invalid spatial cuing gave rise to significant search costs (see also Tan et al., 2009).

Several studies have highlighted the importance of the spatial correspondence between auditory and visual stimuli in modulating participants' performance in tasks involving visual targets. For example, an earlier study by Bolia et al. (1999) showed that the presentation of free-field auditory spatial cues in a visual search task led to rapid search latencies characteristic of parallel search, whereas similar virtual auditory cues, while also resulting in significant reductions in search latencies (as compared with when no sound was presented), nevertheless led to performance that was most consistent with serial search (cf. Roberts, Summerfield, & Hall, 2009).

A recent study of crossmodal temporal adaptation conducted by Di Luca, Machulla, and Ernst (2009) has shown that, when repeatedly presented, asynchronous auditory and visual stimuli are colocalized with the auditory stimuli presented via loudspeakers placed behind the light source, changes are observed in the perceptual latency of the visual stimuli. However, when the asynchronously presented auditory and visual stimuli are presented from different locations (i.e., when the auditory stimuli are presented via headphones while the visual stimuli are still presented in front of the participant), this results in changes in the perceptual latency of the auditory stimuli instead. Zampini et al. (2005) also found that, when required to judge the simultaneity of auditory and visual stimuli, participants were more likely to report them as having been presented simultaneously when the stimuli were presented from the same spatial position than when they originated from different spatial positions.

Previous research by Ho, Tan, and Spence (2006) suggests that spatial cuing effects may differ as a function of cue modality (i.e., auditory or vibrotactile). In their study, participants performed a visual discrimination task (making speeded discrimination responses concerning the color change of a car's numberplate) following the presentation of a spatially predictive auditory or vibrotactile cue from either the front or the rear. Ho, Tan, and Spence observed significant facilitatory effects following auditory, but not vibrotactile, cuing. They suggested that nonvisual cues directionally congruent with respect to a visual target may primarily give rise to response priming benefits,[3] whereas cues that share (i.e., come from) the same functional region of space as the target can give rise to attentional facilitation as well. Ho, Tan, and Spence concluded that, in order for a spatial cue to be maximally effective, it should be colocalized with the target event.

In the present study, we found that the spatially informative auditory and vibrotactile cues were equally effective in facilitating participants' visual search performance, even though the cues came from different functional regions of space—that is, the auditory spatial cues came from loudspeakers positioned close to the visual display, whereas the vibrotactile cues were presented close to the body and away from the visual display. This suggests that spatially informative cues need not be colocalized with the visual target—at least not for the type of dynamic visual search task (in a small and cluttered visual environment) we used in the present study. It is also possible that the spatial information was not as important as the temporal information provided by the auditory cue in facilitating participants' visual search performance.

The spatial uncertainty with respect to the location of the visual target may have been limited by the fact that the participants only had to search a small visual field in the present study. The temporal uncertainty with respect to the color change of the visual target and distractors, however, was a major contributor to the difficulty of the task. Thus, the participants' use of the temporal or spatial information provided by the cue may have been based on the demands of the task. In this case, the participants may have strategically used the cue attribute/feature (i.e., temporal synchrony) that was most informative in helping them correctly identify the visual target.

Nevertheless, the results of Experiment 4 showed that overall performance was 34% better (in terms of search latencies) when auditory cues were presented via loudspeakers situated to either side of the visual display rather than via headphones (i.e., presented close to the participant), thus supporting the conclusions of the many previous studies insisting on the spatial colocalization of the cue and target (e.g., Bolia et al., 1999; Di Luca et al., 2009; Ho, Tan, & Spence, 2006). Moreover, although the spatial colocalization of the cue and target proved not to be vital in facilitating participants' visual search performance in the present study, it may have contributed to the lack of any observed improvements in participants' performance following the presentation of the bimodal, as compared with the unimodal, cues, considering the fact that the auditory and vibrotactile cues in the bimodal cue condition originated from different locations.

The fact that the tactile and auditory cues were not spatially colocalized with the visual targets may have required additional cognitive processing from the participants, in the sense that the participants may have had to interpret the fact that the tactile cue located on the body (in their personal space) corresponded with the visual target located in front of the participant (in their peripersonal space; cf. Ho, Tan, & Spence, 2006; Làdavas & Farnè, 2004). Furthermore, although it may have been possible for auditory cues to capture participants' attention, causing them to automatically shift their attention toward the location of the auditory stimulus, the position of the tactile cue may have impeded this type of automatic orienting, due to the fact that the tactile stimulus was presented to the participants' torsos. If participants' attention were to be automatically directed to the location of the tactile stimulation, as might have been the case for the auditory stimuli, this would mean that their attention would have been directed to their torsos and not to the computer monitor. It would be interesting to see whether positioning the auditory and tactile cues in the same functional region of space would result in significantly better performance in the bimodal than in the unimodal cuing condition.

Van der Burg et al. (2008, 2009) suggested that the performance benefits observed following the presentation of synchronous auditory or vibrotactile cues were likely due to the crossmodal enhancement of the saliency of the visual targets. Van der Burg and colleagues ruled out the possibility that the cues merely had a crossmodal alerting effect (cf. McDonald, Teder-Sälejärvi, & Hillyard, 2000; Spence & Driver, 1994, 1997; Vroomen & de Gelder, 2000) by showing that auditory cues did not give rise to a visual search benefit if they were presented 150 msec prior to the change of the color of the target (see Van der Burg et al., 2008, Experiment 3). Instead, the greatest improvements in participants' visual search performance were reported when the auditory cue and target color change were presented simultaneously or within 50 msec of each other (see Van der Burg et al., 2008, Experiment 3). Van der Burg et al. (2008) therefore concluded that, rather than simply alerting participants of the color change of the target, the synchronous presentation of the cue and target somehow resulted in the visual target popping out from among the distractor line segments. They suggested that "the tactile signal boosts the saliency of the concurrently presented visual event, resulting in a salient emergent feature that pops out from the cluttered visual environment, and guides attention to the relevant location" (Van der Burg et al., 2009, p. 63), hence facilitating participants' visual search performance.

Van der Burg et al. (2008, 2009) suggested that the pip-and-pop and poke-and-pop phenomena occur via an automatic, low-level process. Their idea was that the pop-out effect observed in the presence of the auditory or vibrotactile cues was consistent with the results of previous research by Stein et al. (1996), in which weak (i.e., low-intensity) visual stimuli (flashes of light) were rated as being significantly brighter when they were accompanied by pulses of white noise than when no sound was present. Stein et al. argued that this increase in perceived brightness was due to the crossmodal enhancement of the visual stimulus by the simultaneously presented auditory stimulus (one might think of it in terms of "superadditivity"; Stein & Meredith, 1993; Stein & Stanford, 2008). It is, however, important to note that, rather than visual stimulus enhancement, Odgaard, Arieh, and Marks (2003) subsequently argued that Stein et al.'s results might simply have reflected response bias, which is considered to occur as a later (i.e., postperceptual) decisional level of information processing.

It therefore appears that the crossmodal- (visual-) stimulus enhancement explanation cannot fully account for the visual search performance benefits observed in the present study (but see Van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2010). An alternative explanation for the pip-and-pop and poke-and-pop effects, however, comes from research reported by Vroomen and de Gelder (2000). Similar to Stein et al.'s (1996) findings, Vroomen and de Gelder also observed a crossmodal auditory enhancement of visual target discrimination performance (but in a somewhat different experimental paradigm). The participants in their study had to search for a visual target presented amid a rapidly presented sequence of four masked visual stimuli accompanied by a sequence of four tones. Vroomen and de Gelder showed that, when the target was synchronized with the presentation of a deviant tone, participants identified the visual target more rapidly and more accurately. Rather than attributing this to visual stimulus enhancement, however, Vroomen and de Gelder explained their results in terms of a "freezing" phenomenon, in which the tone momentarily captured the visual stimulus within the rapidly presented visual display. Rather than pointing to the producing of a perceptual enhancement of the representation of the visual target, Vroomen and de Gelder proposed that the auditory and visual stimuli became bound together and were integrated into a single, coherent, multisensory event. Thus, the suggestion here is that, in the visual search paradigm, the auditory, tactile, and audiotactile cue and the visual target may become perceptually grouped in a manner that aids in the segregation of the target from the surrounding distractors, which ultimately leads to its improved identification. Even though this hypothesis is still speculative, it warrants further research.

Another possible explanation for the observed improvement in participants' visual search performance is that, rather than enhancing the segregation of the target from among the distractors or the enhancement of the saliency of the visual target, the cue acted as a temporal marker signifying the moment when the target changed color (Watson, Humphreys, & Olivers, 2003). In the pip-and-pop paradigm, the various distractors change color frequently throughout each trial. When the target changed color, however, it was the only item in the search display to do so, and the participants were explicitly told that this would be the case. The temporal information provided by the target may not have been sufficient to make the target stand out among the distractors, due to the possible overload in visual information. However, it is possible that the temporal information provided by the nonvisual cue may have allowed the participants to inhibit the selection of the items that changed color at different times from the target (i.e., the distractors).

The observed improvements in the participants' visual search performance following the presentation of temporally synchronous and spatially informative nonvisual cues appears to reflect the operation of somewhat different underlying mechanisms—although, as we have just seen, the exact mechanisms remain unclear. Temporally synchronous nonvisual cues seem to have an influence on some feature of the visual target, in terms of its timing or its saliency, and the spatially informative nonvisual cues seem to have more of an attentional effect, which may result from the shifting of participants' attention toward the location of the spatially informative cue rather than any enhancement of visual saliency. Given that 80% of the spatially informative cue trials were valid, the participants might have strategically chosen to search only one side of the screen, resulting in the halving of the effective set size.

Whatever the correct explanation of the mechanism(s) underlying the effects reported here turns out to be, we believe that they have important implications for the design of nonvisual and multisensory warning signals in the automotive and air traffic management industries. Not only is the timing of the cue and target events vital in terms of producing significant performance improvements in tasks requiring the rapid detection and identification of potential threats (e.g., in air traffic management, pilot operations, vehicular navigation, collision avoidance, and military operations), but the spatial coincidence of the warning signal and target also seems to be necessary in order to produce warning signals that are maximally effective for these situations (Ferris & Sarter, 2008; Jones, Gray, Spence, & Tan, 2008; Sarter, 2000; Spence & Ho, 2008). It appears that the benefits observed following the presentation of temporally synchronous and spatially informative cues result from the operation of somewhat different underlying mechanisms. Eye-movement data may prove valuable in revealing subtle differences in the ways these two cue attributes facilitate human performance.

No further performance enhancements were observed with the temporally synchronous multisensory audiotactile cues than with the unimodal cues in the present study, but other research has shown spatially informative multisensory cues to be highly effective in terms of capturing attention and producing significant performance benefits in laboratory settings, as well as in applied domains, such as the aviation and automotive industries (Fitch et al., 2007; Sarter, 2000, 2001; Spence & Ho, 2008). One suggestion that has emerged recently is that bimodal (i.e., multisensory) cues may preferentially capture a person's spatial attention under conditions where their attention is otherwise engaged in a secondary task (for a review, see Spence & Santangelo, 2009). Therefore, it will be necessary to conduct further research on spatially informative bimodal cues in cluttered visual search settings in order to assess whether there truly are no additional performance benefits from combining auditory and vibrotactile cues in the visual search paradigm.

AUTHOR NOTE

The research presented here was supported in part by a Clarendon Fund Scholarship to the first author from Oxford University. We thank Erik Van der Burg for providing the original pip-and-pop program and for his continued support. Correspondence concerning this article should be addressed to M. K. Ngo, Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, England (e-mail: thuan.ngo@psy.ox.ac.uk).

REFERENCES

Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods, Instruments, & Computers*, **28**, 584-589.

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

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, **160**, 273-282.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, **26**, 211-252.

Chan, A. H. S., & Chan, K. W. L. (2006). Synchronous and asynchronous presentations of auditory and visual signals: Implications for control console design. *Applied Ergonomics*, **37**, 131-140.

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

Di Luca, M., Machulla, T.-K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, **9**(12, Art. 7), 1-16.

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

Ferris, T. K., & Sarter, N. B. (2008). Cross-modal links among vision, audition, and touch in complex environments. *Human Factors*, **50**, 17-26.

Fitch, G. M., Kiefer, R. J., Hankey, J. M., & Kleiner, B. M. (2007). Toward developing an approach for alerting drivers to the direction of a crash threat. *Human Factors*, **49**, 710-720.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 1030-1044.

Ho, C., Reed, N., & Spence, C. (2006). Assessing the effectiveness of "intuitive" vibrotactile warning signals in preventing front-to-rear-end collisions in a driving simulator. *Accident Analysis & Prevention*, **38**, 988-996.

Ho, C., Santangelo, V., & Spence, C. (2009). Multisensory warning signals: When spatial correspondence matters. *Experimental Brain Research*, **195**, 261-272.

Ho, C., Tan, H. Z., & Spence, C. (2006). The differential effect of vibrotactile and auditory cues on visual spatial attention. *Ergonomics*, **49**, 724-738.

Jones, C. M., Gray, R., Spence, C., & Tan, H. Z. (2008). Directing visual attention with spatially informative and spatially noninformative tactile cues. *Experimental Brain Research*, **186**, 659-669.

Làdavas, E., & Farnè, A. (2004). Neuropsychological evidence for multimodal representations of space near specific body parts. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 69-98). Oxford: Oxford University Press.

Lee, J.-H., & Spence, C. (2009). Feeling what you hear: Task-irrelevant sounds modulate tactile perception delivered via a touch screen. *Journal on Multimodal User Interfaces*, **2**, 145-156.

Lindeman, R. W., Yanagida, Y., Sibert, J. L., & Lavine, R. (2003). Effective vibrotactile cuing in a visual search task. In M. Rauterburg, M. Menozzi, & J. Wesson (Eds.), *Proceedings of the Ninth IFIP TC13 International Conference on Human–Computer Interaction* (pp. 89-98). Zurich: IOS.

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

Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. *Perception & Psychophysics*, **65**, 123-132.

Pawlak, W. S., & Vicente, K. J. (1996). Inducing effective operator control through ecological interface design. *International Journal of Human–Computer Studies*, **44**, 653-688.

Perrott, D. R., Cisneros, J., McKinley, R. L., & D'Angelo, W. R. (1996). Aurally aided visual search under virtual and free-field listening conditions. *Human Factors*, **38**, 702-715.

Perrott, D. R., & Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *Journal of the Acoustical Society of America*, **87**, 1728-1731.

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.

Previc, F. H. (1998). The neuropsychology of 3-D space. *Psychological Bulletin*, **124**, 123-164.

Previc, F. H. (2000). Neuropsychological guidelines for aircraft control stations. *IEEE Engineering in Medicine & Biology Magazine*, **19**, 81-88.

Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin*, **129**, 643-673.

Roberts, K. L., Summerfield, A. Q., & Hall, D. A. (2009). Covert auditory spatial orienting: An evaluation of the spatial relevance hypothesis. *Journal of Experimental Psychology: Human Perception & Performance*, **35**, 1178-1191.

Rudmann, D. S., & Strybel, T. Z. (1999). Auditory spatial facilitation of visual search performance: Effect of cue precision and distractor density. *Human Factors*, **41**, 146-160.

Santangelo, V., Ho, C., & Spence, C. (2008). Capturing spatial attention with multisensory cues. *Psychonomic Bulletin & Review*, **15**, 398-403.

Santangelo, V., Van der Lubbe, R. H. J., Olivetti Belardinelli, M., & Postma, A. (2006). Spatial attention triggered by unimodal, crossmodal, and bimodal exogenous cues: A comparison of reflexive orienting mechanisms. *Experimental Brain Research*, **173**, 40-48.

Sarter, N. B. (2000). The need for multisensory interfaces in support of effective attention allocation in highly dynamic event-driven domains: The case of cockpit automation. *International Journal of Aviation Psychology*, **10**, 231-245.

Sarter, N. B. (2001). Multimodal communication in support of coordinative functions in human–machine teams. *Journal of Human Performance in Extreme Environments*, **5**, 50-54.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools.

Spence, C. J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 555-574.

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

Spence, C. [J.], & Ho, C. (2008). Tactile and multisensory spatial warning signals for drivers. *IEEE Transactions on Haptics*, **1**, 121-129.

Spence, C. [J.], Kingstone, A., Shore, D. I., & Gazzaniga, M. S. (2001). Representation of visuotactile space in the split brain. *Psychological Science*, **12**, 90-93.

Spence, C. [J.], McDonald, J., & Driver, J. (2004). Exogenous spatial-cuing studies of human crossmodal attention and multisensory integration. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 277-320). Oxford: Oxford University Press.

Spence, C. [J.], & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing Research*, **258**, 134-142.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, **8**, 497-506.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, **9**, 255-266.

Tan, H. Z., Gray, R., Spence, C., Jones, C. M., & Roslizawaty, R. M. (2009). The haptic cuing of visual spatial attention: Evidence of a spotlight effect. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human Vision and Electronic Imaging XIV* (Proceedings of SPIE Vol. 7240, pp. 1-12). San Jose, CA: SPIE–IS&T Electronic Imaging.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.

Treisman, A. [M.] (1996). The binding problem. *Current Opinion in Neurobiology*, **6**, 171-178.

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

Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 459-478.

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

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

Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2010). *Early multisensory interactions affect the competition among multiple visual objects*. Manuscript submitted for publication.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions of Systems, Man, & Cybernetics*, **22**, 589-606.

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

Watson, D. G., Humphreys, G. W., & Olivers, C. N. L. (2003). Visual marking: Using time in visual selection. *Trends in Cognitive Sciences*, **7**, 180-186.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio–visual simultaneity judgments. *Perception & Psychophysics*, **67**, 531-544.

NOTES

1. Note, though, that, even for the smallest set size, average RTs exceeded 2,000 msec. This might lead one to question whether the visual targets ever really popped out of the display.

2. Perrott et al. (1990) used a spatially colocalized 10-Hz auditory click train whose onset was synchronous with the onset of the visual target. In their study, a visual target was always present in the search field, and its shape, orientation, and color never changed. By contrast, in the present study, we use the term *temporally synchronous* to refer to the fact that the auditory, vibrotactile, or audiotactile cues were presented simultaneously with each change of the color of the target stimulus. Thus, the important difference between the cues used in the present study (and in Van der Burg et al., 2008, 2009) and those used by Perrott et al. (1990) is that, in the former, the cues were presented synchronously with the dynamic visual stimuli (making the cues also dynamic themselves), whereas in the latter, the cues were presented synchronously with only the first onset of the visual stimuli.

3. It is worth noting that, in the applied domain, response priming effects (which tend to be of greater magnitude) are often of more relevance than spatial attentional cuing effects. Here, because of the orthogonal nature of the participants' task (discriminating the orientation of a line segment), response priming effects should not have had any influence on their performance (cf. Spence & Driver, 1997).

(Manuscript received August 14, 2009;
revision accepted for publication April 6, 2010.)