

Auditory–visual contextual cuing effect

JUN-ICHIRO KAWAHARA

National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan

Under incidental learning conditions, a spatial layout can be acquired implicitly and facilitate visual searches (the contextual cuing effect). Whereas previous studies have shown a cuing effect in the visual domain, the present study examined whether a contextual cuing effect could develop from association between auditory events and visual target locations (Experiments 1 and 2). In the training phase, participants searched for a T among Ls, preceded by 2 sec of auditory stimulus. The target location could be predicted from the preceding auditory stimulus. In the test phase, the auditory–visual association pairings were disrupted. The results revealed that a contextual cuing effect occurs by auditory–visual association. Participants did not notice the auditory–visual association. Experiment 3 explored a boundary condition for the auditory–visual contextual cuing effect. These results suggest that visual attention can be guided implicitly by crossmodal association, and they extend the idea that the visual system is sensitive to all kinds of statistical consistency.

It is widely accepted that visual attention is guided to potentially informative locations in two main ways. First, as can be experienced in our daily activity, deployment of attentional focus is subject to endogenous control. For example, when we are driving a car, attentional focus can be allocated endogenously to a particular object, such as a traffic sign or a truck immediately ahead. At the same time, the attentional focus can be captured exogenously by abrupt changes in the visual field, such as the sudden appearance of a motorcycle from behind the truck. These two ways have been investigated in the literature, respectively, as top-down or goal-directed control, and bottom-up or stimulus-driven control of attentional focus (e.g., Egeth & Yantis, 1997; Folk, Remington, & Johnston, 1992; Kim & Cave, 1999; Mackeben & Nakayama, 1993).

However, the factors contributing to the guidance of visual attention are not limited to these two components. Studies have also shown that the visual system is sensitive to consistency within the visual context and uses consistency as a cue to guide visual attention to target locations, resulting in enhanced visual search performance. Specifically, Chun and Jiang (1998) developed a contextual cuing paradigm and suggested that consistent structures, such as repetitively presented global spatial layouts, help spatial attentional focus to be allocated to a specific target location. In a typical experiment, participants searched for a target, T, among several distractors (rotated Ls). The location of the target and the distractors (i.e., the layout of items) defined the spatial context. The participants, who were not informed about the experimental manipulations, received two types of displays, the repeated and nonrepeated conditions. The displays under the repeated condition were presented in every block of trials. The target location in each of this type of display could thus

be predicted from the spatial layout of the distractors. In contrast, the displays under the nonrepeated condition were created with the constraint that the location of the target be the same as in the repeated condition; the layout of distractors, however, was randomly determined from trial to trial. Therefore, the layout of the distractors did not provide clues for target location under this condition; that is, the targets under the repeated condition appeared consistently within the same spatial context, while those under the nonrepeated condition appeared in new spatial contexts. By the time participants had conducted several hundred trials under these conditions, they were faster at finding the targets under the repeated condition than under the nonrepeated condition. Interestingly, they did not notice that some displays (i.e., the repeated condition) were presented many times throughout the session and that they were unable to recognize the repeated displays. On the basis of these results, Chun and Jiang (1998) concluded that participants learned the repeatedly presented spatial context and the acquired context, presumably as implicit knowledge, and that this acted as a cue for the visual system to direct spatial attention to locations where the targets were most likely to appear.

On the basis of this finding, the following studies have revealed that the visual system can extract and learn consistencies such as associations between objects (Chua & Chun, 2003; Chun & Jiang, 1999; Endo & Takeda, 2004; Fiser & Aslin, 2001), spatial structures (Kawahara, 2003), and temporal order (Howard, Howard, Dennis, Yankovich, & Vaidya, 2004; Olson & Chun, 2001). These studies unanimously suggest that the human ability to acquire visual consistency is flexible and strong. Therefore, the effect of contextual cuing is found not only in controlled laboratory settings using abstract objects, but is also com-

J. Kawahara, jun.kawahara@aist.go.jp

mon in the natural environment (Brockmole & Henderson, 2006). Jiang and colleagues found that the learning effect is preserved for at least a week (Chun & Jiang, 2003) and that the learning capacity is so large that participants showed learning effect for 60 different spatial layouts (Jiang, Song, & Rigas, 2005).

The contextual cuing effect is not limited to associations between layout and target locations in a single visual search trial. Recently, Ono, Jiang, and Kawahara (2005) found that the spatial context acquired in one trial influences the visual search for target events in the next trial. In the training phase of their study, the spatial layout of a trial was predictive of the target location in the next trial. When the predictive value was removed in the testing phase, the search performance was impaired. Ono et al. (2005) argued that their findings support the *ubiquitous statistical learning account*, in which the visual system is sensitive to all kinds of statistical consistency and will make use of predictive information, whether such information is presented in a single trial or across trials. Although ample evidence indicates that the human cognitive system is sufficiently powerful to acquire various types of visual consistency (Chun, 2000; Fiser & Aslin, 2001; Schacter & Buckner, 1998), it is unclear whether the cognitive system can associate nonvisual context, such as auditory events, with forthcoming target locations.

In the present study, I investigated whether a contextual cuing effect could be obtained from associations between auditory events and visual target locations. Participants engaged in a training phase, during which they heard an auditory stimulus followed by a visual search trial. The target location of the trial was reliably predictive from a preceding unique auditory stimulus. In the testing phase, participants were subjected to similar stimuli of auditory–visual pairs, but the specific pairings between the leading auditory stimulus and the following target location was disrupted, so that the auditory stimulus was not predictive of the target location. I examined whether search performance (reaction time, or RT) improved during the training phase, and whether such improvement disappeared when the association was removed in the testing phase. On the basis of the ubiquitous statistical learning account as a governing rule applicable to multimodal stimulus pairs, I predicted that search performance would be affected by the experimental phases: The performance would improve during the training phase and would be impaired during the testing phase. However, if the ubiquitous statistical learning account were limited to the events within the visual domain only, auditory–visual contexts would not improve search performance.

As described further below, Experiments 1 and 2 indicated a contextual cuing effect from associations between auditory events and visual target locations, and Experiment 3 showed a boundary condition of this auditory–visual contextual cuing effect.

EXPERIMENT 1

To create an auditory–visual context, I prepared meaningless speech sounds using a digitally recorded voice played

backward, and associated each sound with a specific target location in a visual search display. I used an inconsistent transfer procedure (Olson & Chun, 2001). The experiment consisted of two phases: a training phase, in which the auditory–visual association was consistently provided during a visual search task (Blocks 1–5, 80 trials each) and a testing phase, in which the association was removed (Block 6, 80 trials; the inconsistent transfer condition). Participants heard a meaningless speech sound for 2 sec followed by a visual search display. I ran the same experiment as a control condition using a different group of participants, but for this group, the association created in the training phase was maintained in the testing phase (the consistent transfer condition; see below). I predicted that if participants acquired the associations between the auditory stimuli and the visual spatial locations, their RTs in the training phase would decrease as the training progressed, while the RTs in the testing phase would increase in the inconsistent transfer group but decrease in the consistent transfer group, because deployment of attentional focus would be guided to the target locations by acquired contextual memory.

Method

Participants. Forty experimentally naive students (split equally between the inconsistent and consistent transfer conditions) from Hiroshima University volunteered for course credit. All had normal or corrected-to-normal visual acuity and normal color vision, according to self reports.

Visual stimuli. Visual stimuli were displayed on a CRT monitor controlled by a PC/AT compatible computer. The viewing distance was approximately 60 cm. Each visual search display consisted of a target, T, that was tilted 90° to the left or to the right, and 11 distractors, Ls, that were rotated 0°, 90°, 180°, or 270° from the upright position. The orientation of the target was determined randomly on every trial. Each of these items subtended $1.3^\circ \times 1.3^\circ$ in visual angle.

The items were presented within randomly selected locations from an invisible 8×6 matrix subtended $25.0^\circ \times 18.5^\circ$. To avoid colinearity between items, the location of each item was displaced slightly ($\pm 0.8^\circ$) from the vertical and horizontal axes. The items were presented on a gray background, and colored either red, green, blue, or yellow. Three items for each color were randomly assigned. When a target appeared at a repeated location, the target's color was maintained.

Auditory stimuli. The auditory stimuli were presented through headphones. To create meaningless speech sounds, I saved a recording of a female voice reading in Japanese, played backward, on a computer. The voice was divided into 40 sound files, each of 2-sec duration. Sixteen of the 40 files were randomly chosen to be used for the experiment as auditory context and an additional 16 files were used as fillers for the recognition test.¹ The remaining 8 files were used for the practice trials. These auditory stimuli were clearly perceptible as speech-like in an unfamiliar language.

Experimental design and procedure. One within-subjects factor (block) and one between-subjects factor (consistency) were used. The participants received the training phase, followed by the testing and recognition phases. The training phase consisted of five blocks of 80 trials. At the beginning of the experiment, 16 different target locations were selected randomly from 48 possible locations, with the restriction that each quadrant of the display contained four target locations. Each of these locations was paired with one of the 16 sound files. However, these sound files were not associated with the target orientation or response, because the target orientation was determined randomly on every trial. These pairs served as 16 different but consistent auditory–visual contexts throughout the training phase. The 80 trials in each block of the training phase were

further divided into five subblocks, in each of which the 16 pairs of auditory–visual contexts were presented in random order. The locations of the distractors were determined randomly from trial to trial. The distractor layouts were recorded and the experimental program compared each record with a newly generated layout so that none of the previously presented layout was repeated by chance. I wrote the present computer scripts in MATLAB, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

The experiment started with the visual search trials of the training phase, preceded by eight practice trials. Before each trial, a small fixation cross appeared in the center of the screen. The observers initiated each trial by pressing a space bar. After a delay of 500 msec, an auditory stimulus was presented. Immediately following its offset, a visual search display was presented. Participants were instructed to ignore the auditory stimuli and to indicate the orientation of the target as quickly and accurately as possible. They were given no information about the relationship between the auditory stimuli and the target locations. Participants were allowed to rest at the end of each block and whenever else they wanted.

Of the six blocks in the visual search trials, the last block served as the testing phase. In the last (sixth) block (trials 401–480) under the inconsistent transfer condition, I shuffled the 16 auditory–visual locations that had previously been maintained, so that although the auditory stimuli were repeated and the target locations were also repeated, their consistent association was removed. If the auditory–visual association had been available for crossmodal contextual cuing, the effect should not have transferred to the sixth block under the inconsistent transfer condition. Under the consistent transfer condition, the auditory–visual pairs were maintained. Both groups of participants were blind to these manipulations.

After the training and testing phases, the experimenter asked the participants whether they had noticed the repetition of the auditory stimuli during the visual search trials. The experimenter also asked whether they had noticed any associations between the auditory and visual stimuli. Following these questions, the experimenter informed the participants of the consistency between specific auditory stimuli and target locations. The participants then engaged in the recognition test that consisted of the “old” auditory stimuli that had been presented in the training and testing phases and the “new” auditory stimuli that had not been presented (16 trials each). These 32 auditory stimuli were presented sequentially in random order and the participants indicated whether they had heard each stimulus in the visual search session by pressing one of the two keys on the keyboard. After discriminating “old” or “new” for each auditory stimulus, they were asked to choose a target location out of 48 possible locations associated with the auditory stimulus. None of the responses in the recognition phase was speeded.

Results

Visual search task: RTs. The mean RT for each correct trial within each block was calculated under each condition. The RTs that exceeded 3 standard deviations (*SD*) from each participant’s mean RT were discarded in this and subsequent experiments. The group means are plotted in Figure 1 for each condition as a function of training block. Blocks 1–5 comprised the training phase and Block 6 was the testing phase. An ANOVA on the RTs in the training phase of the inconsistent transfer condition revealed a significant main effect of block [$F(4,76) = 31.6, MS_e = 10,080.1, p < .001$]. A planned comparison between Blocks 5 and 6 indicated that RTs in Block 6 were significantly longer than those in Block 5 [$t(19) = 2.3, p < .05$]. Similarly, I ran the same ANOVA on the RTs in the training phase under the consistent transfer condition and found a significant main effect of block [$F(4,76) = 41.9, MS_e = 9,106.4, p < .001$]. Unlike in the

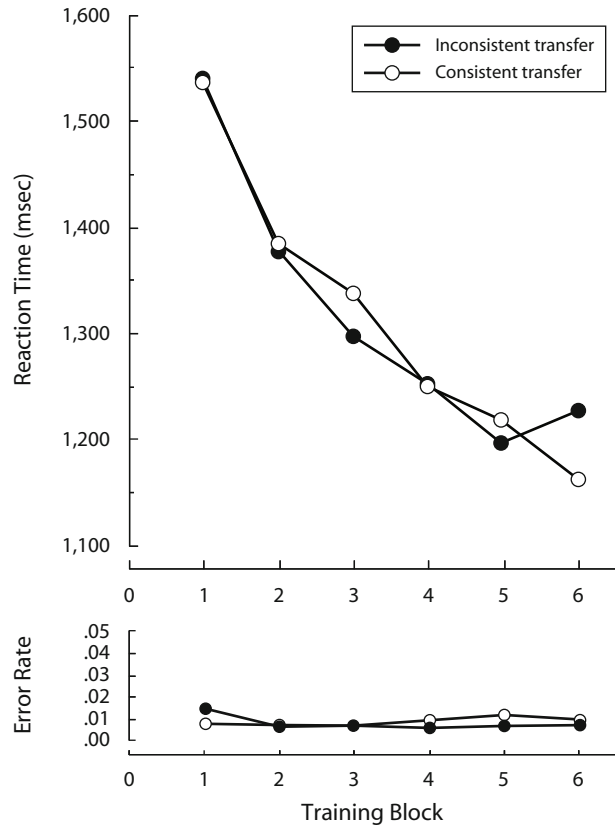


Figure 1. Mean reaction times and error rates for target identification under the inconsistent and consistent transfer conditions in Experiment 1, as a function of block. The training phase comprised the first five blocks, and the testing phase comprised the sixth.

inconsistent transfer condition, under which a significant increase in RT occurred in Block 6, the RT under the consistent transfer condition decreased in Block 6 [$t(19) = 2.1, p < .05$].

To examine the effect of removing the consistent association between the auditory stimuli and the target location, I ran a two-way ANOVA on the RT of the last two blocks, with block (5 vs. 6) as a within-subjects factor and consistency (the inconsistent and consistent transfer conditions) as a between-subjects factor. The analysis indicated that the main effects of the block and the consistency were not significant ($F < 1$ for either condition). More importantly, however, a significant interaction was observed between block and consistency [$F(1,38) = 9.4, MS_e = 4,573.4, p < .005$]. The tests of the simple main effects indicated that from Block 5 to 6, the RT increased significantly under the inconsistent transfer condition [$F(1,38) = 5.2, p < .05$], but decreased significantly under the consistent transfer condition [$F(1,38) = 4.2, p < .05$].

Visual search task: error rates. The bottom of Figure 1 shows the mean error rates in identifying the target under each condition as a function of block. None of the similar analyses in the RT data revealed any significant main effects of blocks and consistency. Furthermore, no interaction was detected between the two factors.

Recognition. Under the inconsistent transfer condition, all but one of the participants reported that they had noticed the repetition of the auditory stimuli. Mean recognition accuracy of the auditory stimuli was 78.1%, which was significantly higher than the chance level [$t(19) = 5.6, p < .01$]. The hit and false alarm rates were 79.1% and 23.0%, respectively. Under the consistent transfer condition, all participants reported that they noticed the repetition of the auditory stimuli. Mean recognition accuracy of the auditory stimuli was 73.1%, which was significantly higher than the chance level [$t(19) = 8.1, p < .01$]. The hit and false alarm rates were 76.8% and 28.5%, respectively. No statistically significant difference was detected in the recognition accuracy of the auditory stimuli between the inconsistent and consistent transfer conditions [$t(38) = 1.3, n.s.$].

Regarding the association between the auditory stimuli and target locations, three of 20 experimental participants reported that they found some consistencies. Out of these three, two participants reported that they had a vague feeling about one or two pairs of sound–location associations, but that they could not specify the pairs. The other reported a false relationship between target and distractor colors and shapes. The mean correct recognition rate of exact location–auditory stimulus pairs was 0.8% (chance was 2.1%). To examine whether participants preserved a coarse memory of target locations, I also scored the data in terms of whether the participants chose the same quadrant of the display as a target location. However, the results indicated that the mean choice rate was 29.3%, which was not significantly higher than chance [$t(19) = 1.3, n.s.$]. For the consistent transfer group, two out of 20 participants reported that they noticed the sound–location consistency. One reported two pairs and the other reported vague feelings about an association. Mean exact location report was 1.2% and the mean of the quadrant choice was 24.2%, which was not significantly higher than chance [$t(19) = 0.3, n.s.$]. No significant difference was detected in the exact target location report rate between the inconsistent and consistent transfer groups, Welch t test [$t(20.7) = 0.7, n.s.$]. Similarly, no significant difference was found in the quadrant match rate between the inconsistent and consistent transfer groups [$t(38) = 1.0, n.s.$].

Discussion

Experiment 1 indicated that when a target location in a visual search display was consistently paired with a preceding auditory stimulus, participants were able to extract the consistency implicitly. As shown in Figure 1, visual search performance improved when the consistency was available during the training phase. Critically, as the interaction between consistency (in the inconsistent vs. consistent transfer conditions) and block (fifth and sixth) indicated, search performance improved further when consistency was maintained during the critical block (sixth). However, when consistency was removed, search performance deteriorated. This is the first study to report that a contextual cuing effect occurs as the result of an association between auditory and visual stimulus events.

The high recognition performance of the auditory stimuli indicates that the participants explicitly remembered the auditory stimuli even though they were told to ignore them. More importantly, while the auditory stimuli themselves were explicitly retrievable, the associations between the auditory stimuli and the visual target locations were not explicitly accessible; in other words, participants were unable to use the auditory stimuli consciously to aid their visual search. This finding is the same as those in the contextual cuing literature. For example, in a typical study requiring participants to search for a T-target among rotated L-distractors (e.g., Chun & Jiang, 1998), participants can consciously recognize that they saw four different types of rotated Ls in the display, but they cannot explicitly report the target locations on the basis of the distractor layouts. In this sense, the fact that participants consciously remember that four different types of specific rotated Ls existed is equivalent to the present result—that the participants could recognize some of the specific auditory stimuli. Similarly, previous findings that participants could not recognize the association between the distractor layout and the target locations are equivalent to the present result that participants were unable to recognize the auditory–visual association. An important point is that the associations between the auditory stimuli and visual target locations remained implicit. This pattern of the results is consistent with visual contextual cuing studies, in that the cuing effect occurred without explicit recognition of the critical association between the target location and other consistencies in the visual environment, such as distractor locations (e.g., Chun & Jiang, 2003).

The present results indicate that humans have the ability to extract consistency information from the environment and to use it as a cue to guide visual attention. Specifically, participants implicitly learn associations between auditory stimuli and target locations. This finding is consistent with the ubiquitous statistical learning account (Ono et al., 2005) and extends it: The present results suggest that the idea is not limited to the visual domain. Rather, the cognitive system monitors various stimulus modalities seeking any consistencies across the domains.

EXPERIMENT 2

This experiment had two main aspects. First, I included a random condition in which the auditory stimuli were not associated with a particular location. This condition was critical for determining whether auditory–visual pairing produced benefits for this kind of visual search performance. Second, in Experiment 1, I found that visual search performance was improved by repeatedly presenting a meaningless sound that was associated with a specific target location, 2 sec before the visual search display. This means that deployment of attention gains advantage from guidance of contextual memory, shifting the attentional focus to a target location sometime between the onset of an auditory stimulus and the identification of the target. To obtain a clue to when this attentional guidance occurs, the auditory stimulus was presented 1 sec before the visual search display. If the contextual cuing process

occurred during the latter half of the presentation of the auditory stimulus, the cuing effect would be greatly reduced or eliminated.

Method

Forty-eight experimentally naive students (16 each under the inconsistent, consistent, and random sound–target location association conditions) participated for pay. All had normal or corrected-to-normal visual acuity and normal color vision, according to self reports. The method was the same as that of Experiment 1, with the following exceptions: First, under the random condition, the auditory stimuli were not associated with a particular location, and the target location was determined randomly for every trial; second, the onset of the auditory stimulus was 1 sec before the onset of the visual search display. Because the duration of the auditory stimuli was 2 sec, the latter half of an auditory stimulus was presented with the visual search display.

Results and Discussion

The mean RT for all correct trials within each block was calculated under each condition. The group means are plotted in Figure 2 under each condition as a function of training block. An ANOVA on the RT data in the training phase of the inconsistent transfer condition revealed a significant main effect of block [$F(4,60) = 36.8, MS_e = 6,914.4, p < .001$]. A *t* test comparing the RTs in Blocks 5 and 6 indicated that RTs in Block 6 were significantly longer than those in Block 5 [$t(15) = 3.2, p < .01$]. An ANOVA on the RTs in the training phase of the consistent transfer condition indicated a significant main effect of block [$F(4,60) = 43.9, MS_e = 5,118.0, p < .001$]. The RT under the control condition did not differ significantly between Blocks 5 and 6 [$t(15) = 1.3, p > .1$].

To compare the effect of consistent transfer, I ran a two-way ANOVA on the RT of Blocks 5 and 6 as a within-subjects factor and consistency (the inconsistent and consistent transfer conditions) as a between-subjects factor. The main effects of block and consistency were not significant. However, a significant interaction was detected between block and consistency [$F(1,30) = 10.0, MS_e = 3,966.0, p < .005$]. The tests of the simple main effects indicated that the RT increased significantly from Block 5 to 6 under the inconsistent transfer condition [$F(1,30) = 9.6, p < .005$], but no significant difference was seen between the RTs in Blocks 5 and 6 under the consistent transfer condition.

To examine whether consistent mapping of the auditory stimuli and target locations benefits visual search performance, a two-way ANOVA was conducted for RT with condition (consistent transfer or random) as a between-subjects factor and block as a between-subjects factor. The main effect of condition was marginally significant [$F(1,30) = 3.98, MS_e = 238443.4, p = .055$], and that of block was significant [$F(5,150) = 35.9, MS_e = 8,573.3, p < .001$]. The interaction between these factors was significant [$F(5,150) = 2.54, MS_e = 8,573.3, p < .05$], indicating that benefits occurred due to auditory–visual pairing in a visual search.

To compare the three consistency conditions directly, a two-way ANOVA was conducted for RT with condition (inconsistent, consistent transfer, or random) as a

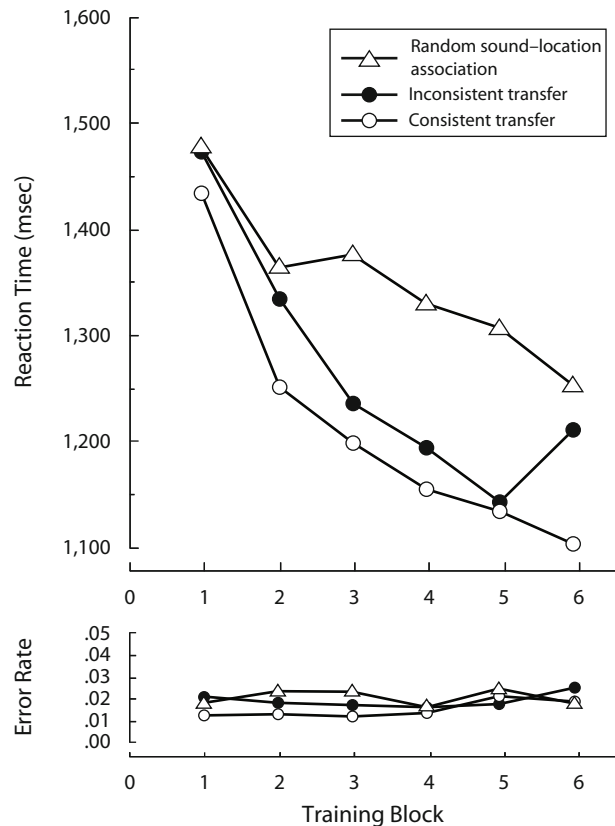


Figure 2. Mean reaction times and error rates for target identification under the inconsistent transfer, consistent transfer, and random conditions in Experiment 2, as a function of block. The training phase comprised the first five blocks, and the testing phase comprised the sixth.

between-subjects factor and block (5 or 6) as a between-subjects factor. The interaction between these factors was significant [$F(2,45) = 3.79, MS_e = 10,798.0, p < .05$]. The tests of simple main effects on this interaction indicated that the RT increased only in the inconsistent transfer condition [$F(1,45) = 4.34, p < .05$]. The RT decreased in the random condition [$F(1,45) = 3.61, p = .06$], but no statistical difference was found in the consistent condition [$F(1,45) < 1, n.s.$]. These results indicate that making auditory–visual pairing inconsistent in the test phase hindered visual search. The pattern of the error rates and recognition was essentially the same as that found in Experiment 1.² Experiment 2 replicated the basic finding of Experiment 1: The results revealed the participants’ ability to learn a consistent association between a target location and its preceding auditory stimuli.

EXPERIMENT 3

In Experiment 3, I examined whether any conditions restricted the acquisition of crossmodal association. Specifically, I focused on one of the distinct characteristics in the results of preceding experiments: relatively high rec-

ognition rate of the auditory stimuli. This is a notable difference from conventional visual contextual cuing studies, in that the displays were not distinctive and could not be explicitly recalled, but could cue the target location. Thus, one could question whether this relatively higher rate of recognizing the auditory stimuli is a prerequisite for successful learning of the auditory–visual context. If the correct recognition were not necessary for the acquisition of the auditory–visual context, such a finding would extend the applicability of the ubiquitous statistical learning. If, however, it turns out that the correct recognition is a necessary condition for the acquisition of the auditory–visual context, such a finding would still be informative, because it would characterize the learning principle and clarify its applicable range of conditions.

To answer this question, I reduced the distinctiveness of the auditory stimuli by replacing the auditory stimuli of reversed playback of a human voice with random sequences of artificial tones from a touch-tone telephone. I expected participants to be unable to recognize the auditory stimuli because—compared with the reversed voice stimuli—the distinctiveness of the tone stimuli was greatly diminished. I expected that a less distinctive stimuli would be less recognizable (Eysenck, 1979; Wolters & Verduin, 1982). If the successful recognition of the auditory stimuli is a prerequisite for the acquisition of an association of crossmodal context, no contextual cuing effect should be obtained in the present experiment. However, if recognition is irrelevant to the acquisition of the association in a crossmodal context, the cuing effect is expected to occur, even when participants are unable to recognize the auditory stimuli.

Method

Twenty-four experimentally naive students (12 each under the inconsistent and consistent transfer conditions) from the National Institute of Advanced Industrial Science and Technology subject pool participated for pay. All had normal or corrected-to-normal visual acuity and normal color vision, according to self reports. The stimuli, experimental design, and procedure were the same as those in Experiment 2, except that (1) the auditory stimuli of reversed playback of a human voice were replaced with electronic tones from a touch-tone telephone (standardized by ITU-T Recommendation Q.24), and (2) the random condition was not included. I recorded a few minutes of a random sequence of keypress tones from a telephone that generated ten sinusoidal tones of two frequencies assigned to each number key. The recorded sequence was divided into 40 sound files each of 2-sec duration. Of the 40 files, 16 were chosen at random to be used as auditory context in the training, and 16 were used as fillers in the recognition test. The remaining 8 files were used for practice trials.

It is important to note that, when they were directly compared with each other, the auditory stimuli used in this experiment displayed differences. As confirmation, ten new participants were presented with two pairs of sequential auditory stimuli used in this experiment, with 1-sec separation between the stimuli. Half of the pairs consisted of the same stimuli, and the other half consisted of different stimuli (50 pairs each). These pairs were presented in random order. The participants were to report whether the two sequential stimuli sounded the same or not by pressing one of two keys on the keyboard. The hit rate was 95.2% and the false alarm rate was 2.8%. This result indicates that the stimuli used in Experiment 3 were clearly distinguishable.

Results

Visual search task: RTs. Figure 3 illustrates the mean RT for all correct trials as a function of training block under the inconsistent and consistent transfer conditions. An ANOVA on the RT data in the training phase under the inconsistent transfer condition revealed a significant main effect of block [$F(4,44) = 20.2$, $MS_e = 8,481.9$, $p < .001$]. No statistical difference was observed between the RTs in Blocks 5 and 6 [$t(11) = 1.9$, n.s.]. An ANOVA on the RTs in the training phase of the consistent transfer condition indicated a significant main effect of block [$F(4,44) = 18.0$, $MS_e = 10,133.3$, $p < .001$]. The RT of Block 6 under the consistent transfer condition was significantly shorter than that of Block 5 [$t(11) = 3.2$, $p < .01$].

As is shown in Figure 3, no effect of consistent transfer was detected. A two-way ANOVA on the RT of Blocks 5 and 6 as a within-subjects factor and consistency (the inconsistent and consistent transfer conditions) as a between-subjects factor revealed a main effect of block [$F(1,22) = 10.3$, $MS_e = 3,200.6$, $p < .005$], indicating that RTs for Block 5 were longer than those of Block 6. The main effect of consistency and the interaction were not significant.

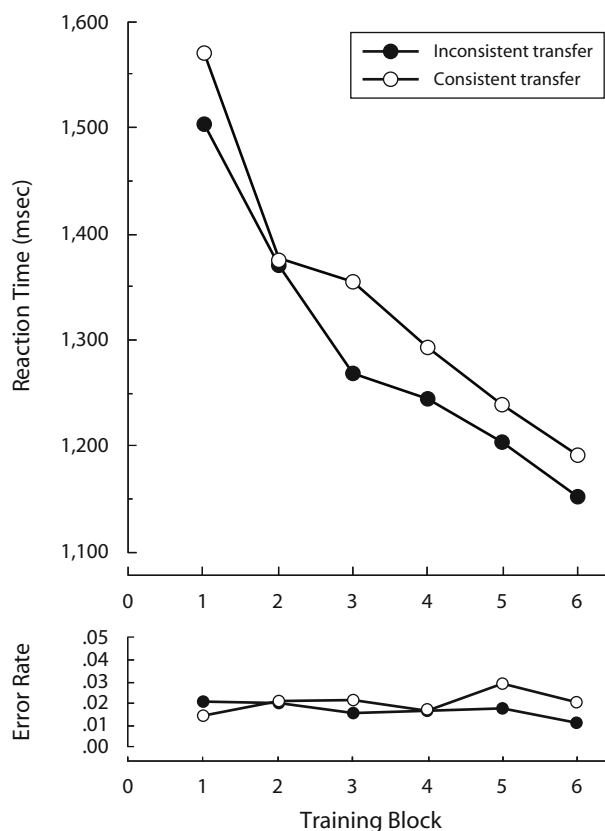


Figure 3. Mean reaction times and error rates for target identification under the inconsistent and consistent transfer conditions in Experiment 3, as a function of block. The training phase comprised the first five blocks, and the testing phase comprised the sixth.

Visual search task: error rates. The mean error rates under each condition are plotted in the bottom of Figure 3 as a function of block. A similar ANOVA conducted for the RT data revealed no significant main effects of blocks or consistency and no significant interaction between the two factors.

Recognition. Of 12 participants under each of the inconsistent and consistent transfer conditions, 3 failed to notice that the auditory stimuli were repeatedly presented under both conditions. Under the inconsistent transfer condition, mean recognition accuracy of the auditory stimuli was 52.6%, which was not significantly higher than chance level [$t(11) = 0.77$, n.s.]. The hit rate was 65.1%, which did not differ from the false alarm rate of 59.9% [$t(11) = 0.80$, n.s.]. Under the consistent transfer condition, mean recognition accuracy of the auditory stimuli was 53.6%, which was not significantly higher than the chance level [$t(11) = 1.05$, n.s.]. The hit rate was 64.6%, which did not differ from the false alarm rate of 57.3% [$t(11) = 1.09$, n.s.]. The recognition accuracy of the auditory stimuli in the inconsistent transfer group did not differ statistically from those under the consistent transfer condition [$t(22) = 0.22$, n.s.].

For the inconsistent transfer group, the mean correct recognition rate of an exact target location and mean quadrant match were 3.6% and 29.2%, respectively. Neither was significantly higher than chance [$t(11) = 0.29$, n.s., and $t(11) = 0.36$, n.s., respectively]. For the consistent transfer group, the mean exact location report was 2.1% and the mean of the quadrant choice was 22.9%, which was not significantly higher than chance [$t(11) = 0.01$, n.s., and $t(11) = 0.22$, n.s., respectively]. No significant difference was detected in the mean exact location report between the inconsistent and consistent transfer groups, Welch's t test [$t(17.0) = 0.8$, n.s.]. Similarly, no significant difference was found in the mean quadrant choice rates between the two groups [$t(22) = 0.1$, n.s.].

Discussion

Using the procedure with reduced auditory distinctiveness, participants were unable to recognize the auditory stimuli. Under such circumstances, no contextual cuing effect occurred, even when the auditory stimulus was predictive of the following target location. These results characterize a boundary condition on ubiquitous statistical learning, in that the mere existence of a consistent association between the auditory stimuli and the visual target locations is not sufficient for the acquisition of auditory–visual context. Although most of the participants noticed that some auditory stimuli were repeatedly presented, their recognition rate on the auditory stimuli was no higher than chance. These results suggest that such a vague memory representation was insufficient; rather, conscious recognition of the auditory stimuli, as demonstrated in Experiments 1 and 2, is required for a successful contextual cuing effect when the present procedure is used. This is in sharp contrast to the results of previous visual contextual cuing studies, in which participants were asked to discriminate previously presented old layouts or new layouts. In the literature, it has been frequently re-

ported that the contextual cuing effect occurred without any conscious recognition of the spatial layouts (e.g., Chun & Jiang, 1998). However, what was common to the previous and the present studies was that participants were unable to access explicitly their knowledge of a consistent association between the background contexts (distractor layouts or auditory stimuli) and the target locations.

GENERAL DISCUSSION

As has been demonstrated in visual contextual cuing studies, the human cognitive system is sufficiently powerful to extract various types of consistency and to learn them. The system was able to use learned context to assist with the deployment of visual attention. Because previous studies have focused exclusively on the learning of visual environments, it has remained unclear whether the ability to learn environmental consistencies as a cue for visual attention is restricted to the visual domain or open to other stimulus domains. Thus, the present study examined whether a context cuing effect could be obtained between auditory events and visual target locations.

The results of Experiment 1 indicated that when a target location in a visual search display was consistently paired with a preceding meaningless sound, participants were able to learn the consistency. The visual search performance improved when the consistency was available (i.e., the training phase) and it improved further when the consistency was maintained (the testing phase of the consistent transfer group). However, when the association between the sound and the target locations became inconsistent (the testing phase of the inconsistent transfer group), search performance deteriorated. It should be noted that in the testing session under the inconsistent transfer condition, the consistency was removed by disconnecting the existing sound–location pairs and shuffling them to create new pairs. Therefore, no new sounds or new locations were introduced in the testing session, and the only difference between the inconsistent and consistent transfer conditions in the testing session was the availability of a consistently maintained association between a sound and a location during the training session.

The amount of cuing effect observed as a result of breaking the auditory–visual association between the 5th and 6th block was about 70 msec (Experiment 1) to 110 msec (Experiment 2), which is very similar to the contextual cuing effect in the visual domain obtained in the previous studies in our laboratory (Kawahara, 2003; Ono et al., 2005). Importantly, comparison of the random and consistent transfer condition in which consistent mapping of the auditory stimuli and target location was maintained clearly indicated that such pairing benefits visual search performance under the present experimental setting. As is illustrated in Figure 2, the visual search performance, after the auditory–visual association was destroyed, was essentially the same as that of the random condition. The performance from Blocks 5 and 6 improved equally in the random and consistent transfer conditions. There is also a general improvement across blocks, as can be seen in typical contextual cuing paradigms. The difference between

the consistent and inconsistent transfer conditions is due to the manipulation of the consistency of the auditory–visual association.

This is the first report of the ability to acquire an association of consistency between auditory stimulus and visual stimulus locations through repetitive presentations. Because the task used in the present study was a typical visual search task that required focal attention (Wolfe, 1997), the results suggest that the preceding auditory stimuli guided visual attention to a specific target location.³ To the best of our knowledge, only one previous study reported that the human cognitive system can use nonvisual contextual cues. Nabeta, Ono, and Kawahara (2003) found that participants who trained with repetitive visual layouts showed the spatial contextual cuing effect when they were tested with a haptic search task. Whereas Nabeta et al.'s study showed that visually acquired spatial context could be transferred to another stimulus modality, no previous studies have shown that the cognitive system can associate nonvisual context with visual target locations.

Of course, studies have shown crossmodal links for exogenous and endogenous covert spatial orienting of attention in terms of auditory cuing and visual judgments (Buchtel & Butter, 1988; McDonald, Teder-Sälejärvi, & Hillyard, 2000; Spence & Driver, 1997). The crossmodal contextual cuing effect obtained in the present study differs from the cuing effect found in those studies in several major ways. First, the auditory–visual contextual cuing effect occurs as a result of learning, whereas the traditional auditory cuing effect on visual performance occurs without training, presumably suggesting built-in orienting mechanisms. Second, the time course of the cuing effect seems to differ across these types of cuing; whereas the auditory exogenous cuing effect lasts only a short time in general (e.g., Ward, McDonald, & Golestani, 1998), the present auditory–visual contextual cuing is obtained with much longer cue–target interval (2 sec; Experiment 1). Finally, location specificity also differs substantially, mainly because the spatial resolution of auditory cues in the previous studies was relatively low. In most studies, the auditory cue simply indicates the left or right side (or, at most, the center) of the visual field. In contrast, the present auditory contextual cue enhanced identification of visual targets distributed at 48 possible target locations. The result—that the contextual cuing effect was not obtained when the auditory–visual association was disconnected (Experiment 1) and when auditory cues were associated with random target locations (Experiment 2)—suggests that the auditory cue was finely associated with the target locations.

The present results extend the ubiquitous statistical learning account of contextual cuing. On the basis of the finding that the cuing effect occurs if statistical consistency exists in a single trial or across trials, Ono et al. (2005) argued that the visual system is sensitive to all kinds of statistical consistency and will make use of predictive information. The present finding is consistent with this idea and suggests that the account is not limited to the visual domain but applies to associations between auditory–visual domains, because if the ubiquitous sta-

tistical learning account were limited to the events only within the visual domain, auditory–visual contexts should not have improved the search performance. However, the present finding also suggests that the ability of the cognitive system to learn contextual cues is not boundless; for example, the learning effect was not obtained when participants ignored consistent spatial contexts (e.g., Jiang & Chun, 2001; Kawahara, 2003). Ono et al. (2005) also found that when the spatial layouts of distractors (or target location) in trial $N-1$ predicted trial N 's target locations, but the target (or the distractor layout) in trial $N-1$ changed its location randomly, the contextual cuing effect was not observed. This suggests that the visual system is sensitive to statistical noise as well as to consistency. The presence of such boundary conditions clearly indicates that the rule of ubiquitous statistical learning in the visual context does not always apply. Therefore, it is reasonable to expect some limitations to the acquisition of auditory–visual crossmodal context. Similarly, the result of Experiment 3 revealed a constraint on the benefit of such auditory–visual context, in that—for the cuing to be effective—conscious recognition of the auditory stimuli is required. Such high recognition performance of auditory stimuli is consistent with the finding that participants can discriminate repetitively presented meaningless auditory stimuli (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997).

Comparison of the results of Experiments 1 and 2 provides a clue to the character of the temporal aspect of the contextual cuing effect. The procedures in these two experiments were identical, except that the auditory stimuli preceded the visual display by 2 sec in Experiment 1 and by 1 sec in Experiment 2. This difference did not affect the pattern of the results: The removal of consistency increased RTs in both experiments, suggesting that the critical period for the contextual cuing to be effective is between 1 sec before the onset of the visual display and the actual onset. Although we have insufficient evidence to determine the time course of the contextual cuing effect, a recent study on the visual contextual cuing effect showed that the effect develops at least about 300 msec following the onset of the contextual information (Ogawa, Takeda, & Kumada, 2007). Ogawa et al. modified an ordinary spatial contextual cuing procedure so that, in some trials, the participants had to detect a probe dot that appeared either on a target or on a distractor location. The results indicated that the RTs to detect a probe dot in repeated layouts were faster at search target locations and slower at distractor locations than they were in nonrepeated layouts. Because it is not clear whether the crossmodal contextual cuing effect develops in a similar manner as the unimodal cuing effect, further investigation is necessary to specify a detailed profile of the time course of the crossmodal cuing effect.

The present results indicate that contextual cuing with auditory–visual consistency seems to be slightly different from that with visual consistency, in that explicit recall of the auditory stimuli is required to gain effective cuing, although the associations between the auditory stimuli and the visual target locations remain implicit. One possibil-

ity is that auditory–visual contextual cuing is less easily acquired than cuing within the visual domain. In the everyday world, many possible combinations of auditory–visual associations exist, some valid and some invalid; for example, one would not want to combine the sound of a plane with the sight of a bird flying. Thus, the cognitive system may require more perceptual evidence (i.e., explicit recall of sounds) before it links auditory events with a visual locus. Another possibility is that more auditory evidence was required, because the present auditory stimuli were presented with visual distractors that varied their locations randomly for every trial. On the basis of the finding that the visual system is sensitive to statistical noise as well as to consistency (Ono et al., 2005), the presence of such visual noise may reduce learning efficiency of auditory–visual association, resulting in the need for more auditory evidence. Since many other possibilities exist, further examination is needed to clarify whether explicit recall of auditory stimuli is required for any types of crossmodal contextual cuing involving the auditory domain.

In summary, the present study showed that participants were able to learn an association between auditory stimuli and immediately successive visual target locations. This finding extends the ubiquitous statistical learning account of the contextual cuing (Ono et al., 2005): The cognitive system is sensitive to auditory–visual consistency. It is important to note that the associations between the auditory stimuli and the visual target locations were not explicitly accessible, although the auditory stimuli themselves were explicitly retrievable. This pattern of the results shares the same characteristics of contextual cuing as have been obtained in the visual domain: The cuing effect occurs without conscious awareness or explicit recognition of the critical association between the target location and other consistencies in the environment, such as distractor locations. Chun and his colleagues (Chun & Jiang, 1998, 2003; Chun & Phelps, 1999) have provided ample evidence to suggest that the spatial contextual cuing effect is driven by implicit memory representations. The present study extended this notion to the auditory–visual association with the constraint that the explicit recognition of the auditory stimuli was necessary when the present procedure was used. It remains a matter for future research to specify why no cuing effect was obtained when participants were unable to recognize the auditory stimuli. One possibility is that the number of trials in the training phase in Experiment 3 was not sufficient for the cuing effect to be effective. Thus, extensive practice might be required for a successful contextual cuing effect from less distinctive stimuli, as used in Experiment 3.

AUTHOR NOTE

I thank Kaori Takahashi and Tomohiro Tsukamoto for data collection, Mario Kleiner for technical assistance, and Melina Kunar and two anonymous reviewers for their helpful comments. Correspondence concerning this article should be addressed to J. Kawahara, National Institute of Advanced Industrial Science and Technology, Central 6, 1-1-1 Higashi, Tsukuba 305-8566, Japan (e-mail: jun.kawahara@aist.go.jp).

REFERENCES

- BRAINARD, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436.
- BROCKMOLE, J. R., & HENDERSON, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, **13**, 99-108.
- BUCHTEL, H. A., & BUTTER, C. M. (1988). Spatial attentional shifts: Implications for the role of polysensory mechanisms. *Neuropsychologia*, **26**, 499-509.
- CHUA, K.-P., & CHUN, M. M. (2003). Implicit scene learning is view-point dependent. *Perception & Psychophysics*, **65**, 72-80.
- CHUN, M. M. (2000). Contextual cuing of visual attention. *Trends in Cognitive Sciences*, **4**, 170-178.
- CHUN, M. M., & JIANG, Y. (1998). Contextual cuing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, **36**, 28-71.
- CHUN, M. M., & JIANG, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, **10**, 360-365.
- CHUN, M. M., & JIANG, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 224-234.
- CHUN, M. M., & PHELPS, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, **2**, 844-847.
- EGETH, H. E., & YANTIS, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, **48**, 269-297.
- ENDO, N., & TAKEDA, Y. (2004). Selective learning of spatial configuration and object identity in visual search. *Perception & Psychophysics*, **66**, 293-302.
- EYSENCK, M. W. (1979). Depth, distinctiveness, and elaboration. In L. Cermak & F. I. M. Craik (Eds.), *Levels of processing: An approach to memory*. Hillsdale, NJ: Erlbaum.
- FISER, J., & ASLIN, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, **12**, 499-504.
- 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.
- HOWARD, J. H., HOWARD, D. V., DENNIS, N. A., YANKOVICH, H., & VAIDYA, C. J. (2004). Implicit spatial contextual learning in healthy aging. *Neuropsychology*, **18**, 124-134.
- JIANG, Y., & CHUN, M. M. (2001). Asymmetric object substitution masking. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 895-918.
- JIANG, Y., SONG, J.-H., & RIGAS, A. (2005). High-capacity spatial contextual memory. *Psychonomic Bulletin & Review*, **12**, 524-529.
- KAWAHARA, J. (2003). Contextual cuing in 3-D layouts defined by binocular disparity. *Visual Cognition*, **10**, 837-852.
- KIM, M. S., & CAVE, K. R. (1999). Top-down and bottom-up attentional control: On the nature of the interference from a salient distractor. *Perception & Psychophysics*, **61**, 1009-1023.
- KUNAR, M. A., FLUSBERG, S., HOROWITZ, T. S., & WOLFE, J. M. (in press). Does contextual cuing guide the deployment of attention? *Journal of Experimental Psychology: Human Perception & Performance*.
- MACKEBEN, M., & NAKAYAMA, K. (1993). Express attentional shifts. *Vision Research*, **33**, 85-90.
- MCDONALD, J. J., TEDER-SÄLEJÄRVI, W. A., & HILLYARD, S. (2000). Involuntary orienting to sound improves visual perception. *Nature*, **407**, 906-908.
- NABETA, T., ONO, F., & KAWAHARA, J. (2003). Transfer of spatial context from visual to haptic search. *Perception*, **32**, 1352-1358.
- OGAWA, H., TAKEDA, Y., & KUMADA, T. (2007). Probing attentional modulation of contextual cuing. *Visual Cognition*, **15**, 276-289.
- OLSON, I. R., & CHUN, M. M. (2001). Temporal contextual cuing of visual attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1299-1313.
- ONO, F., JIANG, Y., & KAWAHARA, J. (2005). Inter-trial contextual cuing: Association across successive visual search trials guides spatial attention. *Journal of Experimental Psychology: Human Perception & Performance*, **31**, 703-712.

- PELLI, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.
- SAFFRAN, J. R., NEWPORT, E. L., ASLIN, R. N., TUNICK, R. A., & BARRUECO, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, **8**, 101-105.
- SCHACTER, D. L., & BUCKNER, R. L. (1998). On the relations among priming, conscious recollection, and intentional retrieval: Evidence from neuroimaging research. *Neurobiology of Learning & Memory*, **70**, 284-303.
- SPENCE, C., & DRIVER, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, **59**, 1-22.
- WARD, L. M., McDONALD, J. J., & GOLESTANI, N. (1998). Cross-modal control of attention shifts. In R. D. Wright (Ed.), *Visual attention* (pp. 232-268). New York: Oxford University Press.
- WOLFE, J. M. (1997). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13-73). Hove, U.K.: Psychology Press.
- WOLTERS, G., & VERDUIN, C. J. (1982). Retention performance as a function of the distinctiveness of memory traces and retention task. *Psychological Research*, **44**, 257-267.

NOTES

1. I am happy to share the auditory stimuli used here upon request.
2. The mean error rates under each condition are plotted in Figure 2 as a function of block. A similar ANOVA as that conducted for the RT data revealed no significant main effects of blocks and consistency. No interaction was observed between these two factors. All participants under the inconsistent transfer condition noticed that the auditory stimuli were repeatedly presented. The mean recognition accuracy of the auditory stimuli was 81.4%, which was significantly higher than chance level [$t(15) = 9.8, p < .001$]. The hit and false alarm rates were 83.2% and 10.5%, respectively. Under the consistent transfer condition, all participants reported that they noticed the repetition of the auditory stimuli. The mean recognition accuracy of the auditory stimuli was 78.1%, which was significantly higher than chance level [$t(15) = 10.1, p < .001$]. The

hit and false alarm rates were 75.8% and 10.4%, respectively. The recognition accuracy of the auditory stimuli under the inconsistent transfer group did not differ statistically from those in the consistent transfer condition [$t(30) = 0.8, n.s.$]. None of the participants reported noticing the consistency between the auditory stimuli and the target locations. For the inconsistent transfer group, the mean correct recognition rate of an exact target location and mean quadrant match were 4.8% and 25.0%, respectively. Neither were significantly higher than chance [$t(15) = 1.5, n.s.$, and $t(15) = 0.01, n.s.$, respectively]. For the consistent transfer group, the mean of exact location report was 2.8% and the mean of the quadrant choice was 24.4%, which was not significantly higher than chance [$t(15) = 0.8, n.s.$, and $t(15) = 0.17, n.s.$, respectively]. No significant difference was detected in the mean exact location report between the inconsistent and consistent transfer groups, Welch's t test [$t(23.4) = 1.0, n.s.$]. Similarly, no significant difference was found in the mean quadrant choice rates between the two groups [$t(30) = 0.1, n.s.$].

3. Note that a recent study (Kunar, Flusberg, Horowitz, & Wolfe, in press) questions the conventional view that the facilitated visual search performance in contextual cuing experiments is due to guidance of attention to the target location. They suggest that the main facilitation in the visual contextual cuing effect is not due to facilitation in attentional guidance. In their study, they showed that the search efficiency, as indexed by the reaction time \times set size function, did not improve over accumulation of learning trials; moreover, they found a contextual cuing effect even when no need seemed to exist for attentional guidance (i.e., set size = 1). On the basis of these results, they concluded that other factors such as response selection contributed to the contextual cuing effect. In this respect, a possibility exists that response selection process in the present study was influenced by the repetitive presentation of auditory-visual consistency in the present study. However, some other studies have shown improvement in search efficiency (e.g., Chun & Jiang, 1998). Therefore, a decisive conclusion regarding this issue awaits further investigation.

(Manuscript received December 18, 2006;
revision accepted for publication May 25, 2007.)