Object-based auditory facilitation of visual search for pictures and words with frequent and rare targets

Lucica Iordanescu[1], **Marcia Grabowecky**[1,2], and **Satoru Suzuki**[1,2]

[1]Department of Psychology, Northwestern University, Evanston, Illinois

[2]Interdepartmental Neuroscience Program, Northwestern University, Evanston, Illinois

Abstract

Auditory and visual processes demonstrably enhance each other based on spatial and temporal coincidence. Our recent results on visual search have shown that auditory signals also enhance visual salience of specific objects based on multimodal experience. For example, we tend to see an object (e.g., a cat) and simultaneously hear its characteristic sound (e.g., "meow"), to name an object when we see it, and to vocalize a word when we read it, but we do not tend to see a word (e.g., cat) and simultaneously hear the characteristic sound (e.g., "meow") of the named object. If auditory-visual enhancements occur based on this pattern of experiential associations, playing a characteristic sound (e.g., "meow") should facilitate visual search for the corresponding object (e.g., an image of a cat), hearing a name should facilitate visual search for both the corresponding object and corresponding word, but playing a characteristic sound should not facilitate visual search for the name of the corresponding object. Our present and prior results together confirmed these experiential-association predictions. We also recently showed that the underlying object-based auditory-visual interactions occur rapidly (within 220 ms) and guide initial saccades towards target objects. If object-based auditory-visual enhancements are automatic and persistent, an interesting application would be to use characteristic sounds to facilitate visual search when targets are rare, such as during baggage screening. Our participants searched for a gun among other objects when a gun was presented on only 10% of the trials. The search time was speeded when a gun sound was played on every trial (primarily on gun-absent trials); importantly, playing gun sounds facilitated both gun-present and gun-absent responses, suggesting that object-based auditory-visual enhancements persistently increase the detectability of guns rather than simply biasing gun-present responses. Thus, object-based auditory-visual interactions that derive from experiential associations rapidly and persistently increase visual salience of corresponding objects.

Introduction

Outside of the controlled psychology laboratory, perceptual experience is fundamentally multisensory. In particular, visual and auditory signals are often closely related. For example, when a dog barks, the visual image of the dog and its barking sound are coincident in space, dynamically synchronized (the barking sound is modulated in synchrony with the mouth/head movement of the dog), and both carry object-identity information. These space-

based, synchrony-based, and object-based auditory-visual associations facilitate visual search. For example, visual detection is facilitated when a sound is simultaneously presented at the location of a visual target (e.g., Bolognini, Frassinetti, Serino, & Làdavas, 2005; Driver & Spence, 1998; Frassinetti, Bolognini, & Làdavas, 2002; Stein, Meredith, Huneycutt, & McDade, 1989) and when a sound is synchronized with the dynamics of a visual target (e.g., Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008).

Our recent research has focused on the effects of object-based auditory-visual associations. We have reported that presenting spatially uninformative characteristic sounds speeds localization of target objects in visual search (Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008). On each trial, the participant was asked to look for a specific object; the target was always presented, and the participant's task was to indicate the visual quadrant in which the target was located. The target, for example a cat, was localized faster when its characteristic sound, "meow," was simultaneously presented compared to when a sound of an object not present in the search display (an unrelated sound), a sound of a distractor object in the display (a distractor-consistent sound), or no sound was presented. Importantly, although characteristic sounds speeded visual search for pictures of objects, they did not affect visual search for names of objects. For example, the search for a word "cat" among other words was not speeded when a "meow" sound was simultaneously presented. This result demonstrated that the facilitative effect of characteristic sounds on object-picture search was mediated neither by semantic associations nor by the sounds serving as a reminder of the target object (as the target was different from trial to trial). If characteristic sounds facilitated visual search through semantic and/or reminder effects, they should have facilitated object-name search as well as object-picture search. We thus concluded that characteristic sounds facilitate visual target localization through object-based auditory-visual interactions.

These interactions are likely to be mediated by high-level brain areas that respond especially strongly when a visual object is presented together with its characteristic sound (e.g., Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005; Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Beauchamp, Lee, Argall, & Martin, 2004). Auditory-visual interactions at the object-processing level could facilitate spatial localization of the target object via cross-connections between the ventral (object processing) and dorsal (spatial and action-related processing) cortical visual pathways (e.g., Felleman & Van Essen, 1991) and/ or via feedback connections from high-level polysensory and visual object-processing areas to lower-level retinotopic visual areas (e.g., Rockland & Van Hoesen, 1994; Roland, Hanazawa, Undeman, Eriksson, Tompa, Nakamura, Valentiniene, & Ahmed, 2006).

Interestingly, whereas target-consistent sounds speeded target localization, distractor-consistent sounds did not slow target localization compared to unrelated sounds or to no sounds. This suggests that object-based auditory effects on visual search require goal-directed top-down signals; whereas a characteristic sound of a search target facilitates localization of the target, a characteristic sound of a distractor object inconsistent with the top-down goal-directed signal does not divert attention to the distractor. Object-based auditory-visual interactions may generally selectively facilitate visual perception of task-relevant stimuli (e.g., Molholm, Ritter, Javitt, & Foxe, 2004).

Our recent results using saccades as the mode of response have provided further support for the idea that characteristic sounds facilitate the process of visual search rather than post-search processes such as response confirmation. Participants fixated the target object as quickly as possible. Characteristic sounds significantly increased the proportion of fast target fixations (< 220 ms), and also narrowed the tuning of the initial saccade towards the target (Iordanescu, Grabowecky, Franconeri, Theeuwes, & Suzuki, 2010).

Our converging evidence thus suggests that characteristic sounds facilitate direction of attention to target objects in visual search via object-based auditory-visual interactions. This paper presents research concerning two remaining questions. The first question is why characteristic sounds did not influence object-name search. We hypothesize that effects of characteristic sounds on visual search are driven by experiential associations between seeing visual objects and hearing their characteristic sounds. According to this *experiential association* hypothesis, a characteristic sound (e.g., "meow") should facilitate visual search for the corresponding object (e.g., a cat's picture) because people tend to see a cat and hear it meow at the same time. In contrast, a characteristic sound (e.g., "meow") should not facilitate visual search for the name of the corresponding object (e.g., a word "cat") because people do not tend to read a word "cat" and hear "meow" at the same time. Our previous study confirmed these predictions (see above). The experiential association hypothesis makes two additional predictions. Hearing a spoken name (e.g., "cat") should facilitate visual search for both the corresponding object (e.g., a cat picture presented among pictures of other objects) and the corresponding written name (e.g., a word "cat" presented among other words). This is because, for example, people tend to say "cat" and point at a cat, and read the word "cat" and overtly or covertly pronounce it. To systematically test all four experiential-association based predictions, we investigated the effects of two types of auditory stimuli, characteristic object sounds and spoken object names, on two types of visual search stimuli, object pictures (Experiment 1) and object names (Experiment 2).

The second question is whether the facilitative effects of characteristic sounds on visual search that we demonstrated using a target localization task with a different target object presented on each trial, generalize to a task where participants responded as to whether a target was present or absent (a typical search task in the literature). We were especially interested in determining whether characteristic sounds facilitate visual search when the target is always of a specific category (e.g., a gun) and rare (i.e., when most of the trials are target-absent trials). Searching repeatedly for rare targets is important in practical applications of visual search such as radiological screening and airport baggage screening. It is difficult to maintain a high degree of vigilance when looking for rare targets (e.g., Wolfe, Horowitz, & Kenner, 2005; Van Wert, Horowitz, & Wolfe, 2009; Warm, Parasuraman, & Matthews, 2008). It is also difficult to improve vigilance performance. Visual search studies using realistic x-ray luggage images have shown that vigilance is difficult to improve with behavioral strategies such as allowing participants to correct a previous response or requiring them to confirm responses (e.g., Van Wert at al., 2009; Wolfe et al., 2007). Object-based auditory-visual interactions could provide a means to enhance vigilance for detection of rare but important objects if characteristic sounds (e.g., gun-shot sounds) could facilitate visual search repeatedly for a single target category (e.g., guns) even when targets are rare and the sounds are presented on every trial (i.e., mostly on target-absent trials). This possibility was investigated in Experiment 3.

Experiment 1: Effects of characteristic object sounds and spoken object names on picture search

In this experiment, we determined whether localization of a target object (e.g., a picture of a cat) in visual search was facilitated by concurrent presentation of a characteristic sound (e.g., "meow") and/or spoken name (e.g., "cat") of the target object.

Methods

Participants—Eighteen undergraduate students at Northwestern University gave informed consent to participate for partial course credit. They all had normal or corrected-to-normal

visual acuity, were fluent English speakers, and were tested individually in a normally lit room.

Stimuli—Each search display contained four colored pictures of common objects (scaled to fit within a 6.5° by 6.5° area of visual angle) placed in the four quadrants at 4.7° eccentricity (center to center); see Figure 1A for an example. One of these objects was the target, and the remaining objects were distractors. These objects were selected from a set of 20 objects (bike, bird, car, cat, clock, coins, dog, door, [running] faucet, keys, kiss, lighter, mosquito, phone, piano, stapler, thunder, toilet, train, and wine glass; see Iordanescu et al., 2008 for the full set of images). The durations of characteristic sounds varied due to differences in their natural durations ($M = 862$ ms with $SD = 451$ ms, all sounds < 1500 ms). These heterogeneities, however, should not have affected our measurement of auditory-visual interactions because our design was fully counterbalanced (see below). The spoken names of these objects were recorded by a female speaker. All sounds were clearly audible (~70 dB SPL), presented via two loudspeakers, one on each side of the display monitor; the sounds carried no information about the target's location.

On each trial, a characteristic sound of the target object — *target-consistent sound* (e.g., a "meow" sound for a cat target), a characteristic sound of a distractor object — *distractor-consistent sound* (e.g., a barking sound when a dog was a distractor), the spoken name of the target object — *target name* (e.g., spoken "cat" for a cat target), the spoken name of a distractor object) — *distractor name* (e.g., spoken "dog" when a dog was a distractor), or a beep (440 Hz, 100 ms), was presented simultaneously with the search display.

In the distractor-consistent-sound and distractor-name conditions, the relevant distractor object was always presented in the quadrant diagonally opposite from the target across the central fixation marker so that any potential cross-modal enhancement of the distractor did not direct attention toward the target. Within a block of 100 trials, each of the 20 characteristic sounds was presented once as the target-consistent sound and once as the distractor-consistent sound, and each of the 20 spoken names was presented once as the target name and once as the distractor name. Each picture served as the target equally frequently in conjunction with a target-consistent sound, a distractor-consistent sound, a target name, a distractor name, and a beep. This counterbalancing ensured that any facilitative effect of target-consistent sounds and target names would be attributable to the sounds' and names' associations with the visual targets, rather than to the properties of the pictures, characteristic sounds, and/or spoken names themselves.

We avoided inclusion of objects with similar characteristic sounds (e.g., keys and coins) within the same search display. Aside from these constraints, the objects were randomly selected and placed on each trial. Each participant was tested in four blocks of 100 trials; 10 practice trials were given prior to the experimental trials. The visual stimuli were displayed on a color CRT monitor (1024 by 768 pixels) at 75 Hz, and the experiment was controlled by a Macintosh PowerPC 8600 using Vision Shell software (Micro ML, Inc.). A chinrest was used to stabilize the viewing distance at 61 cm.

Procedure—Participants pressed the space bar to begin each trial. The name of the target (e.g., cat) was aurally presented at the beginning of each trial. After 1070 ms, the search display appeared for 670 ms synchronously with the onset of one of the two types of characteristic sounds (target-consistent or distractor-consistent), one of the two types of spoken names (target name or distractor name), or a beep. Participants were instructed to indicate the location of the target as quickly and accurately as possible by pressing one of the four buttons (arranged in a square array) that corresponded to the quadrant in which the target was presented. Participants used the middle and index fingers of the left hand to

respond to the upper left and lower left quadrants, and used the middle and index fingers of the right hand to respond to the upper right and lower right quadrants (responses were thus ideomotor compatible). Participants were instructed to maintain eye fixation at a central circle (0.46° diameter). Response times (RTs) and errors were recorded.

Results

The target-consistent sounds speeded localization of the corresponding pictures compared to the distractor-consistent sounds ($t[17] = 2.887$, $p < 0.011$) and to the beep (a trend, $t[17] = 2.004$, $p < 0.062$), whereas the distractor-consistent sounds did not slow target localization compared to the beep ($t[17] = 1.143$, $n.s.$) (Figure 1C, solid bars). This replicated our prior result (Iordanescu et al., 2008). Spoken names produced similar results. The spoken target names speeded localization of the named pictures compared to the spoken distractor names ($t[17] = 3.575$, $p < 0.003$) and to the beep ($t[17] = 2.618$, $p < 0.020$). The spoken distractor names somewhat slowed target localization compared to the beep (a trend, $t[17] = 1.851$, $p < 0.082$) (Figure 1C, striped bars). To confirm that the auditory-visual compatibility effect was equivalent for the characteristic sounds and spoken names, sound type (characteristic sounds vs. spoken names) did not interact with auditory-visual compatibility (target-consistent vs. distractor-consistent) in a 2-factor $ANOVA$ without the beep condition, $F(1,17) = 1.944$, $n.s.$ The overall error rate was low (less than 3%); there were no significant effects on error rates or any evidence of a speed-accuracy trade-off (Figure 1C).

Thus, both characteristic sounds and spoken names of target objects facilitated localization of the corresponding pictures. Although distractors' characteristic sounds did not significantly affect picture search (compared to the beep), spoken distractor names produced a trend toward slowing picture search.

Experiment 2: Effects of characteristic object sounds and spoken object names on name search

We determined whether localization of a target name (e.g., "cat") in visual search was facilitated by the presentation of a characteristic sound (e.g., "meow") of the named object and/or by the spoken name (e.g., spoken "cat").

Methods

Eighteen new participants were recruited for this experiment. The stimuli and procedure were the same as in Experiment 1, except that the pictures of objects used in Experiment 1 were replaced by the written names of those objects; the horizontal extent of the words was similar to those of the pictures (Figure 1B).

Results

Neither the target-consistent nor distractor-consistent sounds had an effect on the localization of target names. The response times were no faster on the target-consistent-sound trials compared to the distractor-consistent-sound trials ($t[17] = 0.685$, $n.s.$) or to the beep trials ($t[17] = 0.752$, $n.s.$), and no slower on the distractor-consistent-sound trials compared to the beep trials ($t[17] = 0.105$, $n.s.$) (Figure 1D). Although the responses were slightly faster on the target-consistent-sound trials compared to the distractor-consistent-sound trials (Figure 1D, black bars), this was driven by one participant who produced an unusually large compatibility gain (greater than the three standard deviations from the mean). Without this outlying participant, the mean response time on the target-consistent-sound trials was identical to that on the distractor-consistent-sound trials (both 526 ms). In contrast, the spoken target names speeded visual localization of the corresponding words compared to the spoken distractor names ($t[17] = 6.024$, $p < 0.0001$) and to the beep (a

trend, $t[17] = 1.978$, $p < 0.065$), whereas the spoken distractor names did not slow target localization compared to the beep ($t[17] = 1.451$, *n.s.*) (Figure 1D). To confirm that auditory-visual compatibility effect was present for spoken names but absent for characteristic sounds, sound type (characteristic sounds vs. spoken names) interacted with auditory-visual compatibility (target-consistent vs. distractor-consistent) in a 2-factor *ANOVA* without the beep condition, $F(1,16) = 9.511$, $p < 0.008$ (without the above mentioned outlier; $F[1,17] = 3.252$, $p < 0.09$ including the outlier). The overall error rate was low (less than 3%); there were no significant effects on error rates or any evidence of a speed-accuracy trade-off (Figure 1D).

Thus, characteristic sounds of target and distractor objects had no effect on visual localization of target names. In contrast, spoken names of target objects facilitated visual localization of the corresponding names. However, spoken names of distractor objects did not slow localization of target names.

In summary, we have demonstrated in Experiments 1 and 2 that spoken names of target objects facilitate localization of both the corresponding objects (in picture search) and names (in name search), whereas characteristic sounds of objects facilitate only localization of the corresponding objects (in picture search) with no effects on localization of object names (in word search).

Experiment 3: Effects of characteristic object sounds on picture search when targets are rare

Do characteristic sounds facilitate search even for rare targets? In this experiment, participants repeatedly looked for the same category of rare targets. For example, participants looked for a gun on every trial, with the gun presented on only 10% of the trials. If a gun-shot sound automatically increased visual salience of a gun image, presenting a gun-shot sound should facilitate gun detection on the rare target-present trials even with the gun-shot sound presented on every trial, including during the target-absent trials which constituted 90% of all trials.

To determine object-based auditory-visual effects on vigilance over and above the effects of specific sounds per se (e.g., arousing, calming), we employed two types of sounds, gun-shot sounds and cat sounds. In the *relevant-sound* condition, participants searched for a cat while hearing a meow sound on every trial, or searched for a gun while hearing a gun-shot sound on every trial. In the *irrelevant-sound* condition, participants searched for a cat while hearing a gun-shot sound on every trial, or searched for a gun while hearing a meow sound on every trial. The overall effects of target type (cat or gun picture) and sound type (meow or gun-shot sound) would be attributable to differential visual salience of the cat and gun pictures and differential arousal levels induced by the meow and gun-shot sounds. Importantly, the remaining effect of sound-picture compatibility (i.e., relevant-sound condition vs. irrelevant-sound condition) would demonstrate object-specific facilitative effects of characteristic sounds on visual search with rare targets.

Methods

Participants—Thirty-six undergraduate students at Northwestern University gave informed consent to participate for partial course credit. They all had normal or corrected-to-normal visual acuity, and were tested individually in a normally lit room.

Stimuli—The number of pictures per search display was increased to eight (compared to four in Experiments 1 and 2) to simulate a more realistic situation for a typical vigilance task

(e.g., baggage screening). The centers of the eight pictures were evenly placed along an approximate iso-acuity ellipse (21° horizontal by 16° vertical, the aspect ratio based on Rovamo & Virsu, 1979); see Figure 2 for an example. The target (a gun or a cat) was presented on only 10% of the trials. The distractors were randomly selected from a set of 50 commonly encountered and portable objects: apples, a basketball, a belt, a white book, a brown book, a brush-scissors pair, a camera, a CD case, a portable MP3 player, a cell phone, a can of cleanser, a soda can, a dollar bill, a brown hat, a black hat, a pair of sun glasses, a pair of gloves, a beer bottle, a pack of chewing gum, a hair clipper, headphones, a lipstick, a long-sleeve shirt, markers, a mug, a necklace, a newspaper, a perfume bottle, a watch, a pen, a tennis racket, a sandal, a scarf, a shampoo bottle, running shoes, shorts, socks, a spray bottle, a brown sweater, a striped sweater, a wallet, tea bags, toilet paper, a tooth brush, a tube of tooth paste, towels, keys, a lighter, a stapler, and a bird. Each target was randomly selected from 15 examples of cats or 15 examples of guns. Each picture was confined within a 4.33° by 4.35° rectangular region.

In the meow-sound block of trials, each meow sound was randomly selected from 15 different meow sounds. Similarly, in the gun-shot-sound block of trials, each gun-shot sound was randomly selected from 15 different gun-shot sounds. All sounds were clearly audible (~70 dB SPL), were presented via loud speakers placed on each side of the display monitor, and carried no target-location information (as in Experiments 1 and 2).

Half of the participants always searched for cats and the remaining participants always searched for guns. Each participant was tested in three blocks of 150 trials with 10 practice trials given at the beginning of each block. In the *relevant-sound* block, participants heard sounds characteristic of the target category on every trial. In the *irrelevant-sound* block, participants heard sounds characteristic of the other category on every trial. In the *no-sound* block, participants heard no sounds. The block order was counterbalanced across participants using all six permutations.

The stimuli were displayed on a color CRT monitor (1024 × 768) at 75Hz, and the experiment was controlled with a Macintosh PowerPC 8600 using Vision Shell software (micro ML, Inc.). A chin rest was used to stabilize the viewing distance at 80 cm.

Procedure—At the beginning of the experiment, participants were informed of the target category to search for (cat or gun). The experimenter pressed the space bar to start each block of trials. After 2008 ms, the search display appeared synchronously with a sound of the search category (in the relevant-sound block), a sound of the other category (in the irrelevant-sound block), or no sound (in the no-sound block). The display remained for 2008 ms or until participants responded. Participants indicated whether a target was present or absent by pressing a corresponding response button as quickly and accurately as possible. The next search display appeared 1500 ms following the response.

Results

We performed *ANOVA*s on the response time data from the target-present and target-absent trials, with target type (cat or gun) and block sequence (6 counterbalancing sequences of the three blocks) as the between-participant factors, and sound relevance (relevant sound, irrelevant sound, or no sound) and block order (1st, 2nd, or 3rd) as the within-participant factors. For both target-present and target-absent trials, there was a significant order effect in that response times speeded in later blocks, probably due to practice effects ($F[2, 48] = 29.017$, $p < 0.0001$, for target-present trials, and $F[2,48] = 79.317$, $p < 0.0001$, for target-absent trials). There was also a significant effect of target type in that participants responded to the cat targets faster than they responded to the gun targets ($F[1, 24] = 23.973$, $p <$

0.0001, for target-present trials, and F[1, 24] = 15.812, p < 0.0006, for target absent trials), suggesting that our cat pictures were generally more distinctive than our gun pictures.

Importantly, sound relevance produced significant effects on both target-present and target-absent trials (F[2, 48] = 11.386, p < 0.0001, for target-present trials and F[2, 48] = 6.174, p < 0.005, for target-absent trials). For target-present trials, the target-relevant sounds significantly speeded visual detection of rare targets compared to the target-irrelevant sounds (t[35] = 2.166, p < 0.037), demonstrating an object-specific auditory-visual facilitation of visual search for rare targets (Figure 3A, black bars). The fact that target-absent trials were no slower (if anything they were faster) with the target-relevant sounds than with the target-irrelevant sounds (t[35] = 0.710, $n.s.$) (Figure 3B, black bars) indicates that the facilitative effect of the target-relevant sounds on target-present trials is not due to response bias. If the target-relevant sounds simply biased participants to make a target-present response, the target-relevant sounds should have speeded target-present responses but slowed target-absent responses.

Furthermore, presenting a sound per se (whether target relevant or irrelevant) had a large impact. Compared to no sounds, both the target-relevant and target-irrelevant sounds speeded responses on both target-present trials (Figure 3A) and target-absent trials (Figure 3B) (for responses on target-present trials, t[35] = 4.973, p < 0.0001, for the target-relevant sounds vs. no sounds, and t[35] = 2.630, p < 0.012, for the target-irrelevant sounds vs. no sounds; for responses on target-absent trials, t[35] = 3.731, p < 0.0007, for the target-relevant sounds vs. no sounds, and t[35] = 2.734, p < 0.009, for the target-irrelevant sounds vs. no sounds). Thus, any coincident sound (irrespective of target relevance) facilitated visual search for rare targets likely by increasing arousal and/or providing temporal cue for the onset of a search array.

The overall error rates were low (3.3% misses and 0.2% false positives for the cat search, and 5.2% misses and 1.1% false positives for the gun search), and there were no significant effects involving errors.

Discussion

Visual experience in the real world is often accompanied by closely associated auditory experience. Our prior research suggests that cross-modal interactions develop through consistent and repeated multisensory experience (Smith, Grabowecky, & Suzuki, 2007, 2008). Here we investigated the possibility that auditory and visual processing of objects and their names are associated based on their frequent co-occurrence in the real world, and that these experience-based auditory-visual associations can be used to facilitate visual search. We tested this hypothesis by comparing the effects of characteristic object sounds and spoken object names on searches for pictures and names. People tend to see an object (e.g., a key chain) and concurrently hear its characteristic sound (e.g., a jingling sound), hear an object name and simultaneously look at the corresponding object (e.g., point at a dog and say "Look at the dog!"), or see an object name and overtly or covertly pronounce it. In contrast, people usually do not read an object name and simultaneously hear a characteristic sound of the named object. We thus predicted that a characteristic object sound should facilitate search for a picture of the corresponding object (Iordanescu et al., 2008, 2010), a spoken object name should facilitate search for both a picture and name of the corresponding object, but a characteristic object sound should not affect search for a name of the corresponding object (Iordanescu et al., 2008). We confirmed these predictions in Experiments 1 and 2. Our results thus suggest that auditory-visual interactions from experiential associations facilitate visual search for common objects and their names.

As in our prior studies (Iordanescu et al., 2008, 2010), target-consistent sounds facilitated target localization, but distractor-consistent sounds did not significantly slow search. However, whereas distractor-consistent sounds showed little evidence of slowing target localization compared to other objects' sounds (sounds of objects not in the current search display) or to no sounds in our prior studies (Iordanescu et al., 2008, 2010), the distractor-consistent sounds in the current study modestly slowed search (although not significantly) compared to the beep sound (Figures 1C and D). It is possible that the beep sound was not an optimum control because it could have speeded search by increasing arousal and/or by providing an especially effective temporal cue for the onset of a search array due to its sharp auditory onset; had we used no sound as the control, the search times with the distractor-consistent sounds may have been virtually equivalent to those with no sound (as in Iordanescu et al., 2008, 2010). Nevertheless, we replicated our prior results in that distractor-consistent sounds did not significantly slow search even compared to the beep sound. This is consistent with the idea that top-down goal-directed signals (e.g., Reynolds & Chelazzi, 2004) might play an important role in mediating object-based auditory-visual interactions (e.g., Molholm et al., 2004; also see Iordanescu et al., 2008 and 2010 for discussion of neural mechanisms that potentially mediate the goal-directed nature of the object-based auditory-visual facilitation in visual search). It is interesting to note that spoken names of distractors trended toward slowing picture search (compared to the beep sound; Figure 1C). It might be the case that spoken names have privileged influences on the salience of visual objects because when people call out an object (e.g., "Snake!"), it is generally beneficial to direct attention to the corresponding object.

If auditory-visual interactions generally develop through experiential associations, characteristic sounds of materials (e.g., glass, wood, plastic, etc.) may also direct attention to the corresponding materials irrespective of object information. In fact, some of the stimuli we used, for example, a wine glass and a clinking sound, a key chain and a jingling sound, and a door and a squeaking hinge sound, could be mediated by material-based rather than object-based auditory-visual associations. We are currently investigating how material-consistent sounds generated in various ways (e.g., by tapping on materials or breaking materials) facilitates visual search for specific materials, such as localizing a metal-textured patch presented among distractor patches showing other material textures.

In our prior studies (Iordanescu et al., 2008, 2010) and also in Experiments 1 and 2, we presented a different search target on each trial and asked the participants to localize the target. In Experiment 3, we extended our results to the case where participants persistently looked for a single category of target objects while the target object was rarely presented and participants responded as to whether a target object was present or absent. This is an important extension because a challenging case of visual search involves vigilance, repeatedly looking for a rare target (e.g., a gun) as in airport baggage screening. The results from Experiment 3 have provided some useful insights.

First, presenting either a meow or gun-shot sound simultaneously with a visual search display substantially speeded responses on both target-present and target-absent trials, compared to presenting no sound. It is possible that presenting any sound, even beeps, might increase arousal and/or provide a temporal cue to speed visual search with rare targets. Alternatively, it might be the case that meaningful and affectively-charged sounds such as a meow or gun-shot sound are particularly effective. Second, in addition to the general benefit of presenting a sound with the search display (compared to presenting no sound), there was an advantage of presenting characteristic sounds of the target category. The target-relevant sounds facilitated search compared to target-irrelevant sounds. The fact that the target-relevant sounds speeded target-present responses without slowing target-absent responses

suggested that the target-relevant sounds increased the salience of target objects rather than biasing target-present responses.

It is remarkable that the target-relevant sounds speeded search despite the fact that they were completely uninformative as to the presence of a target (in fact they were mis-informative because targets were absent on most trials), and participants should have therefore ignored them. Our participants performed 450 search trials in about hal an hour. It would be interesting to extend the duration of the experiment. If target-relevant sounds persistently facilitated visual search for rare targets for a period of hours, the technique might provide a means to improve performance in baggage screening and other situations that require persistent search for rare targets. For example, repeatedly presenting appropriate cracking sounds might facilitate detection of cracks during inspections of buildings or machines.

In summary, visual search in the real world occurs in a multisensory environment. Visual objects are often experienced along with their characteristic sounds and spoken names. Consequently, both characteristic sounds and spoken names of objects facilitated localization of objects in visual search. Written names of objects are often experienced along with their spoken versions, but are not experienced along with the named objects' characteristic sounds. Consequently, spoken names but not characteristic sounds of objects facilitated visual localization of object names. Our results thus suggest that coincident experience of object-related visual and auditory signals lead to object-specific auditory-visual associations through which auditory signals can facilitate visual search. This object-based auditory-visual facilitation is persistent in that characteristic sounds speeded visual search even when targets were rare and the sounds of a single target category were presented on every trial (i.e., primarily on target-absent trials). These results are consistent with our recent results demonstrating that correlated multisensory experience leads to facilitative cross-modal sensory interactions (Smith et al., 2007, 2008).

Acknowledgments

References

Amedi A, von Kriegstein K, van Atteveldt MN, Beauchamp MS, Naumer MJ. Functional imaging of human crossmodal identification and object recognition. Experimental Brain Research. 2005; 166:559–571.

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nature Neuroscience. 2004; 7(11): 1190–1192.

Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron. 2004; 41:809–823. [PubMed: 15003179]

Bolognini N, Frassinetti F, Serino A, Làdavas E. "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. Experimental Brain Research. 2005; 160:273–282.

Driver J, Spence C. Attention and the crossmodal construction of space. Trends in Cognitive Sciences. 1998; 2(7):254–262. [PubMed: 21244924]

Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex. 1991; 1:1–47. [PubMed: 1822724]

Frassinetti F, Bolognini N, Làdavas E. Enhancement of visual perception by crossmodal visuo-auditory interaction. Experimental Brain Research. 2002; 147:332–343.

Iordanescu L, Grabowecky M, Franconeri S, Theeuwes J, Suzuki S. Charactersitic sounds make you look at target objects more quickly. Attention, Perception, and Psychophysics. (In Press).

Iordanescu L, Guzman-Martinez E, Grabowecky M, Suzuki S. Characteristic sounds facilitate visual search. Psychonomic Bulletin & Review. 2008; 15(3):548–554. [PubMed: 18567253]

Molholm S, Ritter W, Javitt DC, Foxe JJ. Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cerebral Cortex. 2004; 14:452–465. [PubMed: 15028649]

Reynolds JH, Chelazzi L. Attentional modulation of visual processing. Annual Review of Neuroscience. 2004; 27:611–647.

Rockland KS, Van Hoesen GW. Direct temporal-occipital feedback connections to striate cortex (V1) in the macaque monkey. Cerebral Cortex. 1994; 4:300–313. [PubMed: 8075534]

Roland PE, Hanazawa A, Undeman C, Eriksson D, Tompa T, Nakamura H, Valentiniene S, Ahmed B. Cortical feedback depolarization waves: a mechanism of top-down influence on early visual areas. Proceedings of the National Academy of Sciences, U.S.A. 2006; 103:12586–12591.

Rovamo J, Virsu V. Visual resolution, contrast sensitivity, and the cortical magnification factor. Experimental Brain Research. 1979; 37:475–494.

Smith E, Grabowecky M, Suzuki S. Auditory-visual crossmodal integration in perception of face gender. Current Biology. 2007; 17:1680–1685. [PubMed: 17825561]

Smith EL, Grabowecky M, Suzuki S. Self-awareness affects vision. Current Biology. 2008; 18(10):R414–R415. [PubMed: 18492469]

Stein BE, Meredith ME, Huneycutt WS, McDade LW. Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. Journal of Cognitive Neuroscience. 1989; 1:12–24.

Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J. Pip and Pop: nonspatial auditory signals improve spatial visual search. Journal of Experimental Psychology: Human Perception and Performance. 2008; 34(5):1053–1065. [PubMed: 18823194]

Van Wert MJ, Horowitz TS, Wolfe JM. Even in correctable search, some types of rare targets are frequently missed. Attention, Perception, & Psychophysics. 2009; 71(3):541–553.

Warm JS, Parasuraman R, Matthews G. Vigilance requires hard work and is stressful. Human Factors. 2008; 50:433–441. [PubMed: 18689050]

Wolfe JM, Horowitz TS, Kenner NM. Cognitive psychology: rare items often missed in visual searches. Nature. 2005; 435:439–440. [PubMed: 15917795]

Wolfe JM, Horowitz TS, Van Wert MJ, Kenner NM, Place SS, Kibbi N. Low target prevalence is a stubborn source of errors in visual search tasks. Journal of Experimental Psychology: General. 2007; 136:623–638. [PubMed: 17999575]
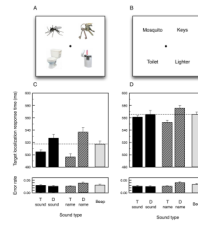
Figure 1.
(A & C) Picture search (Experiment 1). (B & D) Name search (Experiment 2). An example of a picture search display (A) and a name search display (B). The effects of target-consistent sounds (T sound), distractor-consistent sounds (D sound), spoken target names (T name), spoken distractor names (D name), and a beep, on target localization response times (upper panel) and error rates (lower panel) for picture search (C) and name search (D). The error bars represent ± 1 *SEM* adjusted for within-subjects design.

Figure 2.
An example of target-present trials for (A) the gun search task and (B) the cat search task (Experiment 3).
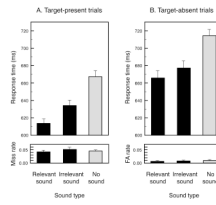
Figure 3.
The effects of sounds consistent with the target category (Relevant sound), sounds consistent with another object category (Irrelevant sound), and no sounds, on picture search (Experiment 3). (A) Response times (upper panel) and miss rates (lower panel) for target-present trials. (B) Response times (upper panel) and false alarm rates (lower panel) for target-absent trials. The error bars represent ±1 *SEM* adjusted for within-subjects design.