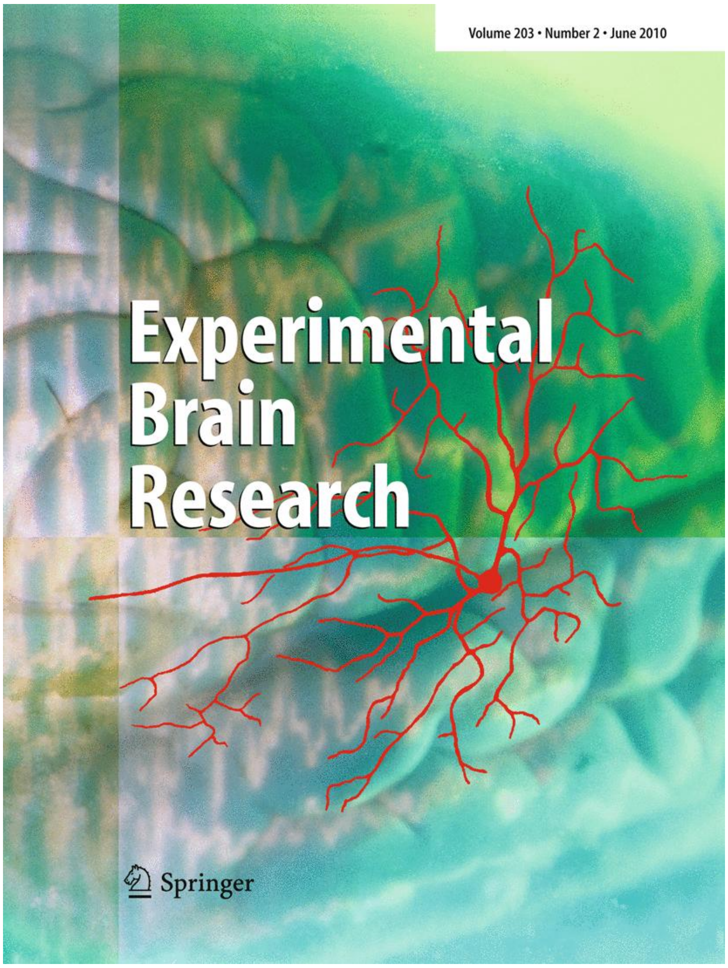
ISSN 0014-4819, Volume 203, Number 2



Exp Brain Res (2010) 203:261–270
DOI 10.1007/s00221-010-2226-2

RESEARCH ARTICLE

Fundamental differences in change detection between vision and audition

Laurent Demany · Catherine Semal ·
Jean-René Cazalets · Daniel Pressnitzer

Received: 10 August 2009 / Accepted: 9 March 2010 / Published online: 6 April 2010

Abstract We compared auditory change detection to visual change detection using closely matched stimuli and tasks in the two modalities. On each trial, participants were presented with a test stimulus consisting of ten elements: pure tones with various frequencies for audition, or dots with various spatial positions for vision. The test stimulus was preceded or followed by a probe stimulus consisting of a single element, and two change-detection tasks were performed. In the "present/absent" task, the probe either matched one randomly selected element of the test stimulus or none of them; participants reported present or absent. In the "direction-judgment" task, the probe was always slightly shifted relative to one randomly selected element of the test stimulus; participants reported the direction of the shift. Whereas visual performance was systematically better in the present/absent task than in the direction-judgment task, the opposite was true for auditory performance. Moreover, whereas visual performance was strongly dependent on selective attention and on the time interval separating the probe from the test stimulus, this was not the case for auditory performance. Our results show that small auditory changes can be detected automatically across relatively long temporal gaps, using an implicit memory system that seems to have no similar counterpart in the visual domain.

Keywords Change detection · Temporal binding · Sensory memory · Attention

Introduction

A basic challenge for the brain is to establish links between successive auditory or visual stimuli and to detect changes between them. If the stimuli are not contiguous in time, this requires memory. When two successive complex stimuli, such as two views of a landscape or two orchestral chords, differ from each other with respect to some detail, the psychological outcome of the brain's comparison can take several forms. The subject may perceive the change as the addition of some new entity to the initial stimulus, or the disappearance of an entity; a third possibility is a percept of motion or shift, for instance an object changing position in the landscape or a tone changing pitch in the chord. These three types of change can be perceived auditorily (Bregman 1990) as well as visually (Rensink 2002). In both sensory modalities, moreover, change detection is to some extent automatic: under certain conditions, subjects can detect a change even if they did not pay attention to the relevant part of the sensory field (Theeuwes et al. 1998; Tiitinen et al. 1994; Yantis and Hillstrom 1994). One does not know, however, if the sensori-mnemonic mechanisms of auditory and visual change detection are essentially similar or significantly different.

In the domain of vision, precise proposals have been made regarding the neural processes responsible for the automatic detection of changes (Beck et al. 2001; Muckli et al. 2005; Newsome et al. 1986; Phillips and Singer 1974a, b). Psychophysical studies have demonstrated that human observers possess a fine-grain "iconic" memory thanks to which small local changes in a complex visual

L. Demany (✉) · C. Semal · J.-R. Cazalets
Laboratoire Mouvement, Adaptation,
Cognition (UMR CNRS 5227), Université de Bordeaux,
146 rue Leo Saignat, 33076 Bordeaux, France
e-mail: laurent.demany@u-bordeaux2.fr

D. Pressnitzer
Laboratoire Psychologie de la Perception (UMR CNRS 8158),
Université Paris-Descartes and Ecole
Normale Supérieure, Paris, France

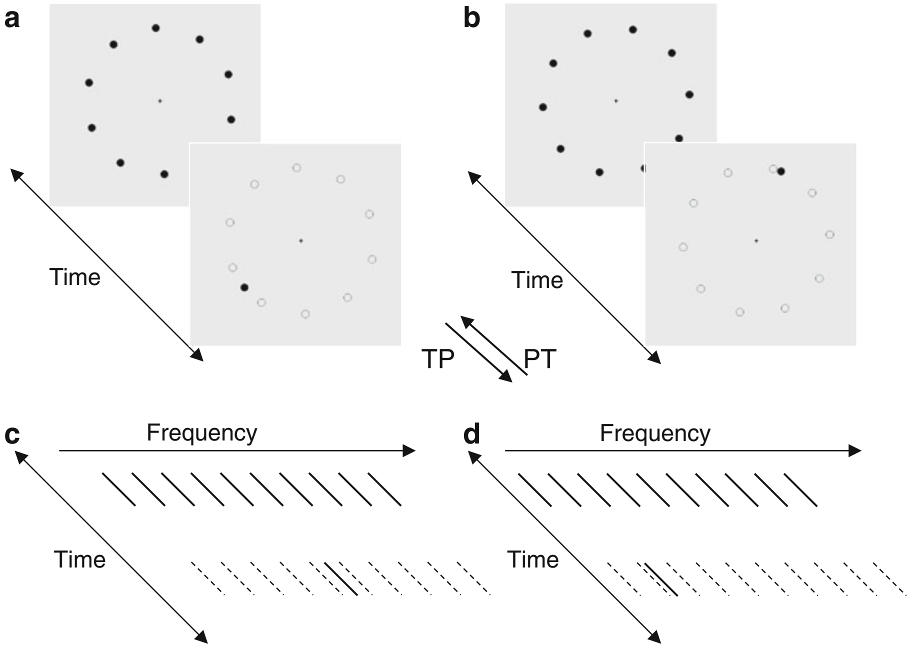


Fig. 1 Examples of test and probe stimuli used in the experiments. **a** A visual trial in the present/absent task, on which the correct response is "absent". The test stimulus (*top, left*) consists of 10 dots regularly arranged on a circle centered on the visual fixation cross. The probe (*bottom, right*) consists of a single dot, positioned halfway between two randomly selected elements of the test stimulus. The test stimulus could be presented before the probe (TP condition) or after it (PT condition). **b** A visual trial in the direction-judgment task, on which the correct response is "clockwise" for the TP condition, and "anticlockwise" for the PT condition. **c** An auditory trial in the present/absent task, on which the correct response is "absent". The test stimulus (*top, left*) consists of 10 synchronous pure tones regularly spaced on a logarithmic frequency scale. The probe (*bottom, right*) consists of a single tone, positioned halfway between two randomly selected elements of the test stimulus. Again, the test stimulus could be presented before the probe (TP condition) or after it (PT condition) **d** An auditory trial in the direction-judgment task, on which the correct response is "up" for the TP condition, and "down" for the PT condition

image can be detected effortlessly in the presence of a temporal gap (Pashler 1988; Phillips 1974; Stelmach et al. 1984). However, the storage time of iconic memory appears to be quite short: less than 100 ms. When details of a complex image have to be held for more than 100 ms, change detection is no longer automatic but requires voluntary attention, and human observers show a surprising "change blindness" (Luck and Vogel 1997; O'Regan et al. 1999; Rensink et al. 1997).

Concerning audition, numerous papers have reported neural correlates of acoustic changes in auditory cortical areas (Brosch and Schreiner 2000; Chait et al. 2007; Downar et al. 2000; McKenna et al. 1989; Näätänen et al. 2001), but the neural mechanisms of change detection remain mysterious. It has been concluded from some studies that auditory memory behaves like visual memory (Cowan 1995; Massaro and Loftus 1996; Visscher et al. 2007) and that change blindness has an auditory homologue, "change deafness" (Eramudugolla et al. 2005; Gregg and Samuel 2008; Pavani and Turatto 2008; Vitevitch 2003). In contrast, other experiments on auditory memory (Demany et al. 2008) have suggested that the auditory counterpart of iconic memory has a much longer storage time than iconic memory. This suggests, in turn, that automatic change detection might be achieved by different means in audition and vision. Whether visual and auditory change detection recruit essentially similar or fundamentally different processes, a basic question with clear implications for multimodal integration, thus remains unsettled.

In the present psychophysical study, auditory change detection and visual change detection were directly compared using stimuli and tasks which were matched as closely as possible. On each trial, in our main experiments, the subject was presented with a brief "test" stimulus consisting of 10 elements. On visual trials, the elements were dots arranged along a circle (see Fig. 1a, b). On auditory trials, the elements were simultaneous pure tones without harmonic relationships (Fig. 1c, d). Participants had to perform two different change-detection tasks, in separate blocks of trials. They always compared the test stimulus to a single element, called the "probe". In the "present/absent" task, the probe could be the exact replica of a randomly selected element of the test stimulus, or it could be equidistant from two randomly selected neighbouring elements of the test stimulus (Fig. 1a, c). The task was then to indicate whether the probe was present in the test stimulus or absent from it. In the "direction-judgment" task, on the other hand, the probe was always a shifted replica of a randomly selected element of the test stimulus, and one had to identify the direction of this shift. In the visual modality, the probe dot could be shifted either clockwise or anticlockwise along the circle formed by the test stimulus (Fig. 1b). In the auditory modality, the probe tone could be shifted either up or down in frequency (Fig. 1d). In both modalities, the magnitude of the shift always corresponded to 18% of the distance between neighbouring test elements.

The present/absent task required subjects to detect changes which were large, but such that there was no obvious relation between test and probe. Smaller changes had to be detected in the direction-judgment task, but these changes were such that one element of the test stimulus was close to the probe. Given that change magnitude was larger in the present/absent task, one would expect subjects to perform better in that task than in the direction-judgment task. For the auditory modality, however, a different outcome could be predicted. In a previous auditory experiment (Demany and Ramos 2005, experiment 1), listeners had performed both the present/absent task and the direction-judgment task with test stimuli akin to those used here. Surprisingly, performance was found to be better in the direction-judgment task than in the present/absent task. This paradoxical result was accounted for by assuming that the auditory system contains automatic detectors of frequency shifts. The detectors in question were thought of as an auditory counterpart of the automatic spatial-motion detectors of the visual system (Ditterich et al. 2003). Under that hypothesis, it could be expected that we would be able to observe here an advantage of the direction-judgment task over the present/absent task for vision as well as audition.

The two tasks were performed in a condition where the test stimulus preceded the probe (“TP” condition) and also in a condition where the probe preceded the test stimulus (“PT” condition). In the TP condition, the subject did not know at the onset of a trial which element of the test stimulus was going to be the relevant element for that trial. Hence, the entire set of 10 elements had to be memorized, as far as possible. Crucially, the 10 elements had to be memorized individually since the probe consisted of a single element; in other words, change detection could not be based on “Gestalt” cues. In the PT condition, by contrast, the probe drew the subject’s attention onto the appropriate region of the test stimulus, and only one element (the probe) had to be memorized. Since identical stimuli and tasks were used in the two conditions, a performance drop from the PT condition to the TP condition was wholly attributable to a shortage of memory and/or attention in the latter condition. By comparing the performance drops observed for the visual stimuli to those observed for the auditory stimuli, we determined whether memory and attention limit change detection similarly or not for vision and audition.

The choice to compare the perception of spatial changes for vision to the perception of frequency changes for audition was primarily motivated by the fact that space and frequency are the topographic dimension of the sensory epithelium in vision and audition, respectively. We will come back to this choice of dimensions in the Discussion, but let us already point out that several authors have already argued that the appropriate auditory analogy to visual space is auditory frequency, not auditory space (Kubovy and van Valkenburg 2001; Scholl 2001; Shamma 2001; Griffiths and Warren 2004).

Experiment 1

Experiment 1 consisted of two parallel sub-experiments—one on vision (experiment 1-VIS) and the other on audition (experiment 1-AUD)—in which the test stimulus and the probe were separated by a 500-ms inter-stimulus interval (ISI) in the TP and PT conditions.

Methods

Stimuli

In experiment 1-AUD, each test stimulus was an inharmonic “chord” of ten synchronous pure tones separated by frequency intervals of 5.5 semitones (i.e., 0.458 octave). From trial to trial, this chord was randomly transposed within a logarithmically scaled frequency range bounded by two rules: the lowest element of the chord had a minimum frequency of 125 Hz, and the highest element had a maximum frequency of 4,000 Hz. Due to these rules, the chord as a whole varied from trial to trial in a 10.5-semitone range. On each direction-judgment trial, the probe tone preceding or following the chord was positioned 1 semitone away from a randomly chosen element of the chord. In the present/absent task, when the correct response was “absent”, the probe was 2.75 semitones away from two adjacent elements of the chord. Whatever the task or condition (PT or TP), any trial began with the presentation of a 1.5-s random sequence of five pure tones, with frequencies drawn between 125 and 4,000 Hz. The next stimulus (test or probe) followed this random sequence after a 600-ms ISI. Each stimulus had a total duration of 300 ms and was gated on and off using raised-cosine amplitude ramps of 20 ms. Each pure tone had a nominal sound pressure level of 65 dB. The stimuli were presented binaurally (diotically), via headphones (Sennheiser HD265 or HD250 Linear II).

In experiment 1-VIS, each test stimulus was an array of 10 circular dots regularly distributed on an imaginary circle centered on a visual fixation cross. The visual angle subtended by the whole array was 17.3°. Two shades of gray were used to produce the dots (luminance: 17.6 cd/m^2) and their background (luminance: 32.6 cd/m^2). The visual angle subtended by a dot was 0.8°. The computer screen displaying the stimuli was at a distance of 49.5 cm from the eyes of the subject, whose head was immobilized by means of a chinrest. From trial to trial, the orientation of the array was randomly varied. On each trial, the probe dot preceding or following the array was positioned on the same circle as the array elements: in the direction-judgment task, a randomly chosen element of the array was rotated by $2\pi/55$ radians, clockwise or anticlockwise; in the present/absent task, when the correct response was "absent", the corresponding rotation angle was $2\pi/20$. As in experiment 1-AUD, all stimuli had a total duration of 300 ms, and a 1.5-s random sequence of five elements (in this case single dots at random positions along the imaginary circle referred to above) was presented at the beginning of each trial; there was again a 600-ms ISI between this sequence and the test stimulus (in the TP condition) or the probe (in the PT condition).

Procedure

In experiment 1-AUD, subjects were tested in a double-walled soundproof booth, where they gave their responses by means of mouse clicks on two virtual buttons displayed on a computer screen. In experiment 1-VIS, subjects were tested in a dimly lit room and responses were given by pressing one of two keys on a keyboard.

In both sub-experiments, the two possible responses on a given trial ("present" vs. "absent", "up" vs. "down", or "clockwise" vs. "anticlockwise") had the same a priori probability of being correct. Response time was unlimited. Subjects were informed of their results following each block of trials. A given experimental session consisted of one block of 50 trials in each of the four possible sub-conditions [2 conditions (TP or PT) × 2 tasks], which were randomly ordered. Within a given block, any trial started immediately after the response given on the previous trial. For each subject and sub-experiment, eight sessions were run, following an initial practice phase (see the next section, "Participants"); 400 trials per subject were therefore performed in each sub-condition.

In experiment 1-VIS, subjects were instructed to keep their eyes on the fixation cross throughout the 1,100-ms interval encompassing, on each trial, the presentation of the test and probe stimuli. It was important to check that this instruction was followed in the PT condition, where it would have been advantageous for the subject to fixate the probe visually (rather than only attentionally) when it was presented, and then to wait for the test stimulus with the eyes in that position. We thus used an eye-tracking system, based on infrared reflection (R6, Applied Science Laboratories, Bedford, MA). Gaze direction (dominant eye) was acquired at a rate of 50 Hz. A trial was excluded from the analysis if any one of the following two criteria was met: (1) the distance between gaze and fixation cross exceeded 1.2° of visual angle (1 cm on the screen), in any direction, for at least one sample during the crucial time interval; (2) more than six samples were missing for the whole duration of the trial. Based on these criteria, we discarded 14.7% of the trials run in experiment 1-VIS (15.2% in the TP condition and 14.2% in the PT condition).

Participants

Each sub-experiment was carried out on seven subjects, among whom three took part in both sub-experiments. Thus, 11 subjects were tested overall. Nine of them were students in their twenties; the remaining two subjects were author DP (tested in experiment 1-VIS) and author LD (tested in both sub-experiments). All participants in experiment 1-AUD had normal hearing (audiometric thresholds lower than 20 dB HL at both ears, up to at least 4,000 Hz). Among these seven subjects, five had previously served as subjects in related auditory studies; they were given only one short practice session initially; the remaining two subjects were initially trained in the two tasks for a longer period (2 or 3 h). Participants in experiment 1-VIS had normal or corrected-to-normal vision and no previous experience in related tasks. They were initially trained for about 2 h.

Results and discussion

Figure 2 shows subjects' performance in terms of the sensitivity index d' (Macmillan and Creelman 1991) for each sensory modality (audition or vision), task (present/absent or direction-judgment), and condition (TP or PT). A three-way ANOVA of these data revealed a significant main effect of condition [$F(1, 48) = 65.1$, $P < 0.001$]: as could be expected, performance was globally poorer when the probe was presented after, rather than before, the test stimulus. In contrast, there was no significant main effect of sensory modality or task ($P > 0.10$ in each case). The absence of a global effect of sensory modality is noteworthy but merely indicates that we had successfully matched the prothetic parameters of the visual and auditory stimuli. More importantly, the ANOVA revealed the presence of two—and only two—significant interactions: one between sensory modality and condition [$F(1, 48) = 16.4$, $P < 0.001$], the other between sensory modality and task [$F(1, 48) = 26.6$, $P < 0.001$].

The first of these interactions reflects the fact that presenting the probe after the test stimulus rather than before

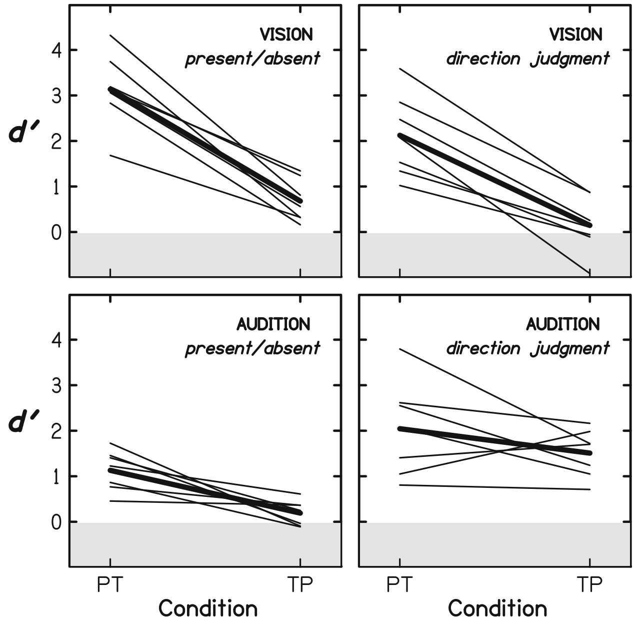


Fig. 2 Performance (d') in experiment 1-VIS (*upper panels*) and experiment 1-AUD (*lower panels*), as a function of the task (present/absent or direction-judgment) and the order in which the test and probe stimuli were presented (PT or TP). *Thin lines* represent individual subjects, and *thick lines* represent means across subjects. Performance was measured in terms of d', the sensitivity index of signal detection theory. In the visual experiment, subjects were much more efficient when the probe preceded the test stimulus than when the test stimulus was presented first, and performance was better in the present/absent task than in the direction-judgment task. In the auditory experiment, on the other hand, the order in which the two stimuli were presented had only a weak effect on performance, and subjects were more efficient in the direction-judgment task than in the present/absent task

the test stimulus impaired performance much more for vision than for audition, whatever the task. The mean difference between the d' values measured in the two conditions was three times larger for vision (2.22) than for audition (0.74). This implies that in the TP condition, the extent to which memory factors and/or attentional factors imposed limits on change detection was considerably larger for visual changes than for auditory changes.

The second interaction stems from opposite task effects for the two sensory modalities: with the visual stimuli, the present/absent task was significantly easier than the direction-judgment task ($P = 0.005$, Tukey test); with the auditory stimuli, the reverse was true ($P < 0.001$). This is not because the visual version of the direction-judgment task was intrinsically more difficult than the auditory version: in the PT condition, the mean d' obtained for the visual direction-judgment task (2.12) was very similar to (and even slightly larger than) the mean d' obtained for the auditory direction-judgment task (2.04). Thus, when attention was focused on the relevant part of the test stimuli, the auditory and visual direction-judgment tasks were matched in difficulty. Nevertheless, performance was always better for present/absent judgments in the visual modality and always better for direction judgments in the auditory modality. This pronounced contrast suggests that visual change detection and auditory change detection are based on dissimilar mechanisms.

In addition to the ANOVAs reported above, we also performed a two-way ANOVA (Sensory modality × Condition) on subjects' response biases in the present/absent task. (The direction-judgment task had to be discarded from the analysis since there was no logical way to match a given visual direction with a given auditory direction). Our index of bias was the c statistic of signal detection theory (Macmillan and Creelman 1991). The ANOVA did not reveal any main or interaction effect [$F(1, 24) < 1$ in each case]. Thus, in the present/absent task, there was no intermodal difference regarding response strategy.

Experiment 2

In experiment 1, the test stimulus and the probe were always separated by a 500-ms ISI. Such an ISI might be too long for the automatic and direction-sensitive motion detectors of the visual system (Newsome et al. 1986), but not too long for auditory detectors of frequency shifts (Demany and Ramos 2005; Demany et al. 2008). We thus had to consider the possibility that the interaction between task and sensory modality observed in experiment 1 revealed only a quantitative difference between the two modalities, rather than a basic, qualitative difference. Besides, it could be thought that the counterintuive task effect found with the auditory stimuli would perhaps disappear for ISIs well below 500 ms. These considerations motivated experiment 2, which was a replication of experiment 1 with ISIs of 0, 12, 24, 94, and 188 ms.

Methods

The stimuli were identical to those of experiment 1 except that each auditory stimulus was now gated on and off with 6-ms amplitude ramps. The tasks were exactly the same as before. However, we omitted the PT condition, and for this reason, the eye-tracking system was no longer used. There was a visual TP condition (experiment 2-VIS) and an auditory TP condition (experiment 2-AUD). Ten sub-conditions (5 ISIs × 2 tasks) were run in each case. As in experiment 1, all sub-conditions were run once per session, in a random order, within blocks of 50 trials. Again, each sub-experiment consisted of eight sessions, i.e., a total of 400 trials per participant in each sub-condition.

The two experimental groups consisted each of five subjects and had two common members. The eight subjects tested overall included six students, author DP (tested in experiment 2-VIS), and author LD (tested in both subexperiments). Three subjects had previously participated in experiment 1. The five new subjects were initially familiarized with the tasks as in experiment 1. All subjects had normal hearing and/or normal or corrected-to-normal vision.

Results

The obtained values of d' are displayed in Fig. 3. As in our previous experiment, subjects behaved very differently in the visual and auditory situations. In experiment 2-VIS, the present/absent task was performed better than the direction-judgment task whatever the ISI; however, subjects were also more efficient when the ISI was short than when it was long. In experiment 2-AUD, by contrast, the direction-judgment task was always performed better than the present/absent task and performance did not depend on the ISI. A three-way ANOVA of the individual data (Sensory modality × Task × ISI) confirmed the existence of significant interactions between Sensory modality and Task [$F(1, 80) = 98.4$, $P < 0.001$] and between Sensory modality and ISI [$F(4, 80) = 6.5$, $P < 0.001$]. There was no significant three-way interaction [$F(4, 80) = 1.5$, $P = 0.22$]. In the direction-judgment task, for the shortest ISI (0 ms), the average visual performance ($d' = 1.18$) was very similar to the average auditory performance ($d' = 1.00$); but for the longest ISI (188 ms), the average auditory performance ($d' = 1.02$) was definitely better than the average visual performance ($d' = 0.10$), and this difference was statistically significant [$t(8) = 4.4$, $P = 0.002$, two-tailed test]. Overall, performance was significantly better for the visual stimuli than for the auditory stimuli [$F(1, 80) = 29.6$, $P < 0.001$], because the visual present/absent task was easy when the ISI was very short.

As in experiment 1, a two-way ANOVA (Sensory modality × ISI) of the values of c for the present/absent task did not reveal any main or interaction effect ($F < 1$ in each case).

The foregoing analyses of subjects' behavior, in experiment 1 as well as experiment 2, used the conceptual framework of signal detection theory. In this conceptual framework, the internal representation of each stimulus element is a *continuous* variable: change detection is limited by Gaussian internal noise. Such a model has been justified by experimental evidence (Wilken and Ma 2004). However, other models are conceivable (Micheyl et al. 2008). As proposed by Pashler (1988), it may be assumed that a given stimulus element is either processed, and then encoded perfectly, or not processed, and then not encoded at all. Under this model, performance in the TP condition is determined by the number of test stimulus elements that the subject is able to process. This *capacity* can be computed as $N * (H - F)/(1 - F)$, where N is the total number of test elements, H is the probability of a hit (e.g., a correct "absent" response), and F is the probability of a false alarm. Figure 4 displays the results of experiment 2 in terms of capacity (we also plot in this figure the results obtained for the TP condition of experiment 1). Satisfactorily, it can be seen that all the trends described above with respect to d' are again observable with respect to capacity.

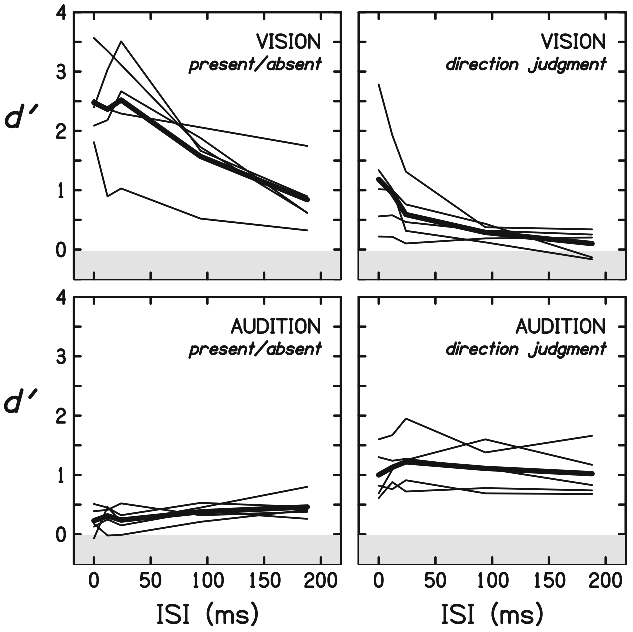


Fig. 3 Performance (d') in experiment 2-VIS (*upper panels*) and experiment 2-AUD (*lower panels*), as a function of the task (present/absent or direction-judgment) and the time interval (ISI) separating the test stimulus from the probe. The test stimulus was always presented first. *Thin lines* represent individual subjects and *thick lines* mean data. In the visual experiment, performance was better when the ISI was very short than when it was longer, and subjects were more efficient in the present/absent task than in the direction-judgment task. In the auditory experiment, by contrast, increasing the ISI did not impair performance and subjects were more efficient in the direction-judgment task than in the present/absent task

Discussion

The task effects found in experiment 2 replicate those of experiment 1. In the visual present/absent task, mean performance was good when the ISI was 0–24 ms, but much worse for a 188-ms ISI. This is consistent with the results of several previous studies in which observers had to detect small differences between complex visual stimuli separated by variable ISIs, ranging from 0 to a few hundreds of ms (Phillips 1974; Stelmach et al. 1984; Pashler 1988). It was found in all these studies that detection performance decreased very rapidly as the ISI increased. This was the case when the detected changes were experienced as movements (Pashler 1988) as well as when the change was

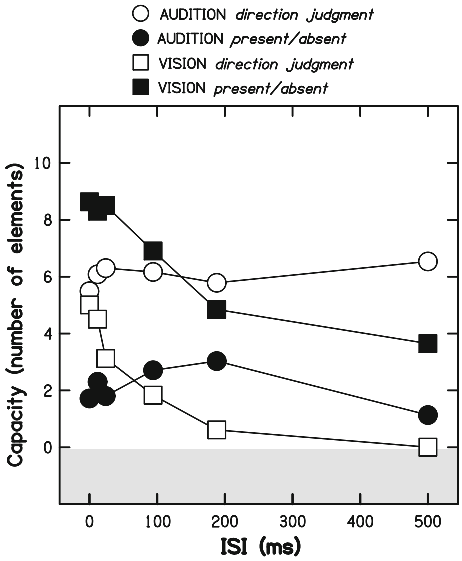


Fig. 4 Capacity measures of performance in experiment 1 (TP condition; ISI = 500 ms) and experiment 2 (ISI = 0–188 ms). Capacity was computed as suggested by Pashler (1988). Each data point represents the average capacity obtained for the tested subjects. The trends observable here are similar to those observable in Figs. 2 and 3

instead experienced as an appearance or disappearance (Phillips 1974).

Crucially, the fact that performance was good for the shortest ISIs indicates that for longer ISIs (including the 500-ms ISI used in experiment 1), subjects' difficulty in the visual TP condition reflects memory limitations. In the auditory TP condition, on the other hand, performance was apparently not limited by memory. It is remarkable that in this condition, we found an advantage of the direction-judgment task over the present/absent task even for ISIs below 100 ms, because for such ISIs a sound can perceptually mask a subsequent sound; since this phenomenon, called forward masking, shows a sharp frequency selectivity (Moore 2004), exploiting it in experiment 2-AUD would have facilitated the present/absent task much more than the direction-judgment task.

Experiment 3

This short final experiment was intended to determine whether, for visual test stimuli consisting of more than 10 dots, the direction-judgment task could become easier than the present/absent task. Such a reversal of trend was a priori conceivable because increasing the number of dots was expected to increase the difficulty to perceive a given dot individually (Intriligator and Cavanagh 2001). We used in experiment 3 circular arrays of 20 dots. The arrays had the same diameter as in experiments 1-VIS and 2-VIS, but the dots now subtended a visual angle of 0.5°. In the direction-judgment task, the spatial shifts of the dots were again equal to 18% of the distance between neighbouring test elements. Five subjects, including authors LD and DP, were tested in the TP condition, with ISIs of 0, 12, and 24 ms. Four of these subjects had been previously tested in experiment 1-VIS and/or experiment 2-VIS. For each combination of ISI and task, 200 trials per subject were run. From the shortest to the longest ISI, d' took mean values of 2.47, 1.44, and 1.82 for the present/absent task, versus 0.29, 0.28, and 0.12 for the direction-judgment task. These data are basically similar to those obtained in experiment 2-VIS.

General discussion

Contrary to the common belief that auditory memory and visual memory function in the same way (Cowan 1995; Massaro and Loftus 1996; Visscher et al. 2007), our study suggests that there are profound differences between the sensori-mnemonic processes involved in visual versus auditory change detection. The evidence is threefold.

One piece of evidence is the differential ISI effect observed in experiment 2, indicating that fine-grain sensory information is stored for a longer time period in auditory memory than in visual memory. This finding was actually predictable from previous research. The capacity limits of the memory system called visual short-term memory are notoriously narrow (Alvarez and Cavanagh 2004; Awh et al. 2007; Bays and Husain 2008; Luck and Vogel 1997; Zhang and Luck 2008), and it has often been argued that they severely restrict the detectability of changes between two complex visual images separated by more than a few tens of ms. Our results are consistent with this notion, as well as with the idea that visual short-term memory is complemented by an "iconic" memory system with an essentially unlimited capacity but a very short storage time (Phillips 1974). Regarding auditory memory, on the other hand, the present results concur with those of Demany et al. (2008) in showing that complex static sounds are memorized quasi-perfectly for hundreds of ms. As pointed out in this previous paper, the idea that humans are afflicted with "change deafness", in addition to change blindness, stems from studies in which the stimuli to be compared by the participants were not static sounds but acoustic patterns already including temporal changes.

A second difference observed here between visual and auditory change detection is a differential effect of attention. In experiment 1, visual performance was markedly better when the focus of attention was directed to the relevant element of the test stimulus (PT condition) than when the relevant element was unpredictable (TP condition); auditory performance, in contrast, was only weakly affected by this manipulation. Thus, auditory change detection appears to be more "automatic" than visual change detection. What we found for vision tallies with previous demonstrations of the major role of attention in visual change detection (O'Regan et al. 1999; Rensink et al. 1997). The absence of a similar auditory effect is more surprising insofar as attention can markedly affect performance in auditory tasks of various types (Best et al. 2008; Hafter et al. 2008). However, the fact that attention had no significant effect in our auditory direction-judgment task is logically consistent with the fact that this task was performed much better than the auditory present/absent task: both of these results indicate that a pitch change between two tones can be consciously heard even when one of them has not been consciously perceived individually.

Whereas, in the auditory modality, the direction-judgment task was always performed better than the present/absent task, the reverse was always true in the visual modality. This third differential effect is our main finding.

The advantage of the present/absent task in our visual experiments can be explained in two ways. Most importantly, the changes to be detected in the present/absent task were larger than those to be detected in the direction-judgment task. A secondary factor, when the ISI was very short, may have been an additional perceptual cue produced by local interactions between transient neural responses to onsets and offsets. As shown by Phillips and Singer (1974a, b), when a given visual stimulus (such as a dot) is presented twice at the same place with a short ISI between the two presentations, the neural response to the onset of the second presentation is affected by the offset of the first presentation. This presumably provided a flicker cue in the present/absent task of experiment 2-VIS. Such a cue could not occur in the direction-judgment task. The poor performance in the latter task shows that the automatic motion-detectors of the visual system generally failed to provide reliable perceptual cues when direction had to be judged. This does not imply that no motion was perceived: it is also possible that many motion signals were generated, but failed to indicate reliably the direction of the change. An overintegration of motion signals from large parts of the visual field could thus explain our results. This interpretation is supported by the subjective reports of participants, who claimed to respond without hesitation in the direction-judgment task because they could see motion, even though their performance was in fact close to chance.

In our auditory experiments, by contrast, the direction judgments were always more accurate than the present/absent judgments. This is a striking finding since it means that relatively small changes were better detected than larger changes. One can account for such a finding with a simple model of frequency-shift detectors (FSDs) (Demany and Ramos 2005; Demany et al. 2009). Suppose that each FSD responds optimally to small shifts in a given direction, the upward direction for one subset of FSDs and the downward direction for the remaining subset. On a direction-judgment trial, one of the two subsets will tend to respond more vigorously than the other subset, and this asymmetry will point to the correct response. However, on both "present" and "absent" trials, there will be no definite response asymmetry. Hence, if the consciously available information about the FSDs' activity is only the relative difference between the response strengths of the "upward" and "downward" subsets, the FSDs will provide no helpful cue in the present/absent task. The model that we just described is strongly reminiscent of a recent conceptualization of visual motion perception (Ditterich et al. 2003). Yet, our study as a whole suggests that the analogy between auditory FSDs and visual motion detectors cannot be carried far. In any case, the FSDs appear to operate on internal representations of frequency that are not accessible to consciousness and persist in some implicit auditory memory for at least 500 ms after the end of the stimulus. A similar memory system does not seem to exist for vision.

To be of interest, the differences that we observed between vision and audition should not depend on peculiarities of the visual and auditory stimuli. While we cannot rule out this possibility entirely, efforts were made to select comparable stimuli. We compared the detection of spatial shifts in dots to the detection of frequency shifts in pure tones. Two points can be made to justify such a choice. First, it is reasonable to consider dots and pure tones as subjectively homologous since these two types of stimuli are subjectively elementary, that is, impossible to split up into subparts. Second, as noted previously by a number of authors (e.g., Kubovy and van Valkenburg 2001), spatial position in vision is better likened, in the auditory domain, to frequency than to spatial position itself. Whereas dots differing in spatial position as well as tones differing in frequency activate different neurons in the sensory periphery, this is not the case for tones differing only in spatial position. Indeed, simultaneous tones differing only in spatial position are impossible to perceive as separate entities. Thus, frequency is the primary dimension of sounds for the auditory system, and as such it is the appropriate "auditory translation" of spatial position in vision. One might think that by using chords of tones differing both in frequency and in spatial position, it is possible to investigate the perception of auditory spatial shifts in an experimental paradigm similar to the one used here for vision. However, this is in fact difficult due to the limited spatial resolution of simultaneous tones by the binaural system (see, e.g., Divenyi and Oliver 1989). Moreover, such an experiment would not be a logically satisfactory transposition of our visual experiments since the shifts to be detected would occur in a dimension that differs from the one in which the elements of the test stimuli are primarily individualized.

In a chord of pure tones differing in frequency, hearing out a given tone as a separate entity can be difficult even when the frequency differences are large enough to prevent significant overlaps of the tones' neural representations in the auditory system. This results from a phenomenon called informational masking (Kidd et al. 2008). Auditory informational masking has a visual counterpart, called crowding in the visual literature (Intriligator and Cavanagh 2001). Both effects tend to increase with the number and proximity of the elements making up the stimulus. In experiments 1 and 2, it may well be that the amount of crowding affecting the perception of our 10-element visual stimuli was smaller than the amount of informational masking affecting the perception of the corresponding auditory stimuli. However, when we used instead 20-element visual stimuli, in experiment 3, we found again that the present/absent task was performed better than the direction-judgment task; in addition, the advantage of the present/absent task was not weaker than the advantage previously observed for 10-element stimuli, and performance in the direction-judgment task was very close to the chance level. So, it seems impossible to design a configuration of dots for which the direction-judgment task would be easier than the present/absent task. With tones rather than dots, on the contrary, one obtains such a trend quite invariably, without having to use peculiar tonal configurations.

Although we cannot completely rule out the existence of visual phenomena akin to the auditory phenomena reported here, our study suggests that auditory memory is, at least in some respects, more sophisticated than visual memory. Specifically, auditory memory appears to be a more efficient tool for the automatic identification of relations between sensory events that differ from each other and are separated by relatively long temporal gaps. This makes sense from an ecological point of view. In everyday life, memorizing the details of a complex visual scene for hundreds of ms is superfluous because the visual world constitutes an external memory (O'Regan and Noë 2001): the scene will generally remain available for scrutiny. Moreover, it is unlikely that an element of the scene will disappear and reappear some time later at a different position; instead, changes in position are most often continuous motions. By contrast, humans typically extract auditory information from fleeting sequences of sounds that differ from each other in frequency content and can be separated by substantial silent pauses. These successive sounds need to be linked perceptually in order to form recognizable words, sentences or tunes (Demany and Semal 2008). An essential function of the FSDs is presumably to create a link between successive sounds which are spectrally close to each other; that would be useful because such sounds are likely to originate from one and the same acoustic source rather than from separate and concurrently active sources (Bregman 1990). In any case, the binding of successive sounds into higher-order auditory entities such as those which are meaningful for human listeners clearly calls for an elaborate and largely automatic auditory memory.

Acknowledgments Authors LD and DP contributed equally to this work. We thank Prof. Patrick Cavanagh and Dr. Andrei Gorea for helpful discussions about the visual experiments. We also thank Dr. Etienne Guillaud for technical assistance and Dr. Trevor Agus for suggestions on the manuscript.

References

Alvarez GA, Cavanagh P (2004) The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychol Sci 15:106–111

Awh E, Barton B, Vogel EK (2007) Visual working memory represents a fixed number of items regardless of complexity. Psychol Sci 18:622–628

Bays PM, Husain M (2008) Dynamic shifts of limited working memory resources in human vision. Science 321:851–854

Beck DM, Rees G, Frith CD, Lavie N (2001) Neural correlates of change detection and change blindness. Nat Neurosci 4:645–650

Best V, Ozmeral EJ, Kopco N, Shinn-Cunningham BG (2008) Object continuity enhances selective auditory attention. Proc Nat Acad Sci USA 105:13173–13177

Bregman AS (1990) Auditory scene analysis. MIT Press, Cambridge

Brosch M, Schreiner CE (2000) Sequence sensitivity of neurons in cat primary auditory cortex. Cereb Cortex 10:1155–1167

Chait M, Poeppel D, de Cheveigné A, Simon JZ (2007) Processing asymmetry of transitions between order and disorder in human auditory cortex. J Neurosci 27:5207–5214

Cowan N (1995) Attention and memory. Oxford University Press, New York

Demany L, Ramos C (2005) On the binding of successive sounds: perceiving shifts in nonperceived pitches. J Acoust Soc Am 117:833–841

Demany L, Semal C (2008) The role of memory in auditory perception. In: Yost WA, Popper AN, Fay RR (eds) Auditory perception of sound sources. Springer, New York, pp 77–113

Demany L, Trost W, Serman M, Semal C (2008) Auditory change detection: simple sounds are not memorized better than complex sounds. Psychol Sci 19:85–91

Demany L, Pressnitzer D, Semal C (2009) Tuning properties of the auditory frequency-shift detectors. J Acoust Soc Am 126:1342–1348

Ditterich J, Mazurek ME, Shadlen MN (2003) Microstimulation of visual cortex affects the speed of perceptual decisions. Nat Neurosci 6:891–898

Divenyi PL, Oliver SK (1989) Resolution of steady-state sounds in simulated auditory space. J Acoust Soc Am 85:2042–2052

Downar J, Crawley AP, Mikulis DJ, Davis KD (2000) A multimodal cortical network for the detection of changes in the sensory environment. Nat Neurosci 3:77–283

Eramudugolla R, Irvine DRF, McAnally KI, Martin RL, Mattingley JB (2005) Directed attention eliminates 'change deafness' in complex auditory scenes. Curr Biol 15:1108–1113

Gregg MK, Samuel AG (2008) Change deafness and the organizational properties of sounds. J Exp Psychol Hum Percept Perform 34:974–991

Griffiths TD, Warren JD (2004) What is an auditory object? Nat Rev Neurosci 5:887–892

Hafter ER, Sarampalis A, Loui P (2008) Auditory attention and filters. In: Yost WA, Popper AN, Fay RR (eds) Auditory perception of sound sources. Springer, New York, pp 115–142

Intriligator J, Cavanagh P (2001) The spatial resolution of visual attention. Cogn Psychol 43:171–216

Kidd G, Mason CR, Richards VM, Gallun FJ, Durlach NI (2008) Informational masking. In: Yost WA, Popper AN, Fay RR (eds) Auditory perception of sound sources. Springer, New York, pp 143–189

Kubovy M, van Valkenburg D (2001) Auditory and visual objects. Cognition 80:97–126

Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. Nature 390:279–281

Macmillan NA, Creelman CD (1991) Detection theory: a user's guide. Cambridge University Press, Cambridge

Massaro DW, Loftus GR (1996) Sensory and perceptual storage. In: Bjork EL, Bjork RA (eds) Memory. Academic Press, San Diego, pp 67–99

McKenna TM, Weinberger NM, Diamond DM (1989) Responses of single auditory cortical neurons to tone sequences. Brain Res 481:142–153

Micheyl C, Kaernbach C, Demany L (2008) An evaluation of psychophysical models of auditory change perception. Psychol Rev 115:1069–1083

Moore BCJ (2004) An introduction to the psychology of hearing. Elsevier, Amsterdam

Muckli L, Kohler A, Kriegeskorte N, Singer W (2005) Primary visual cortex activity along the apparent-motion trace reflects illusory perception. PLoS Biol 3:e265

Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) 'Primitive intelligence' in the auditory cortex. Trends Neurosci 24:283–288

Newsome WT, Mikami A, Wurtz RH (1986) Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion. J Neurophysiol 55:1340–1351

O'Regan JK, Noë A (2001) A sensorimotor account of vision and visual consciousness. Behav Brain Sci 24:939–1011

O'Regan JK, Rensink RA, Clark JJ (1999) Change-blindness as a result of 'mudsplashes'. Nature 398:34

Pashler H (1988) Familiarity and visual change detection. Percept Psychophys 44:369–378

Pavani F, Turatto M (2008) Change perception in complex auditory scenes. Percept Psychophys 70:619–629

Phillips WA (1974) On the distinction between sensory storage and short-term visual memory. Percept Psychophys 16:283–290

Phillips WA, Singer W (1974a) Function and interaction of On and Off transients in vision. I. Psychophysics. Exp Brain Res 19:493–506

Phillips WA, Singer W (1974b) Function and interaction of On and Off transients in vision. II. Neurophysiology. Exp Brain Res 19:507–521

Rensink RA (2002) Change detection. Ann Rev Psychol 53:245–277

Rensink RA, O'Regan JK, Clark JJ (1997) To see or not to see: the need for attention to perceive changes in scenes. Psychol Sci 8:368–373

Scholl BJ (2001) Objects and attention: the state of the art. Cognition 80:1–46

Shamma SA (2001) On the role of space and time in auditory processing. Trends Cog Sci 5:340–348

Stelmach LB, Bourassa CM, Di Lollo V (1984) Detection of stimulus change: the hypothetical roles of visual transient responses. Percept Psychophys 35:245–255

Theeuwes J, Kramer AF, Hahn S, Irwin DE (1998) Our eyes do not always go where we want them to go: capture of the eyes by new objects. Psychol Sci 9:379–385

Tiitinen H, May P, Reinikainen K, Näätänen R (1994) Attentive novelty detection in humans is governed by pre-attentive sensory memory. Nature 372:90–92

Visscher KM, Kaplan E, Kahana MJ, Sekuler R (2007) Auditory short-term memory behaves like visual short-term memory. PloS Biol 5:e56

Vitevitch MS (2003) Change deafness: the inability to detect changes between two voices. J Exp Psychol Hum Percept Perform 29:333–342

Wilken P, Ma WJ (2004) A detection theory account of change detection. J Vis 4:1120–1135

Yantis S, Hillstrom AP (1994) Stimulus-driven attentional capture: evidence from equiluminant visual objects. J Exp Psychol Hum Percept Perform 20:95–107

Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. Nature 453:233–235