Change perception in complex auditory scenes

Francesco Pavani and Massimo Turatto
University of Trento, Rovereto, Italy

In four experiments, we examined the role of auditory transients and auditory short-term memory in perceiving changes in a complex auditory scene comprising multiple auditory objects. Participants were presented pairs of complex auditory scenes that were composed of a maximum of four animal calls delivered in free field; participants were instructed to decide whether the two scenes were the same or different (Experiments 1, 2, and 4). Changes to the second scene consisted of either the addition or the deletion of one animal call. Contrary to intuitive predictions based on results from the visual change blindness literature, substantial deafness to the change emerged without regard to whether the scenes were separated by 500 msec of masking white noise or by 500 msec of silence (Experiment 1). In fact, change deafness was not even modulated by having the two scenes presented contiguously (i.e., 0-msec interval) or separated by 500 msec of silence (Experiments 2 and 4). This result suggests that change-related auditory transients played little or no role in change detection in complex auditory scenes. Instead, the main determinant of auditory change perception (and auditory change deafness) appears to have been the capacity of auditory short-term memory (Experiments 3 and 4). Taken together, these findings indicate that the intuitive parallels between visual and auditory change perception should be reconsidered.

Change perception in the visual modality—and particularly its counterpart, change blindness—have received much interest in the past decade (Rensink, 2002; Simons & Levin, 1997). In particular, change blindness studies have shown that, contrary to our impression of seeing everything around us, human observers often fail to see even large changes in the scene if attention is not directed to the object or location of change (O'Regan, Rensink, & Clark, 1999; Rensink, O'Regan, & Clark, 1997). Change blindness does not directly address the conscious perception of objects per se, but addresses only the awareness of change (Rensink, 2000); yet it supports the view that attention might be crucial for the contents of our conscious visual perception, or, put differently, that we consciously see what we attend to (Mack & Rock, 1998). In the present study, we examined an auditory analogue of change blindness—namely, *change deafness*—in the context of complex auditory scenes.

In the visual modality, the mechanisms allowing correct change perception and those leading to change blindness are relatively well understood (O'Regan, Deubel, Clark, & Rensink, 2000). Any sudden change in the environment that can be registered by the visual system gives rise to a *local* visual transient. When such a local transient is detected, attention is automatically attracted to the change location, and the novel aspect of the visual scene is reported (Rensink et al., 1997; Turatto & Bridgeman, 2005). However, when the local visual transient is masked or made ineffective in capturing visual attention, the change location remains unattended, and change blindness occurs. Although different visual events can mask the change-related transient (Rensink, 2002), this has typi-

cally been achieved by interposing a blank screen between two consecutive images of the same scene differing in only one particular (Rensink et al., 1997). The appearance of the second scene after the blank creates a *global* visual transient (i.e., all locations in the second scene change with respect to the previous blank image), so that detection of the *local* visual transient that is associated with the changed object is no longer available to summon attention. Under these conditions, change blindness is often observed, and the change can be detected only through a time-consuming serial scanning of the scenes. According to this account, selective attention is the mechanism that permits information from the two scenes to be stored in visual short-term memory (vSTM; see Luck & Vogel, 1997; Phillips, 1974) for comparisons across views that would then allow conscious change detection. Note that this account does not exclude the possibility that change detection could, to some extent, occur in the absence of awareness (Fernandez-Duque & Thornton, 2003; Laloyaux, Destrebecqz, & Cleeremans, 2006; Thornton & Fernandez-Duque, 2000).

The fact that people can easily fail to notice visual changes is a remarkable result. Yet, human beings are often confronted with multisensory perceptions, rather than just unimodal visual ones, which raises the important question of whether an analogue of change blindness exists in other sensory modalities. Recently, an analogue of change blindness has been documented in the tactile domain, with participants failing to detect location changes between consecutive tactile patterns presented across the whole body surface (Gallace, Tan, & Spence, 2006, 2007; see

F. Pavani, francesco.pavani@unitn.it

also Gallace, Auvray, Tan, & Spence, 2006, for a visuo-tactile version of the paradigm). The existence of change *deafness* has also been described: A listener's ability to detect a change in the auditory scene declines systematically as the complexity of the scene increases—specifically, when attention cannot be directed to the changing feature or object in advance (Eramudugolla, Irvine, McAnally, Martin, & Mattingley, 2005; see also Vitevitch, 2003).

The first study to introduce the expression *change deafness* was conducted by Vitevitch (2003). Participants were instructed to repeat single words, presented through headphones by a male voice, as quickly and accurately as possible. One half of the participants heard the same talker throughout the entire word list. The remaining participants unexpectedly heard a different male talker pronouncing the words in the second part of the list. Regardless of whether the two parts of the list were separated by a 1-min break (Experiment 1) or were contiguous (Experiment 2), approximately 50% of the participants who were exposed to a different talker during the experiment failed to notice such a change. Vitevitch described this surprising inability to detect a voice change as being an example of *change deafness*. However, it should be noted that listeners failed to perceive auditory changes in a context in which they were not instructed to expect them. This is unlike the typical change blindness paradigm, in which participants are explicitly instructed to search for changes. Therefore, the phenomenon appears to be more similar to inattentional blindness (Mack & Rock, 1998) than to change blindness. In addition, it should be noted that the paradigm adopted by Vitevitch did not entail any multiple and concurrent auditory objects, but only single words presented in sequence. In this respect, it did not reproduce the complexity of the scene usually used in change blindness experiments.

A subsequent study (Eramudugolla et al., 2005) more closely paralleled the procedure and scene complexity that characterize change blindness. In that study, participants monitored two consecutive naturalistic auditory scenes, each lasting 5 sec and comprising multiple auditory objects, to detect whether one of the auditory objects present in the first scene disappeared from the second. The two scenes were always separated by a 500-msec interval of white noise, and participants' attention was manipulated as follows. In the *directed attention* condition, participants were cued to the potentially changing object before the scenes were presented (the name of the object was shown at the center of the screen). In contrast, in the *nondirected attention* condition, no cue was provided, and, therefore, participants were left with the strategy either of directing their attention randomly to different objects at any given moment, or of distributing their attention over all objects in the auditory scene. The results showed a strong modulation of auditory change perception by attention. Specifically, when attention was focused on the correct object in advance, listeners had no difficulty detecting the object's disappearance, and no change deafness occurred, regardless of the number of objects (four, six, or eight) that were presented in the scene. However, when attention was not

cued to the potentially changing object, profound change deafness emerged; listeners' sensitivity (d') to auditory changes dropped from a value of almost 3 to 1 as the number of objects increased from four to eight (recall that a d' of zero indicates chance-level performance).

Instead of cycling the two scenes continuously until listeners found the change—as is the case in the standard Flicker technique adopted for many change blindness studies (Rensink et al., 1997)—the authors used the "one-shot technique" (Gallace et al., 2007; Phillips, 1974; Rensink, 2000). This technique consists of the presentation of a single pair of events (here auditory scenes) in each experimental trial, and participants are required to detect any difference between the two. It is interesting to note that Eramudugolla et al. (2005) made the explicit assumption that a 500-msec burst of noise between the two auditory scenes was critical to eliciting the phenomenon of change deafness in their study. As they clearly stated, the noise burst was added between the scenes "to mask any [auditory] transient or echoic memory trace that might cue the listener's attention to the change" (Eradumugolla et al., 2005, p. 1108).

The assumption that the role of auditory transients for change perception in auditory scenes is analogous to the role of visual transients for change perception in visual scenes is not straightforward. Acoustic analysis of complex auditory scenes entails two processes (Alain & Izenberg, 2003; Bregman, 1990). One concerns the segregation of acoustic data within hundreds of milliseconds in order to extract acoustic elements from concurrent sound sources. The other entails sequential integration of acoustic data over several seconds in order to extract meaningful *auditory objects*, defined as auditory experience that "produces a two-dimensional image with frequency and time dimensions" (Griffiths & Warren, 2004, p. 891; see also Kubovy & Van Valkenburg, 2001). According to this perspective, any sudden change in the pattern of auditory stimulation that can be detected by the auditory system (i.e., *auditory transients*, analogous to the concept of visual transients) could have more impact on the real-time aspect of auditory scene analysis (i.e., concurrent sound segregation) than on the sequential sound integration that develops through time. In fact, it can also be hypothesized that auditory transients become ineffective at orienting attention to the auditory change after they are incorporated into a sequence that is perceived as being a whole auditory object or event.

In the present study, we explored whether change deafness relies on the same mechanisms that produce the inability to perceive changes in the visual modality. Specifically, we examined the role of auditory transients and auditory short-term memory (aSTM) in perceiving changes in a complex auditory scene comprising multiple auditory objects. Rather surprisingly, we found that, contrary to what has been repeatedly demonstrated for vision (e.g., Rensink et al., 1997; Turatto & Bridgeman, 2005), and more recently for touch (Gallace, Tan, & Spence, 2006), auditory transients play a less crucial role in the change deafness phenomenon when complex auditory scenes are considered.

EXPERIMENT 1

The aim of Experiment 1 was twofold. First, we wanted to examine whether the masking effect of white noise is really necessary in order to induce the change deafness phenomenon. To this end, we compared listeners' performance during the auditory change-detection task (in the one-shot version, similar to Eramudugolla et al., 2005) when the 500-msec interval between the two scenes was either empty or filled with white noise. Second, since previous studies of change blindness in vision have shown that change detection and identification may be better for deletion than for addition (Agostinelli, Sherman, Fazio, & Hearst, 1986; Mondy & Coltheart, 2000), we also examined whether the same asymmetry holds true for change deafness. We intermingled change trials in which an object that was present in the first auditory scene was removed in the second one (deletions) with change trials in which an object that was absent in the first auditory scene was added to the second one (additions).

Unlike Eramudugolla et al. (2005), who presented auditory objects in a virtual auditory space created by means of head-related transfer functions, we used real auditory scenes, in which auditory objects were actually presented from different free-field locations in front of the listener (see the Method section and Figure 1 for details). In addition, in the present study, participants were subjected only to the nondirected attention condition, because the Eradumogolla et al. study clearly showed that this was the only condition in which listeners exhibited a clear change deafness effect, whereas participants performed at ceiling in the directed attention condition. Any advantage of the silent interval over the white-noise interval should emerge as a reduced change deafness effect in the former condition.

Method

Participants. Fourteen undergraduate students (10 females and 4 males; mean age = 23.6 years; SE = 1.0) at the University of Trento participated in the study. All were unaware of the purpose of the experiment and self-reported having normal or corrected-to-normal vision and normal hearing. The study was conducted in accordance with the guidelines of the Declaration of Helsinki and after approval by the Ethical Committee at the Department of Cognitive Sciences and Education of the University of Trento.

Stimuli and Apparatus. A library of 12 animal calls (cat, horse, sparrow, crow, hawk, rooster, dog, walrus, pig, cow, sheep, chicken) was created for the experiment. All auditory stimuli in the library were duration matched (all lasted 2 sec). The sound pressure level of each stimulus ranged between 60 and 65 dB, as measured from the location of the participant's head using a digital phonometer (TES 1350A). In addition, although their onset was not synchronized, they all started within 200 msec of one another. Note that this small temporal asynchrony never resulted in phenomenological experiences of apparent motion in the auditory display, presumably due to the substantial difference between the presented sounds.

Each auditory scene was created by drawing a subset of auditory stimuli (either three or four) from the library and simultaneously presenting each of them through separate loudspeakers. Selection of stimuli from the library was pseudorandomized, with the only constraint being that the same auditory stimulus could not occupy the same location in space (i.e., the same loudspeaker) in successive trials. All trials comprised two auditory scenes lasting 2 sec each, presented one after the other and separated by a 500-msec interval. For half of the trials, the interval was empty (i.e., 500 msec of silence), and for the remaining half it was filled with 500 msec of white noise. On no-change trials, all four stimuli selected for the first auditory scene were repeated in the second scene. On item-deletion
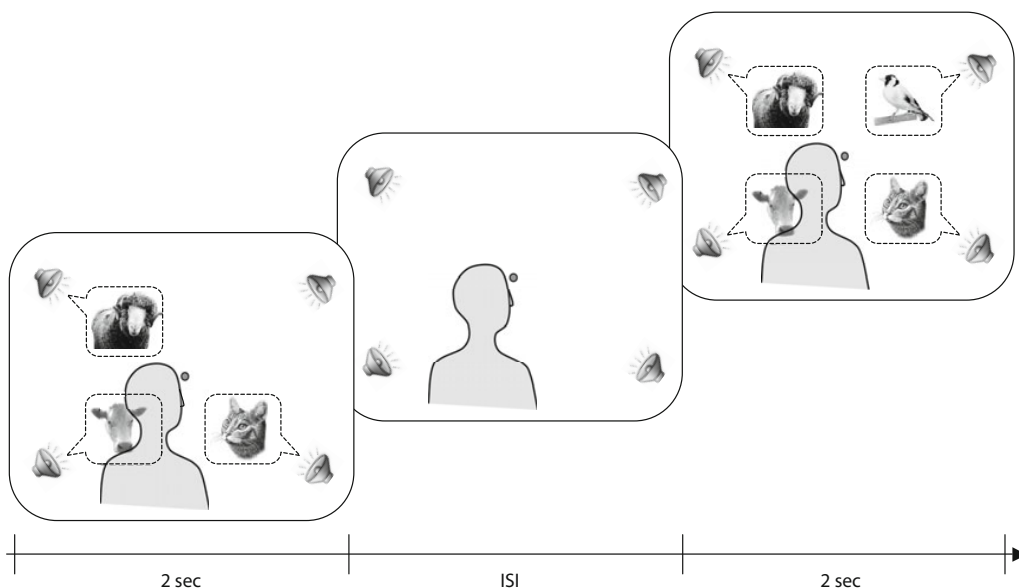

Figure 1. Schematic illustration of the experimental setup and a change-trial sequence. In the example, the bird's call is added from the top right speaker in the second auditory scene. Note that, across experiments, the interstimulus interval (ISI) was filled either with 500 msec of white noise (Experiment 1) or with 500 msec of silence (Experiments 1, 2, and 4), or was reduced to 0 msec (i.e., the two auditory scenes were presented contiguously; Experiments 2 and 4).

trials, only three out of the four stimuli presented in the first auditory scene were repeated in the second scene. On item-addition trials, the first auditory scene consisted of three stimuli, which were all repeated in the second scene with the addition of a new, fourth item. When repeated across the two auditory scenes, stimuli never changed their location—that is, they were always presented from the same loudspeaker within each trial.

Because item deletion or addition would have produced overall volume differences between the two auditory scenes, random volume attenuation was applied to either or both of the scenes to prevent any between-scenes comparison based exclusively on intensity changes. Specifically, each scene was independently presented either at full volume (68 dB on average when the scene comprised three items, 71 dB on average when the scene comprised four items) or attenuated by 3–5 dB. Note that this random volume attenuation easily covered the maximal difference in loudness between the three- and four-item scenes. All of the participants were explicitly informed about the unpredictability of these volume variations.

The experiment was conducted inside a dimly lit, sound-attenuated anechoic booth (Amplifon G2x2.5; floor area = 200 × 250 cm, height = 220 cm). One ultrabright green light-emitting diode (LED) was attached to the wall in front of the participant and served as a visual fixation. Six amplified loudspeakers (Genius SP-Q06S; 120 W, frequency response = 75 Hz–20 kHz, input impedance = 21 kΩ) were used in the study. Four loudspeakers were mounted on the widest wall of the booth, at the four corners of an imaginary square of 120 cm, centered vertically and horizontally with respect to the floor, ceiling, and lateral walls (see Figure 1). These were used during the experimental phase to present auditory scenes (see Procedure). The remaining two loudspeakers (not shown in Figure 1) were also mounted on the wall, 40 cm on either side of the visual fixation, and were used only during the instruction phase for presenting examples of each of the animal calls (see Procedure). The entire apparatus was hidden from view by means of a large polyester curtain—an acoustically transparent close-weave cloth, width = 250 cm, height = 200 cm, specifically designed for mounting on loudspeaker grilles; model: KS50E, Maplin, U.K.) suspended from the ceiling. The visual fixation LED, which was illuminated throughout the experiment, was the sole piece of equipment visible through the cloth.

The entire apparatus was controlled by a personal computer (Dell GX270). Loudspeakers were interfaced by means of three external soundcards (Creative Sound Blaster MP3+) connected to the USB ports, and LEDs were interfaced through the parallel port. A computer mouse (Logitec) served for response collection. Timing of the stimuli and response collection were controlled by custom software written in MATLAB 6.5 (MathWorks), using the Cogent library (www.fil.ion.ucl.ac.uk/cogent2000).

Procedure and Design. Participants sat on an adjustable chair inside the booth, facing the apparatus at a distance of approximately 180 cm. The height of the chair was adjusted to align each participant's ear level with the visual fixation (i.e., a vertical midpoint with respect to the loudspeakers); the participants were instructed to maintain their gaze toward the fixation.

Before the experiment, all 12 stimuli in the library were stereophonically presented, one at a time, through the two loudspeakers flanking the visual fixation (note that these were not used during the experiment, and therefore are not shown in Figure 1). The participants were asked to name each of the animal calls, and the experimenter ensured that all stimuli were clearly identifiable before proceeding with the experiment. Participants were then informed that their task was to indicate whether the two auditory scenes presented in each experimental trial were the same or different by pressing the left or right button on a computer mouse, respectively. It was made explicit that the task was unspeeded and that the volume of the two auditory scenes would change randomly from presentation to presentation.

Participants performed 6 practice trials, followed by four experimental blocks of 48 trials each, resulting in 192 trials overall (i.e., 64 trials for each of the experimental conditions: no-change, item addition, item deletion). Thus, a change occurred on two thirds of the trials. When an item was added or deleted, the location of the change was counterbalanced across the experiment. The type of interval (white noise or silence) between the two auditory scenes was counterbalanced between blocks. Half of the participants started with white-noise interval trials, the other half with silent interval trials.

Results

Change detection performance expressed as a percentage of correct trials (i.e., overall mean of correct detections and correct rejections) was 66% (SE = 5%). Thus, substantial change deafness occurred in our task (50% is the chance level). However, in order to obtain a criterion-free measure of the participants' capacity to detect a change between the two auditory scenes, data in this and the following experiments were analyzed according to signal detection theory. Specifically, we used a measure of signal sensitivity (d'; Green & Swets, 1974). The proportion of trials in which participants correctly detected a change between the scenes (hits) and the proportion of trials in which they reported a change on no-change trials (false alarms) were used to calculate sensitivity (d') and criterion (c) measures for each experimental condition.

Mean sensitivity (d') across all conditions was significantly greater than zero [M = 1.0, SE = 0.3; $t(13)$ = 3.5, p < .004], revealing better than chance performance overall. Mean d' values for each participant were also entered into an ANOVA with two within-participants factors: interval type (silence or white noise) and type of change (addition or deletion). The analysis revealed no main effect or interaction (all Fs < 1.6), indicating that sensitivity did not vary as a function of type of change (deletion, M = 0.9, SE = 0.2; addition, M = 1.1, SE = 0.3) or interval type (white noise, M = 1.0, SE = 0.3; silence, M = 1.1, SE = 0.3).

A further ANOVA, similar to the one used for sensitivity, was applied to the measure of criterion (c). No main effect or interaction was significant in this analysis (all Fs < 1.6), indicating that criterion did not vary as a function of type of change (deletion, M = 0.5, SE = 0.1; addition, M = 0.4, SE = 0.1) or of interval type (white noise, M = 0.4, SE = 0.1; silence, M = 0.4, SE = 0.1). Nonetheless, criterion was, on average, significantly greater than zero [M = 0.4, SE = 0.1; $t(13)$ = 3.9, p < .002], revealing that participants were more prone to no-change responses overall.

Discussion

Three main results emerged from the analysis of the results of Experiment 1. First, a rather modest sensitivity to auditory changes was observed, despite our seemingly simple situation, in which no more than four auditory objects were presented at any one time (note that Eramudugolla et al., 2005, observed a comparable sensitivity only when eight objects were presented). Second, change detection was as low when the change pertained to the addition of a new auditory object as it was when it pertained to the deletion of an existing auditory object. Third, and most important, substantial deafness to auditory changes emerged regardless of whether the interval between the auditory scenes was silent or filled with white noise.

This last finding challenges the assumption put forward by Eramudugolla et al. (2005, p. 1108), that masking "any transient or echoic memory trace that might cue the listener's attention to the change" is a prerequisite for change deafness. In fact, adding backward white noise masking at the end of the first scene and forward white noise masking at the beginning of the second scene was not critical to producing change blindness in Experiment 1. The 500-msec silent interval alone proved sufficient to prevent any attention shift to the auditory transients that signaled the change location. Although apparently counterintuitive, this result is actually compatible with what has been found in the visual literature on change blindness. As discussed earlier, the mere presence of a blank screen between the two visual scenes is sufficient to create change blindness because it makes the appearance of the second scene a global transient capable of hiding the local transient that is associated with the change. Extrapolating this mechanism to the auditory context, it can be argued that change deafness also occurred in the silent interval condition because the onset of the second auditory scene constituted a global transient with respect to the preceding silent interval. This, in turn, prevented attention shifts to the auditory transient that accompanied the change and resulted in change deafness.

In the visual modality, when the blank between the two scenes is removed, the local transient immediately attracts attention, and the change becomes visible (see, e.g., Rensink et al., 1997; Turatto, Bettella, Umiltà, & Bridgeman, 2003). In the same vein, one might suppose that simply removing the interval from between the two auditory scenes in our task should result in a dramatic reduction of change deafness. More specifically, the sudden disappearance of an animal call (deletion trials) or the sudden appearance of a new call (addition trials) should generate an auditory transient—that is, a sudden change in the auditory pattern of stimulation that can be registered by the auditory system—and attract attention. To examine this issue directly, in Experiment 2, we compared change detection performance when the auditory scenes were separated by a silent interval of 500 msec or when they were contiguously presented (i.e., the silent interval was reduced to 0 msec). In the latter condition, any auditory transient associated with the addition or deletion of an object should be unmasked, leading to virtually no change deafness.

EXPERIMENT 2

Method

Participants. Twelve undergraduate students (8 females and 4 males; mean age = 23.3 years; SE = 1.1) at the University of Trento participated in the study. All were unaware of the purpose of the experiment and self-reported having normal or corrected-to-normal vision and normal hearing. One participant had taken part in Experiment 1.

Stimuli, Apparatus, Procedure, and Design. These were identical to those in Experiment 1, with the following exceptions. All trials comprised two auditory scenes that lasted 2 sec each and were presented one after the other. On half of the trials, the auditory scenes were contiguous (i.e., no interval was introduced), whereas on the remaining half, the two auditory scenes were separated by 500 msec of silence. Trials with and without the silent interval were now intermingled unpredictably in each experimental block.

Results

Two participants performed very poorly on the no-change trials (i.e., more than 80% of the responses were false alarms) and were excluded from further analysis. On average, the participants' performance (percentage correct) was 78% (SE = 2%). As for Experiment 1, all analyses were conducted on sensitivity and criterion measures.

Mean sensitivity (d') across all conditions was significantly above zero [M = 2.0, SE = 0.1; $t(9)$ = 16.6, p < .0001], indicating better than chance performance overall. Mean d' for each participant was entered into an ANOVA with two within-participants factors: interval (0 msec of silence or 500 msec of silence) and type of change (addition or deletion). The analysis revealed a significant main effect of type of change [$F(1,9)$ = 9.92, p < .01] caused by higher perceptual sensitivity for item addition (M = 2.2, SE = 0.2) than for item deletion (M = 1.8, SE = 0.1). However, neither the main effect of interval [$F(1,9)$ = 1.3, p = .3] nor the two-way interaction was significant [$F(1,9)$ = 0.2, p = .7]. This indicates that sensitivity to change was unaffected by whether the two scenes were contiguously presented (M = 1.9, SE = 0.2) or separated by 500 msec of silence (M = 2.1, SE = 0.2).

Mean criterion (c) across all conditions was also greater than zero [M = 0.3, SE = 0.1; $t(9)$ = 8.3, p < .0001], revealing a tendency toward no-change responses overall. An ANOVA similar to that performed for d', now applied to criterion measures, revealed a main effect of type of change [$F(1,9)$ = 10.1, p < .01] due to participants' being more prone to no-change responses on deletion (M = 0.8, SE = 0.1) than on addition (M = 0.6, SE = 0.1) trials. However, the main effect of interval [$F(1,9)$ = 3.4, p = .1] and the two-way interaction [$F(1,9)$ = 0.2, p = .7] were not significant.

Discussion

Consistent with the results of Experiment 1, we once again found that changes in auditory scenes of just four items can be difficult to perceive. In addition, we found that change detection was more difficult when an item was deleted than when it was added. However, the most notable result of Experiment 2 was that no difference in change detection performance emerged as a function of whether the auditory scenes were presented contiguously or separated by a silent interval. This finding reveals that the mechanisms involved during change perception in the auditory modality are different from those involved in change perception in the visual modality. As noted earlier, visual changes are usually detected immediately, if not identified, when the local transients accompanying the change are not hidden by means of a blank screen interposed between the two scenes (O'Regan et al., 2000; Turatto & Bridgeman, 2005). This is either because visual transients can be detected in a diffuse attentional state or because they attract spatial attention to the corresponding position (Breitmeyer & Ganz, 1976; Jonides, 1981), thus becoming immediately visible. In contrast, the results of the present experiment clearly show that auditory changes can remain largely unnoticed, even when the two auditory scenes are contiguous (0-msec interstimulus interval).

In the General Discussion, we will discuss why auditory transients, at least in the present paradigm, seem to play no role in auditory change detection.

EXPERIMENT 3

When combined with the previous observation—that change deafness is comparable regardless of whether 500 msec of white noise or of silence is added between the scenes (Experiment 1)—the results of Experiment 2 suggest that auditory transients or echoic memory traces do not contribute to change perception in the relatively complex auditory scenes presented in Experiments 1 and 2. What other mechanism can then be responsible for change perception? Our hypothesis is that change detection was achieved by means of an aSTM system (Cowan, 1984, 1995) and that change deafness reflects the limits of aSTM in encoding and storing multiple auditory objects for comparison with a subsequent scene (as was also suggested by Eramudugolla et al., 2005).

By analogy with the visual modality, when transients turn out to be ineffective at informing the observer about visual changes, the only mechanism that allows change detection between two scenes is vSTM (Rensink et al., 1997). In the flicker paradigm, the observer shifts his or her attention to a specific part of the first scene, stores information at that location, and waits for the second image in order to see whether something has changed. Thus, in the typical change blindness paradigm, the possibility of successful change detection depends critically on the limits of vSTM. If the scene is made up of only very few items, change detection should be relatively easy. On the other hand, if the scene consists of many objects (as is usually the case with real-world scenes), the information that needs to be stored at any given moment exceeds the limits of vSTM. As argued by Luck and Vogel (1997), the possibility of reporting a change depends on the capacity to store as much information as possible from the first image in the vSTM and to compare this information with the second scene. Incidentally, it is interesting to note that the paradigm used by Luck and Vogel (see also Phillips, 1974)—namely, the procedure of interposing a blank between two scenes to be compared—constitutes the core of the flicker paradigm used in the change blindness literature.

In the visual domain, Luck and Vogel (1997) have quantified the limited capacity of vSTM and suggested that only four or five objects can be stored efficiently. In the tactile domain, a recent study by Gallace and colleagues (2007) suggests that tactile displays of two or three items could already prove complex enough to produce a substantial tactile change blindness phenomenon. In Experiment 3, we directly tested the capacity of aSTM for our complex auditory scenes. On each trial, we presented single auditory scenes that were identical to the ones used in Experiments 1 and 2, each comprising three or four items, followed by a single animal call delivered from a pair of loudspeakers directly in front of the participant. Participants were instructed to decide whether the single animal call (i.e., the probe) was present among the calls of the scene (see also Luck & Vogel, 1997).

Method

Participants. Ten undergraduate students (5 females and 5 males; mean age = 25.0 years, $SE = 1.0$) at the University of Trento participated in the study. All were unaware of the purpose of the experiment and self-reported having normal or corrected-to-normal vision and normal hearing. None had taken part in any of the previous experiments.

Stimuli, Apparatus, Procedure, and Design. These were identical to those in Experiment 1, with the following exceptions. On each trial, the presentation of the first auditory scene was followed by 500 msec of silence and the subsequent presentation of a single animal call from the two speakers mounted on each side of the fixation (i.e., the same two speakers used during the stimulus identification phase that preceded all experimental sessions). On half of the trials, the animal call presented alone (probe stimulus) was selected among the animal calls of the preceding auditory scene (*probe-present trials*). On the remaining half of the trials, the animal call was selected among the animal calls that did not compose the preceding auditory scene (*probe-absent trials*). Participants were instructed to indicate whether the probe was present or absent by pressing one of two mouse buttons. Response mapping of the mouse keys was clearly marked on the buttons with the letters P (*present*) and A (*absent*). It was made explicit that the task was unspeeded.

Participants performed 6 practice trials, followed by three experimental blocks of 64 trials each, resulting in 192 trials overall (i.e., 48 trials for each of the experimental conditions: three elements probe present, three elements probe absent; four elements probe present, four elements probe absent). The positioning of the silent speaker in the three-element scenes was fully counterbalanced within participants.

Results

On average, participants correctly reported the presence of the probe in complex auditory scenes in 82% ($SE = 2$%) of the trials. As for the previous experiments, we computed sensitivity (d') and criterion (c) measures for each participant.

For both types of auditory scenes, d' was significantly above zero [three elements, $t(9) = 12.79, p < .0001$; four elements, $t(9) = 11.87, p < .0001$], indicating that, overall, participants could discriminate above chance whether the probe was present or absent in the previous auditory scene. It is important, however, that sensitivity was higher with auditory scenes of three elements ($M = 2.4, SE = 0.2$) than four with those of elements ($M = 1.8, SE = 0.2$) [$t(9) = 3.54, p < .01$], indicating easier detection of probes in simpler auditory scenes.

Criterion also changed between the two scenes. Overall, participants were more prone to saying that the probe was present than that it was absent [$M = 0.4, SE = 0.1$; $t(9) = 4.25, p < .002$]—particularly when the auditory scenes comprised four elements ($M = 0.50, SE = 0.1$) rather than three elements ($M = 0.30, SE = 0.1$) [$t(9) = 5.82, p < .001$].

Discussion

The present results show that the aSTM capacity is severely limited, and they are in line with those of Luck and Vogel (1997) for the visual modality and those of Gallace et al. (2007) for the tactile modality. Although participants

responded above chance overall, a clear drop in performance emerged when set size increased from three to four items. Although some studies have previously probed auditory memory for sounds occurring simultaneously at different spatial locations (e.g., Darwin, Turvey, & Crowder, 1972; Treisman & Rostron, 1972), to our knowledge, the present result provides the first example of an auditory analogue of the Luck and Vogel paradigm. Note that this finding also offers an explanation for the advantage in change detection we found for addition over deletion trials in Experiment 2 (with a similar numerical trend in Experiment 1, albeit not significant). Since the first auditory scene comprised three objects on all addition trials, and comprised four objects on all deletion trials, the better performance observed for addition than for deletion trials might merely reflect the difference in complexity of the respective first auditory scenes. This, in turn, could have affected the amount of information encoded in aSTM and, therefore, the ability to make comparisons with the subsequent scene. Although future research should address whether such an advantage for addition over deletion could remain evident even when complexity of the first auditory scene is held constant (e.g., always comprises four elements), the difference in the complexity of the first scene that characterized the experiments reported here represents the most parsimonious explanation of the performance difference observed between addition and deletion trials.

The present findings are consistent with our hypothesis invoking a role of aSTM to explain change deafness. However, one might still argue that the failure to detect a change between the two scenes reflects the inability to perceive or segregate the auditory objects in the first (and the second) scene, rather than the inability to encode and store these objects in aSTM. In other words, instead of being a problem of memory capacity, change deafness could result, for example, from the reciprocal masking of the multiple and simultaneous animal calls in the scene. We did not exclude this possibility on empirical grounds, and, in addition, it is obvious that the larger the number of calls, the worse the reciprocal masking between auditory streams. However, it is useful to recall that, in the Eramudugolla et al. (2005) study, participants' performance was at ceiling as soon as their attention was directed to the identity of the potential change before the scene was presented (e.g., search for the cello). Thus, it appears unlikely that the multiple auditory objects masked one another in the complex auditory scene. In addition, there is now considerable independent evidence from event-related potential (ERP) studies that the segregation of input to distinct sound streams can occur without attention's necessarily having to be focused on the sounds (e.g., Ritter, Sussman, & Molholm, 2000; Sussman, 2005). This makes it likely that auditory objects were segregated in our scenes comprising a maximum of four elements, each lasting 2 sec.

EXPERIMENT 4

The main findings that emerged from the previous three experiments can be summarized as follows. First, in a complex auditory scene consisting of multiple auditory objects of the same category (i.e., animal calls), change deafness occurred even for a small number of auditory objects. This finding is similar to what has been reported recently in the tactile modality (e.g., Gallace et al., 2007). Second, change deafness was higher for object disappearance than for object appearance, a result that can parsimoniously be related to the lesser complexity of the first auditory scene on addition than on deletion trials. Third, the results of Experiments 1 and 2 showed that change deafness was not the consequence of a white-noise interval interposed between the scenes, but emerged just as strongly when the interval was completely silent, or even when the interval was removed altogether (i.e., the scenes were contiguous). In other words, we observed change deafness even when the auditory transients accompanying the change were neither masked by noise nor made less effective by the introduction of a gap between the two scenes. Fourth, the results of Experiment 3 support the view that, in our task, change deafness resulted mainly from the severely limited capacity of aSTM.

If auditory changes are detected only when the relevant auditory information is stored in aSTM (and not by means of auditory transients' attracting attention), it can be predicted that, whenever a change is detected, participants would also be able to access enough information about the changed item to be able to identify it correctly. Turatto and Bridgeman (2005) recently showed that change detection and change identification may constitute two distinct aspects of change perception in the visual modality. They found that change detection was easier than was change identification (as originally predicted by O'Regan et al., 2000), since change detection can be performed on the sole basis of the perceived local transients, whereas change identification requires the involvement of short-term memory in order to retrieve, for instance, the identity of the object that has disappeared. The aim of our last experiment was therefore to explore change detection and identification also in the auditory modality. The idea was to ascertain whether listeners were good (once a change was detected) at identifying the change as either an addition or a deletion, and whether they could report the identity of the object that changed. If aSTM is the main mechanism leading to change perception, we expected our participants to reach a good performance level on change identification whenever they had previously detected the change.

Method

Participants. Twelve undergraduate students (9 females and 3 males; mean age = 24.4 years; *SE* = 1.0) at the University of Trento participated in the study. All were unaware of the purpose of the experiment and self-reported having normal or corrected-to-normal vision and normal hearing. Three had taken part in one of the previous experiments.

Stimuli, Apparatus, Procedure, and Design. These were identical to those in Experiment 2, with the sole exception that participants were now instructed to detect any change between the two auditory scenes using the mouse button as before, and then, if a change was perceived, to report verbally whether an item was added or deleted and to identify which specific item (i.e., which animal call) had changed.

Results

Change-detection task. On average, participants' performance in the change-detection task was 77% correct (SE = 2%). Mean sensitivity (d') across all conditions was significantly above zero [M = 1.8, SE = 0.2; $t(11)$ = 11.6, p < .0001], indicating better-than-chance detection performance overall. Mean sensitivity scores for each participant were entered into an ANOVA with two within-participants factors: interval (0 msec of silence or 500 msec of silence) and type of change (addition or deletion). The analysis revealed a significant main effect of type of change [$F(1,11)$ = 33.9, p < .0001] caused by higher perceptual sensitivity for item addition (M = 2.1, SE = 0.2) than for item deletion (M = 1.5, SE = 0.2) trials. The main effect of interval was not significant [$F(1,11)$ = 0.4, p = .5], whereas the interaction between interval and type of change approached significance [$F(1,11)$ = 4.33, p = .06]. This reflected a tendency for higher differences in perceptual sensitivity between item addition and deletion for the 0-msec interval (addition, M = 2.2, SE = 0.2; deletion, M = 0.4, SE = 0.2) than for the 500-msec silent interval (addition, M = 2.0, SE = 0.2; deletion, M = 1.5, SE = 0.2).

Mean criterion (c) across all conditions was not greater than zero [M = 0.2, SE = 0.1; $t(11)$ = 1.8, p = .1]. An ANOVA similar to that performed on the d', now applied to criterion measures, also revealed a main effect of type of change [$F(1,11)$ = 33.8, p < .0001], caused by participants' being more prone to no-change responses on deletion (M = 0.8, SE = 0.1) than on addition (M = 0.6, SE = 0.1) trials. Instead, the main effect of interval was not significant [$F(1,11)$ = 1.8, p = .2], and the two-way interaction approached significance [$F(1,11)$ = 4.3, p = .06].

Change-identification task. On average, when a change was detected, the participants were 90% correct (SE = 2%) in deciding whether it was an addition or deletion. In addition, they correctly reported the identity of the changed item on 81% of the trials (SE = 3%). However, the participants' performance differed as a function of whether the change was an addition or a deletion. Participants correctly identified addition changes on 98% of the trials (SE = 1%) and deletion changes on 83% of the trials (SE = 4%) [$t(11)$ = 3.60, p < .004]. Furthermore, participants correctly reported the object's identity on 94% of the trials (SE = 1%) when it was an addition, and only on 68% of the trials (SE = 4%) when it was a deletion [$t(11)$ = 7.67, p < .0001]. Instead, the presence or absence of the silent interval had no effect on participants' ability to recognize the type of change (0-msec interval, M = 90%, SE = 2%; 500-msec interval, M = 91%, SE = 2%) [$t(11)$ = 0.9, p = .4], or on their ability to correctly report the identity of the changed item (0-msec interval, M = 81%, SE = 3%; 500-msec interval, M = 81%, SE = 3%) [$t(11)$ = 0.5, p = .6].

Discussion

Three important results emerged from the analysis of the final experiment. First, the results confirmed our prediction that, whenever a change was detected, participants were also correct at identifying the nature of this change (over 90% correct in deciding whether the change was an addition or a deletion), and were relatively good at identifying the change (over 80% accuracy in deciding which specific animal call was removed or added). Second, this near-ceiling performance in the change identification task was not modulated by the presence or absence of a silent interval between the two scenes. In other words, it did not reflect in any way the potential auditory transient related to the change, which would have been available when the two scenes were presented contiguously, but not when they were separated by 500 msec. This replicates the results of Experiment 2. Third, performance in the change identification task was higher when the first auditory scene comprised three items than when it comprised four items (i.e., was better on addition than on deletion trials). This again endorses the idea that the more the first scene can be encoded in aSTM, the more change perception improves, in terms of both detection and identification.

GENERAL DISCUSSION

The last decade has seen an intense proliferation of studies dedicated to the phenomenon of change blindness (e.g., Rensink et al., 1997; Turatto et al., 2003; see Rensink, 2002, for a review). This general interest stems from the fact that change blindness demonstrates the crucial role of visual attention and visual memory in determining the contents of conscious visual perception (Mack & Rock, 1988; Rensink, 2002; see also Gallace, Auvray, et al., 2006; Gallace, Tan, & Spence, 2006, 2007, for examples in the tactile modality). Two recent studies have addressed whether an analogue of change blindness exists in the auditory domain and have shown that changes in auditory scenes can be missed with remarkable ease when not attended (Eramudugolla et al., 2005; Vitevitch, 2003), a phenomenon that the investigators termed *change deafness*.

Eramudugolla et al. (2005) explicitly linked the observed phenomenon to the 500-msec white-noise interval they had introduced between the two auditory scenes. This white-noise interval was meant to represent the auditory analogue of the intervening blank screen in the typical change blindness paradigm and had the purpose of masking all transients (or *echoic memory traces*, in the authors' words) related to the auditory change. The interpretation offered by Eramudugolla and colleagues suggested that change deafness emerged because participants could not shift their attention to the relevant auditory change, precisely because the change-related auditory transients were masked. This endorsed the assumption that visual and auditory transients can play a very similar role for change perception in complex natural scenes, with easier change perception when transients are available and with change blindness or deafness when they are removed.

The results of the present series of experiments are summarized in Figure 2. Contrary to the conclusion of Eramudugolla et al. (2005), we found that change deafness in complex auditory scenes emerged without regard to auditory transient masking. Specifically, detection of changes
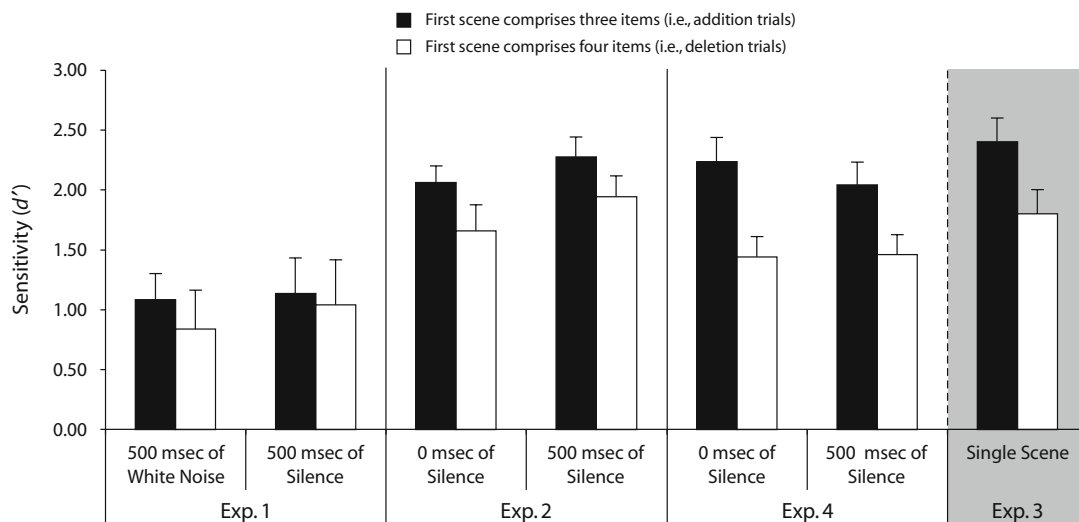
Figure 2. Sensitivity to auditory changes in Experiments 1 through 4, as a function of the number of elements in the first auditory scene. Note that, whereas Experiments 1, 2, and 4 always consisted in the presentation of two auditory scenes (with change trials equally divided among addition and deletion trials; see text for details), in Experiment 3 (highlighted in gray in the plot), the first auditory scene was followed by a single animal call (i.e., the probe stimulus). Error bars indicate the standard error of the mean.

between two consecutive auditory scenes was similarly poor, regardless of whether the scenes were separated by 500 msec of silence or by 500 msec of white noise (Experiment 1). In fact, auditory change perception was not even improved by having the two scenes presented contiguously (i.e., 0-msec interval) rather than having them separated by 500 msec (Experiments 2 and 4). In other words, unmasking the change-related transients between the two auditory scenes was not enough to prevent change deafness. This result indicates that, when processing complex auditory scenes, the auditory system may rely on transients to a much lesser extent than does the visual system, and that the intuitive parallel between visual and auditory change perception should be reconsidered.

This is not to say, however, that auditory transients are not important in audition. A large body of research using an electrophysiological indicator termed *mismatch negativity* (MMN) has revealed that the auditory systems can readily detect changes with respect to some repetitive aspect of the auditory environment (see Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001, for a review). In addition, studies on auditory and cross-modal spatial attention have clearly shown that a sudden auditory event, such as a burst of white noise in a silent background, triggers a shift of attention (auditory and visual) toward the corresponding location (see Spence, McDonald, & Driver, 2004, for a review). In some cases, such orienting appears even to be mandatory (Mazza, Turatto, Rossi, & Umiltà, 2007; but see Santangelo, Olivelli Belardinelli, & Spence, 2007), and several authors have put forward the idea that the primary function of audition is to detect sound sources and direct visual attention for further analysis (e.g., Heffner & Heffner, 1992; Kubovy & Van Valkenburg, 2001; Spence & Driver, 1997). However, it

should be emphasized that in most studies addressing the role of an abrupt sound for the orienting of attention, the auditory stimulus was a transient event in an otherwise silent (or very simplified) auditory scene (e.g., Mazza et al., 2007; Spence & Driver, 1994). Under these conditions, the attention-capturing auditory stimulus competes only minimally with other events in the auditory scene, and, therefore, it is conceivable that it can have privileged access to attentional resources and awareness.

A very different scenario emerges when complex auditory scenes containing multiple auditory objects are considered. Each auditory object in the scene is an acoustic experience that develops along the frequency and time dimensions (Griffiths & Warren, 2004; Kubovy & Van Valkenburg, 2001), with segregation of auditory input into distinct auditory streams occurring preattentively (e.g., Macken, Tremblay, Houghton, Nicholls, & Jones, 2003; Ritter et al., 2000; Winkler et al., 2003). In this respect, auditory objects are no longer series of auditory transients. As elegantly demonstrated by Sussman and colleagues (e.g., Sussman & Gumenyuk, 2005; Sussman, Ritter, & Vaughan, 1998), a frequency-deviant tone embedded in a repetitive tone sequence elicits an MMN when the sequence is perceived to be a series of distinct tones (because it is played at a slow pace; e.g., stimulus onset asyncnchrony [SOA] = 1,300 msec), but not when the sequence is perceived to be a single auditory event (because it is played at a faster pace; e.g., SOA = 100 msec). In other words, auditory transients may no longer be detected when they are incorporated into a sequence perceived as a whole auditory object or event. When applying the same logic to our own findings, one possibility is that auditory transients related to the change were no longer available to attract attention because they remained embedded in

the auditory objects resulting from preattentive auditory stream segregation.

An alternative explanation for why change-related auditory transients were inefficient at attracting attention (thus producing change deafness) should also be considered. Change-related auditory transients occurring at the onset of the second scene could have competed with the multiple amplitude and frequency transients that were related to the nonchanging events that occurred at the same instant in time. In other words, at the moment of change, most (if not all) auditory stimuli in the scene changed with respect to the previous time instant. Adopting the terminology of the literature on change blindness, this would be the auditory analogue of the *mudsplash* phenomenon (Gallace, Auvray, et al., 2006; Gallace et al., 2007; O'Regan et al., 1999), in which the change-related (task-relevant) transient is swamped by multiple and competing "extraneous" transients. An interesting difference to note is that, whereas in the visual and in the tactile change blindness paradigms, the extraneous transients (i.e., mudsplashes) are purposefully and artificially introduced by the experimenter into an otherwise static scene (in vision or touch), in the change deafness paradigm, they are the natural result of the dynamic nature of complex auditory scenes that change and develop over time. From this perspective, change deafness potentially could be more akin to real-life phenomena than is change blindness. (We thank an anonymous reviewer for pointing out this interesting comparison.)

Whatever the reason for change-related auditory transients' ineffectiveness at attracting attention (thus resulting in change deafness), it remains to be determined on what basis the listeners could occasionally detect the auditory changes. Although we did not encourage participants to memorize the auditory objects in the scenes, we suggest that change perception in complex auditory scenes relies on aSTM. This is exactly what happens in the visual modality in change perception tasks when those transients that attract attention to the change location are removed (see, e.g., Luck & Vogel, 1997; Turatto & Bridgeman, 2005). Since we cannot exclude the possibility that the participants could have converted the acoustic objects into verbal descriptions of the corresponding animals, it is not clear whether the memory system involved in the present study is a nonverbal aSTM (as we have hypothesized) or the well-known verbal working memory system (Baddeley, 1986). The working memory model was composed of three main components: the *central executive*, which acts as the supervisory system and controls the flow of information from and to its slave systems; the *phonological loop*; and the *visuospatial sketchpad*. The phonological loop (or *articulatory loop*) as a whole deals with sound or phonological information and consists of two subcomponents: the phonological short-term store and the subvocal rehearsal mechanism. Since items in the store decay over time and become inaccessible after around 2,000–3,000 msec, the subvocal rehearsal mechanism (involving subvocal articulation in real time) is used to refresh items in the store in a serial, time-based manner. The fact that, in all experiments, the overall change perception accuracy ranged from 60% to 70% seems at odds with the verbal working memory hypothesis. Verbal working memory has typically been shown to possess a storage capacity ranging from five to seven items (Baddeley, 1986; but see Cowan, 2001). Thus, one would have expected a better performance with scenes consisting of only four objects if change detection relied mainly on verbal working memory. Note, however, that the brief duration (2 sec) of each auditory scene in our study could have made the subvocal rehearsal mechanism ineffective.

Another possibility is that performance is limited not by aSTM, but by the number of objects that attention can handle at the same time. For instance, as far as visual objects are concerned, evidence exists that the maximum number of objects that can be attended at the same time is around four (Pylyshyn & Storm, 1988; Rensink, 2000). In the terminology of Kahneman and his colleagues, that would mean that attention can keep open no more than four or five *files* of different items at the same time (Kahneman, Treisman, & Gibbs, 1992). So four objects might already be a limit for an efficient allocation of attention in an auditory scene. That would explain why we observed a consistent level of change deafness with just four objects. This interpretation, however, would seem to be less consistent with the results of Eramudugolla et al.'s (2005) study. They reported a level of performance comparable to ours with six objects, and performance was almost perfect with four objects. Note, however, that all our stimuli pertained to the category of animal calls, and one could speculate that the number of open object files that can be handled by attention is reduced when these are semantically related, as they were here.

In a recent review of the literature on auditory objects, Griffiths and Warren (2004) claimed that, with respect to object analysis in the auditory domain, the "single most pressing issue concerns the validity of visual analogies for auditory-object-processing mechanisms" (p. 892). The results of the present study clearly point to the fact that, at least for the mechanisms that govern perception of object change and object access to awareness, a fundamental difference exists between vision and audition in relation to the role of transients. If perceiving changes in visual objects follows directly from perceiving the visual transients associated with them, perceiving changes in auditory objects appears to rely to a greater extent on the short-term memory trace and on the attentional resources that can be enrolled in the concurrent analyses of multiple objects.

REFERENCES

AGOSTINELLI, G., SHERMAN, S. J., FAZIO, R. H., & HEARST, E. S. (1986).

Detecting and identifying change: Addition versus deletions. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 445-454.

ALAIN, C., & IZENBERG, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience*, **15**, 1063-1073.

BADDELEY, A. [D.] (1986). *Working memory*. Oxford: Oxford University Press, Clarendon Press.

BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sounds*. London: MIT Press.

BREITMEYER, B. G., & GANZ, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review*, **83**, 1-36.

COWAN, N. (1984). On short and long auditory stores. *Psychological Bulletin*, **96**, 341-370.

COWAN, N. (1995). *Attention and memory: An integrated framework*. Oxford: Oxford University Press.

COWAN, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, **24**, 87-114.

DARWIN, C. J., TURVEY, M. T., & CROWDER, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, **3**, 255-267.

ERAMUDUGOLLA, R., IRVINE, D. R. F., MCANALLY, K. I., MARTIN, R. L., & MATTINGLEY, J. B. (2005). Directed attention eliminates "change deafness" in complex auditory scenes. *Current Biology*, **15**, 1108-1113.

FERNANDEZ-DUQUE, D., & THORNTON, I. M. (2003). Explicit mechanisms do not account for implicit localization and identification of change: An empirical reply to Mitroff et al. (2002). *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 846-858.

GALLACE, A., AUVRAY, M., TAN, H. Z., & SPENCE, C. (2006). When visual transients impair change detection: A novel case of crossmodal change blindness? *Neuroscience Letters*, **398**, 280-285.

GALLACE, A., TAN, H. Z., & SPENCE, C. (2006). The failure to detect tactile change: A tactile analogue of visual change blindness. *Psychonomic Bulletin & Review*, **13**, 300-303.

GALLACE, A., TAN, H. Z., & SPENCE, C. (2007). Do "mudsplashes" induce tactile change blindness? *Perception & Psychophysics*, **69**, 477-486.

GREEN, D. M., & SWEETS, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Krieger.

GRIFFITHS, T. D., & WARREN, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, **5**, 887-892.

HEFFNER, R. S., & HEFFNER, H. E. (1992). Visual factors in sound localization in mammals. *Journal of Comparative Neurology*, **317**, 219-232.

JONIDES, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. [B.] Long & A. [D.] Baddeley (Eds.), *Attention and performance IX* (pp. 187-203). Hillsdale, NJ: Erlbaum.

KAHNEMAN, D., TREISMAN, A. [M.], & GIBBS, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, **24**, 175-219.

KUBOVY, M., & VAN VALKENBURG, D. (2001). Auditory and visual objects. *Cognition*, **80**, 97-126.

LALOYAUX, C., DESTREBECQZ, A., & CLEEREMANS, A. (2006). Implicit change identification: A replication of Fernandez-Duque and Thornton (2003). *Journal of Experimental Psychology: Human Perception & Performance*, **32**, 1366-1379.

LUCK, S. J., & VOGEL, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, **390**, 279-281.

MACK, A., & ROCK, I. (1998). *Inattentional blindness.* Cambridge, MA: MIT Press.

MACKEN, W. J., TREMBLAY, S., HOUGHTON, R. J., NICHOLLS, A. P., & JONES, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 43-51.

MAZZA, V., TURATTO, M., ROSSI, M., & UMILTÀ, C. (2007). How automatic are audiovisual links in exogenous spatial attention? *Neuropsychologia*, **45**, 514-522.

MONDY, S., & COLTHEART, V. (2000). Detection and identification of change in naturalistic scenes. *Visual Cognition*, **7**, 281-296.

NÄÄTÄNEN, R., TERVANIEMI, M., SUSSMAN, E. [S.], PAAVILAINEN, P., & WINKLER, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences*, **24**, 283-288.

O'REGAN, J. K., DEUBEL, H., CLARK, J. J., & RENSINK, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, **7**, 191-211.

O'REGAN, J. K., RENSINK, R. A., & CLARK, J. J. (1999). Change-blindness as a result of "mudsplashes." *Nature*, **398**, 34.

PHILLIPS, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, **16**, 283-290.

PYLYSHYN, Z. W., & STORM, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, **3**, 179-197.

RENSINK, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition*, **7**, 345-376.

RENSINK, R. A. (2002). Change detection. *Annual Review of Psychology*, **53**, 245-277.

RENSINK, R. A., O'REGAN, J. K., & CLARK, J. J. (1997). To see or not see: The need of attention to perceive changes in scenes. *Psychological Science*, **8**, 368-373.

RITTER, W., SUSSMAN, E. [S.], & MOLHOLM, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *NeuroReport*, **11**, 61-63.

SANTANGELO, V., OLIVETTI BELARDINELLI, M., & SPENCE, C. (2007). The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. *Journal of Experimental Psychology: Human Perception & Performance*, **33**, 137-148.

SIMONS, D. J., & LEVIN, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, **1**, 261-267.

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

SPENCE, C., & DRIVER, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, **59**, 1-22.

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

SUSSMAN, E. S. (2005). Integration and segregation in auditory scene analysis. *Journal of the Acoustical Society of America*, **117**, 1285-1298.

SUSSMAN, E. S., & GUMENYUK, V. (2005). Organization of sequential sounds in auditory memory. *NeuroReport*, **16**, 1519-1523.

SUSSMAN, E. [S.], RITTER, W., & VAUGHAN, H. G., JR. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, **789**, 130-138.

THORNTON, I. M., & FERNANDEZ-DUQUE, D. (2001). An implicit measure of undetected change. *Spatial Vision*, **14**, 21-44.

TREISMAN, M., & ROSTRON, A. B. (1972). Brief auditory storage: A modification of Sperling's paradigm applied to audition. *Acta Psychologica*, **36**, 161-170.

TURATTO, M., BETTELLA, S., UMILTÀ, C., & BRIDGEMAN, B. (2003). The perceptual conditions necessary to induce change blindness. *Visual Cognition*, **10**, 233-255.

TURATTO, M., & BRIDGEMAN, B. (2005). Change perception using visual transients: Object substitution and deletion. *Experimental Brain Research*, **167**, 595-608.

VITEVITCH, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 333-342.

WINKLER, I., SUSSMAN, E. [S.], TERVANIEMI, M., RITTER, W., HORVÁTH, J., & NÄÄTÄNEN, R. (2003). Preattentive auditory context effects. *Cognitive, Affective, & Behavioral Neuroscience*, **3**, 57-77.