Research report

The role of multisensory memories in unisensory object discrimination

Sandra Lehmann[a], Micah M. Murray[a,b,*]

[a]Division Autonome de Neuropsychologie, The Functional Electrical Neuroimaging Laboratory,
Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland
[b]Service Radiodiagnostique et Radiologie Interventionnelle, The Functional Electrical Neuroimaging Laboratory,
Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

Abstract

Past multisensory experiences can influence current unisensory processing and memory performance. Repeated images are better discriminated if initially presented as auditory–visual pairs, rather than only visually. An experience's context thus plays a role in how well repetitions of certain aspects are later recognized. Here, we investigated factors during the initial multisensory experience that are essential for generating improved memory performance. Subjects discriminated repeated versus initial image presentations intermixed within a continuous recognition task. Half of initial presentations were multisensory, and all repetitions were only visual. Experiment 1 examined whether purely episodic multisensory information suffices for enhancing later discrimination performance by pairing visual objects with either tones or vibrations. We could therefore also assess whether effects can be elicited with different sensory pairings. Experiment 2 examined semantic context by manipulating the congruence between auditory and visual object stimuli within blocks of trials. Relative to images only encountered visually, accuracy in discriminating image repetitions was significantly impaired by auditory–visual, yet unaffected by somatosensory–visual multisensory memory traces. By contrast, this accuracy was selectively enhanced for visual stimuli with semantically congruent multisensory pasts and unchanged for those with semantically incongruent multisensory pasts. The collective results reveal opposing effects of purely episodic versus semantic information from auditory–visual multisensory events. Nonetheless, both types of multisensory memory traces are accessible for processing incoming stimuli and indeed result in distinct visual object processing, leading to either impaired or enhanced performance relative to unisensory memory traces. We discuss these results as supporting a model of object-based multisensory interactions.
© 2005 Elsevier B.V. All rights reserved.

1. Introduction

Investigations of memories' or past experiences' influence(s) on the treatment of incoming stimuli have predominantly focused on unisensory memories (e.g., Ref. [27]).

However, multisensory experiences are believed to enrich our memories and influence ongoing sensory processes. Recent studies using hemodynamic measures (fMRI and PET) have examined how experiences in one or multiple senses alter later processing of stimuli in another sensory modality, providing evidence that brain regions involved in the encoding of an experience are also involved during its subsequent retrieval [19,33]. Intracranial microelectrode recordings in monkeys provide similar evidence by demonstrating that neuronal responses in visual object recognition areas are selective for multisensory-learned associations [10].

* Corresponding author. Division Autonome de Neuropsychologie, The Functional Electrical Neuroimaging Laboratory, CHUV, Hôpital Nestlé, 5 Avenue Pierre Decker, 1011 Lausanne, Switzerland. Fax: +41 21 314 1319.
 E-mail address: micah.murray@hospvd.ch (M.M. Murray).
 URL: www.chuv.ch/neuropsy.

These collective data show that neurophysiological responses, both within an area as well as across a brain network, to an incoming unisensory stimulus can vary according to whether it is part of a multisensory or unisensory memory.

Our recent behavioral and electrical neuroimaging study investigated the discrimination between unisensory (visual) and multisensory (auditory–visual) memories, providing data on when and where these effects first take place. While performing a continuous recognition task that required the differentiation of newly and already viewed images, subjects incidentally discriminated the repeated presentations of images according to their prior presentation as either a visual stimulus or auditory–visual (AV) pair. Stimuli with multisensory pasts were more accurately discriminated as having already been seen. This effect was observed in the absence of explicit studying of the auditory–visual pairs. Moreover, this discrimination was present electrophysiologically at just 60 ms and manifested as a change in the active areas of the brain within the lateral–occipital complex [17]. This was taken as evidence of the distinct representation of unisensory and multisensory events. As such, this study demonstrated that the functional consequences of the variations in cerebral activity following multisensory memory representations can be observed both electrophysiologically and, critically for the present study, behaviorally. Unresolved, however, are the kinds of memory traces that support this later discrimination, as well as whether all combinations of multisensory stimuli would suffice.

The aim of the present investigation was to determine what kinds of multisensory experiences are required to produce distinct perceptual/memory traces that can later be differentially retrieved upon repetition of the visual component. The use of meaningful auditory–visual stimulus pairs that always corresponded semantically across sensory modalities obfuscated our ability to address the requisites for establishing distinct perceptual/memory representations. One possibility is that the mere simultaneous presentation (i.e., a purely episodic context) of any auditory stimulus with visual objects would suffice. In which case, one would anticipate similar performance benefits irrespective of the nature of the multisensory experience. A parallel issue concerns whether or not equally effective, distinct perceptual/memory representations result from somatosensory–visual events. Both of these examples address the more general question of whether episodic multisensory experiences, which are orthogonal to the required task, cannot only result in perceptual/memory traces distinct from those for unisensory experiences, but also be later accessible upon presentation of just the visual component. Experiment 1 therefore examined the efficacy of episodic memory traces and whether different sensory combinations are equally effective by pairing visual object stimuli with either pure tones or somatosensory vibrations on distinct blocks of trials. A second (non-exclusive) possibility is that distinct perceptual/memory traces are established only after extensive semantic processing. In this case, performance would only be improved if stimuli presented to the different senses imparted information about the same object. That is, the above-mentioned distinct perceptual/memory traces would be for specific objects, rather than general visual experiences. Experiment 2 tested this by manipulating the congruence between auditory and visual object stimuli within blocks of trials.

2. Materials and methods

2.1. Participants

Experiment 1 included 16 (9 female) volunteers aged 21–31 years (mean ± SD = 26.7 ± 0.8). A different cohort of 11 subjects (8 female), aged 23–32 years (mean ± SD = 25.6 ± 0.9) participated in Experiment 2. All subjects provided written, informed consent to participate in the study, the procedures of which were approved by the Ethical Committee of the University of Lausanne. Of the 27 subjects, 24 were right-handed [20]. No subject had a history of or current neurological or psychiatric illnesses, and all reported normal or corrected-to-normal vision as well as normal hearing and touch.

2.2. Procedures

Both experiments had subjects perform a continuous recognition task comprised of equal numbers of initial and repeated presentations of line drawings that were pseudo-randomized with a block of trials (see also Ref. [17]). The line drawings were of common objects selected from either a standardized set [28] or obtained from an online library (dgl.microsoft.com) and modified to stylistically resemble those from the standardized set. The selected images contained an equivalent number of objects from different semantic categories (e.g., animals, tools, musical instruments, vehicles, miscellaneous household items, etc.) that were equally subdivided across experimental conditions. Images appeared black on a white background and were positioned centrally on a computer monitor (Philips Brilliance 202P4) located 114 cm from the subject. Images subtended an average of ~4.5° (±1.2°) in both the vertical and horizontal planes. On initial presentations, visual stimuli could either be presented with or without a sound or somatosensory vibration (the details of which are provided below).

The image set was equally divided into two groups: those that upon initial presentation appeared only visually and those that appeared as a simultaneous multisensory pair. On repeated presentations, only the visual stimuli from the initial presentations were displayed. Subjects' task was to indicate as quickly and as accurately as possible whether the image was being presented for the initial or repeated time. Thus, there were two classes of repeated presentations: (1)

those having initially been presented exclusively visually and (2) those having initially been presented as multisensory pairs. Subjects were not asked to make this distinction, so that the context (i.e., whether the initial encounter with the image during the experiments was unisensory or multisensory) was completely orthogonal rather than integral to the task. The timing of trials was such that stimuli were presented for 500 ms, followed by a 1200- to 1500-ms randomized inter-stimulus interval (ISI). Stimulus delivery and the recording of behavioral data (reaction time and accuracy) were controlled by E-prime in conjunction with their serial response box (www.pstnet.com; Psychology

Software Tools). The upper panels of Figs. 1 and 2 show schematics of the experimental paradigms.

Two variations of this task were used. Experiment 1 examined the effect of episodic multisensory experiences with images on subsequent visual repetition discrimination performance. These multisensory experiences were induced with either pure auditory tones (1000 Hz sinusoid; 10 ms rise/fall; 16 bit stereo; 44100 Hz digitization) or somatosensory vibrations (50 Hz square wave; 10 ms rise/fall; 44100 Hz digitization) on distinct blocks of trials. Both the auditory and somatosensory stimuli were controlled by E-prime via a Creative SoundBlaster card (www.creative.com).
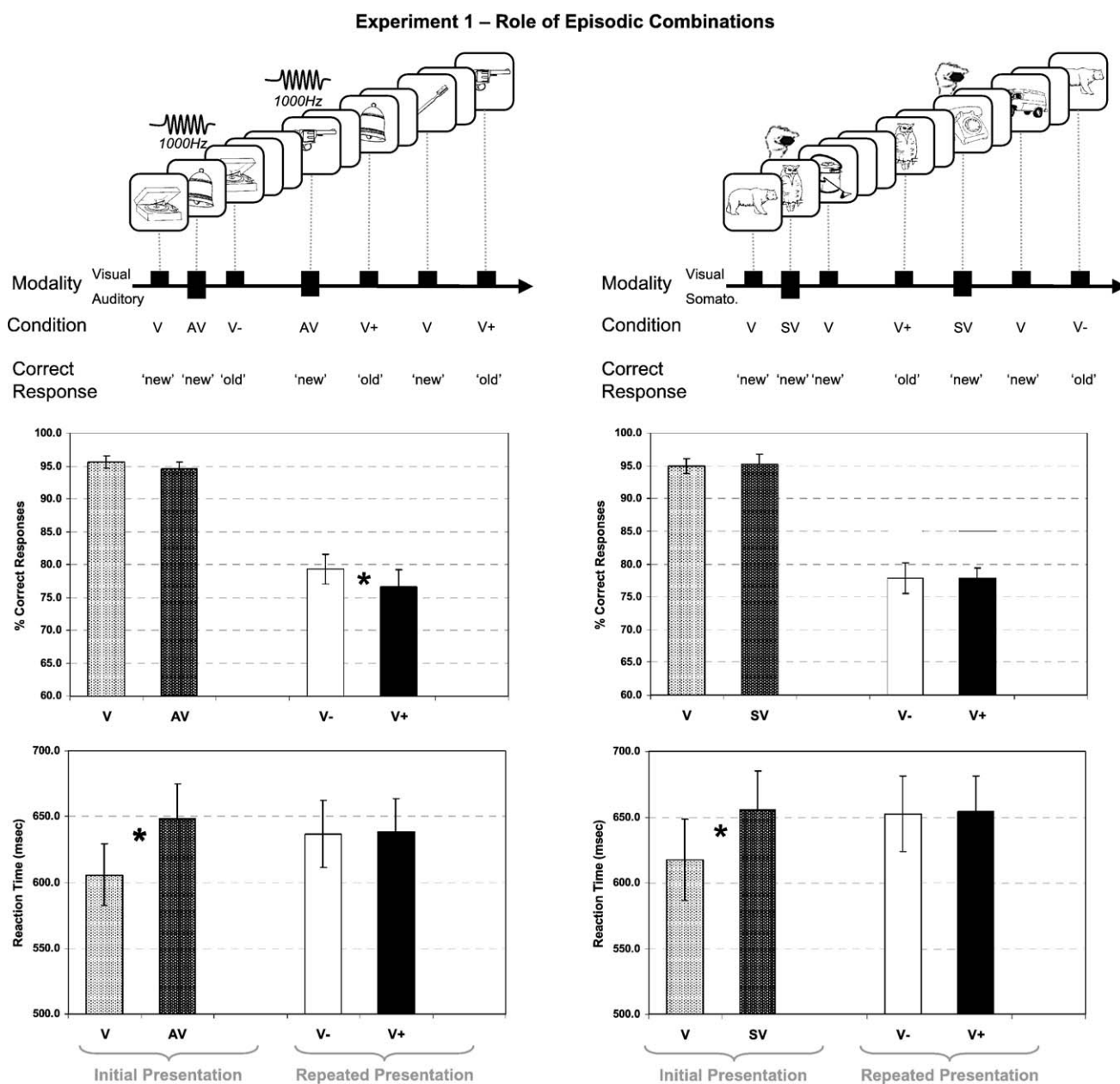

Fig. 1. Paradigm and behavioral results from Experiment 1 with auditory–visual and somatosensory–visual episodic pairings (left and right panels, respectively).
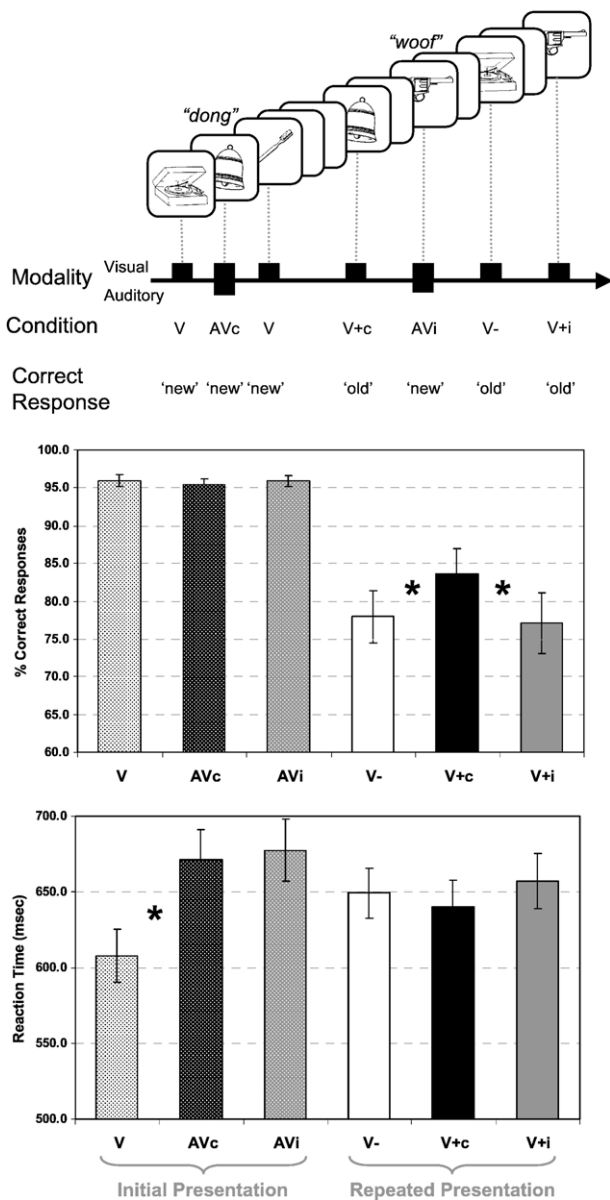
Experiment 2 – Role of Semantic Congruence

Fig. 2. Paradigm and behavioral results from Experiment 2 with auditory–visual semantic congruent and incongruent pairings.

The order of these blocks was counterbalanced across subjects (hereafter 'auditory blocks' and 'somatosensory blocks', respectively).[1] For auditory blocks, the initial presentations of images were subdivided into 2 equally-sized groups: those appearing only visually (V) and those simultaneously appearing with a pure tone (AV). Repeated presentations were comprised of only images. We refer to repetitions of images from the V condition as V− and to repetitions of the AV condition as V+. For somatosensory

blocks, the initial presentations of images were subdivided into 2 equally-sized groups: those appearing only visually (V) and those simultaneously appearing with a somatosensory pulse (SV). Repeated presentations were comprised of only images. We refer to repetitions of images from the V condition as V− and to repetitions of the SV condition as V+.

Experiment 2 examined the effect of semantic memories and auditory-visual congruency on subsequent visual repetition discrimination performance. To do this, the initial presentations of images were subdivided into 3 groups: those appearing only visually (V) and comprising 50% of the initial presentations, those simultaneously appearing with a semantically congruent sound (AV$_c$) and comprising 25% of the initial presentations, and those simultaneously appearing with a semantically incongruent sound (AV$_i$) and comprising the remaining 25% of initial presentations. In this way, there were equal numbers of unisensory and multisensory initial presentations. Auditory stimuli were complex, meaningful sounds (16 bit stereo; 44100 Hz digitization; 500 ms duration). Sounds were obtained from an online library (dgl.microsoft.com) and modified using audio editing software (Adobe Audition version 1.0) so as to be 500 ms in duration. As in Experiment 1, on repeated presentations, only the visual stimuli from initial presentations were displayed. We refer to repetitions of the images from the V condition as V−, to repetitions of images from the AV$_c$ condition as V+$_c$, and to repetitions of images from the AV$_i$ condition as V+$_i$.

For both experiments, stimuli were blocked into series of 136 trials, with equal likelihood of initial and repeated presentations. During a block of trials, each image was repeated once, independently of how the image was initially presented. Within a block of trials, the conditions were pseudo-randomized. The average number of trials between the initial and repeated presentation of any given stimulus was 13 images (range = 5–21 images). In Experiment 1, each subject completed 4 blocks of trials; two of which involved auditory–visual pairings and the other two of which involved somatosensory–visual pairing. The order of blocks was counterbalanced across subjects. In Experiment 2, each subject completed 4 blocks of trials. For blocks involving auditory stimuli, sounds were delivered through stereo speakers located on each side of the computer monitor. Each subject reported that the volume was at a comfortable and comprehensible level that was not disturbing during the task. For blocks involving somatosensory stimuli, an Oticon-A 100Ω bone conduction vibrator (Oticon, Somerset, NJ) with a 1.6 cm × 2.4 cm surface was held between the thumb and index finger of the left hand. It was held away from the knuckles to prevent bone conduction of sound. Likewise, on blocks of trials involving somatosensory stimuli, continuous white noise (the volume of which was adjusted to a comfortable level for each subject) was presented to mask any sounds made by the vibrator. Additionally, subjects placed their hand on a table

---

[1] Analyses of data from auditory and somatosensory blocks were conducted separately since there is no unequivocal means of equilibrating tones and vibrations.

before them so as to be in the same direction as the monitor, but without obstructing their view of the images. This resulted in approximate spatial alignment between visual and somatosensory stimuli. No object was repeated more than once for any subject—that is, each experiment was comprised of distinct stimuli. Likewise, sounds/images used for the incongruent condition were neither previously nor later used for other conditions.

It is also important at this point to comment on the distribution of multisensory and unisensory conditions as well as old/new image presentations within a block of trials. In the case of the auditory blocks of Experiment 1, the first quarter of trials contained 7 and 8 multisensory events, and the final quarter of trials contained 8 and 9 multisensory events (1st and 2nd block, respectively). Similarly, the first quarter of trials contained 18 new images, and the final quarter 15 new images. In the case of somatosensory blocks of Experiment 1, these values were 9 and 8 multisensory events for the first quarter and 11 and 5 multisensory events for the final quarter of trials. Similarly, the first quarter contained 18 new images, and the final quarter 17 and 13 new images (1st and 2nd block, respectively). In the case of Experiment 2, there was an average of 8.0 multisensory events during the first quarter of trials and 8.3 multisensory events during the final quarter of trials across the four blocks. There was an average of 18.0 new images during the first quarter and 15.0 new images during the final quarter. Thus, neither experiment had a clear bias in the distribution of multisensory versus unisensory events nor in terms of old/new images that would readily explain performance differences between images with multisensory versus unisensory pasts.

3. Results

Mean ($\pm$SEM) reaction time and accuracy data for each condition are shown in Fig. 1 for Experiment 1 and Fig. 2 for Experiment 2. Precise values are presented here in the text.

Analysis of performance data from initial stimulus presentations revealed a homogenous pattern of results. First, accuracy did not significantly differ across conditions on either experiment. Subjects performed at ~95% correct, with no significant differences between unisensory and multisensory conditions on Experiment 1 (auditory blocks: 95.6 $\pm$ 0.9% vs. 94.6 $\pm$ 1.1%; $t_{(15)} = 1.136$; $P = 0.274$, somatosensory blocks: 95.0 $\pm$ 1.1% vs. 95.3 $\pm$ 1.5%; $t_{(15)} = 0.303$; $P = 0.766$) and no main effect of condition on Experiment 2 (mean accuracy = 95.9 $\pm$ 0.8%, 95.4 $\pm$ 0.8%, 95.9 $\pm$ 0.7% for the V, AV_c, and AV_i conditions, respectively; $F_{(2,9)} = 0.177$; $P = 0.841$). Second, reaction times for multisensory stimulus presentation were slower than for their unisensory counterparts, with significant differences observed on Experiment 1 (auditory blocks: 606 $\pm$ 24 ms vs. 649 $\pm$ 26 ms; $t_{(15)} = 5.282$; $P < 0.001$, somatosensory blocks:

617 $\pm$ 31 ms vs. 655 $\pm$ 30 ms; $t_{(15)} = 4.305$; $P < 0.001$) and a significant main effect of condition on Experiment 2 ($F_{(2,9)} = 16.387$; $P < 0.001$). Follow-up comparisons showed that reaction times were significantly faster for the visual unisensory condition (V; 608 $\pm$ 18 ms) than either the semantically congruent (AV_c; 671 $\pm$ 20 ms) or incongruent multisensory condition (AV_i; 677 $\pm$ 21 ms) (V vs. AV_c: $t_{(10)} = 5.726$; $P < 0.001$; V vs. AV_i: $t_{(10)} = 5.935$; $P < 0.001$), with the latter two conditions not significantly differing (AV_c vs. AV_i: $t_{(10)} = 1.083$; $P = 0.304$).

Several aspects of the results from initial stimulus presentations during each experiment are worth brief comment here and will be elaborated in Discussion. First, slower reaction times on multisensory trials provide an indication that subjects were unaware of the fact that only initial presentations included sounds or vibrations (which was confirmed in follow-up debriefings). If this were not the case, faster and/or more accurate performance would be expected for any trial containing a non-visual component, since such information reliability indicated a new image. This supports our contention that any effects of these multisensory pairings on repeated image presentations are incidental, in addition to this aspect being orthogonal to the task. Second, performance patterns on initial presentations were not mirrored on repeated presentations.

Analysis of performance data from repeated stimulus presentations revealed a less homogenous pattern of results. On Experiment 1, accuracy was significantly worse for images that had been presented with a pure tone than for images presented only visually during the same blocks of trials (V+ versus V−) (auditory blocks: 76.6 $\pm$ 2.6% vs. 79.3 $\pm$ 2.2%; $t_{(15)} = -2.235$; $P = 0.041$). By contrast, there was no performance difference between V+ and V− conditions when the same subjects completed blocks of trials including somatosensory–visual pairings (77.8 $\pm$ 1.5% vs. 77.8 $\pm$ 2.3%; $t_{(15)} = 0.004$; $P = 0.997$). Thus, the effect of episodic multisensory memories on later unisensory discrimination would appear to be limited to auditory–visual experiences. On Experiment 2, the ANOVA conducted with the accuracy data from the V−, V+$_c$, and V+$_i$ conditions (78.0 $\pm$ 3.5%, 83.6 $\pm$ 3.3%, and 77.1 $\pm$ 4.1%, respectively) yielded a main effect of condition ($F_{(2,9)} = 23.950$; $P < 0.001$). Follow-up comparisons revealed that discrimination accuracy was improved for repetitions of images that had been presented with semantically congruent sounds versus images that had been initially presented only visually (V+$_c$ vs. V−: $t_{(10)} = 4.013$; $P = 0.002$), as well as images that had initially been presented with a semantically incongruent sound (V+$_c$ vs. V+$_i$: $t_{(10)} = 5.036$; $P = 0.001$). Accuracy from these latter two conditions did not significantly differ (V− vs. V+$_i$: $t_{(10)} = 0.413$; $P = 0.688$). In contrast to these effects on performance accuracy, reaction times did not significantly differ between conditions, either on Experiment 1 (auditory blocks: 637 $\pm$ 26 ms vs. 638 $\pm$ 26 ms; $t_{(15)} = 0.311$; $P = 0.760$, somatosensory blocks: 652 $\pm$ 29 ms vs. 654 $\pm$ 27 ms;

$t_{(15)} = 0.383$; $P = 0.707$) or on Experiment 2 where the main effect of condition did not reach our significance criterion (mean reaction times = 649 ± 16 ms, 640 ± 17 ms, 657 ± 18 ms for the V−, V+$_c$, and V+$_i$ conditions, respectively; $F_{(2,9)} = 2.855$; $P = 0.110$). It is important to note that this latter trend towards a significant main effect of condition was due to the tendency of reaction times from the V+$_c$ condition to be faster than either other condition. This parallels the pattern seen for accuracy measures and suggests that the significant improvement of V+$_c$ trials does not follow from a speed/accuracy trade-off.

4. Discussion

The principal finding of this study is that past multisensory experiences can influence the ability to accurately judge image repetitions during a continuous recognition task. This discrimination was according to past multisensory versus unisensory experiences, during the task itself, and was influenced by both episodic and semantic auditory–visual memory traces. Specifically, accuracy in indicating image repetitions (1) was significantly impaired for those images that had been presented with a 1000-Hz tone, (2) was not significantly affected for those images that had been presented with a 50-Hz vibration, (3) was not significantly affected for those images that had initially been presented with a semantically incongruent sound, and (4) selectively improved for images initially presented with a semantically congruent sound. Such performance changes were relative to repetition discrimination accuracy with those images initially presented only visually. These effects replicate and extend our previous work [17] to provide further indications concerning the necessary conditions for multisensory perceptual/memory traces to be established and later accessed upon the repeated presentation of unisensory visual stimuli. The collective results reveal opposing effects of episodic and semantic contexts from auditory–visual multisensory events. Additionally, the effect of episodic multisensory contexts on later unisensory discrimination would appear to be limited to auditory–visual experiences (at least for the stimulus parameters applied in the present study). Clearly, further experimentation will be required to fully assess whether initial pairings between visual and somatosensory stimuli alter later discrimination of repeated visual stimuli. Nonetheless, both unisensory visual and multisensory auditory–visual traces are accessible for processing incoming stimuli and indeed result in distinct visual object processing. We discuss these results as supporting the hypothesis that distinct perceptual/memory traces are established within object-based multisensory integration centers.

In a prior study, we proposed that the discrimination of the repetitions of visual items according to whether they were initially encountered in a multisensory or unisensory context reflects the rapid access to and influence from multisensory perceptual/memory traces [17]. The use of electrical neuroimaging methods in that study allowed us to identify the timing and likely neurophysiological basis for this initial discrimination, which onset at 60 ms in the form of a distinct source configuration within regions of the lateral–occipital complex (LOC). In addition to its role in visual object-related processing [14], the LOC and neighboring visual cortices have been repeatedly implicated in auditory–visual multisensory convergence and interactions [3,5,9,15], including higher-order multisensory object and linguistic processes [1,2,4,9,16,21,24,31]. From such, we concluded that the processing of visual stimuli already has access to perceptual/memory traces during the initial stages of sensory responses, and that these traces are within object-related multisensory integration areas. Whether the presently observed effects occur at a similarly early latency within the LOC will require further investigation with electrical neuroimaging techniques. Such notwithstanding, the present data do demonstrate that the capacity for perceptual/memory traces to serve as a basis for the differentiation of repeated unisensory stimuli is influenced by both the presence of simple auditory stimuli, as well as by the semantic congruence of auditory–visual multisensory object pairings, albeit in qualitatively different manners.

We propose that these traces reflect the consequences of object-based multisensory interactions, which may be distinct from interactions observed between more rudimentary stimuli. By the latter, we refer to interactive, non-linear effects observed between sensory stimuli (e.g., visual flashes, auditory beeps, and somatosensory pulses), often during conditions of passive presentation or simple detection [7,8,15,18]. These varieties of multisensory interactions are thought to depend on overlapping spatial receptive fields and temporal response profiles (reviewed in [30]). In contrast to these effects, we would propose that object-based multisensory interactions are particularly sensitive to the identity and semantic attributes of stimuli [22], which in the present study are principally conveyed within the visual modality. In support of this distinction, recent studies show sensitivity to the semantic attributes of stimuli – that is, whether information conveyed to the different senses derive from the same object – in generating multisensory effects. Some have shown that brain responses modulate in strength as a function of such semantic congruence [2,4,12,13,16]. Others show that the timing of effects varies with semantic congruence [6,16]. Here, we propose the following model to account for the establishment of object-based perceptual/memory traces that yield the pattern of effects observed in the present study.

This proposition is predicated on the notion of distributed object representations [11,25] and the hypothesis that semantic influences from the different senses modulate the fidelity with which these object representations can be activated. In the case of semantically congruent auditory–visual objects, distinct perceptual/memory traces can be

established that can be rapidly reactivated upon subsequent presentation of the visual component. This may arise through the enhanced activation of a singular object representation (e.g., 'cat') via multiple, redundant sources that in turn effectively yields a higher signal-to-noise ratio (in object terms) when the object system is confronted with repetition of just the visual component. That this reactivation has been observed to begin just 60 ms post-stimulus onset during the processing of image repetitions provides one indication of the particular fidelity that such perceptual/memory traces may have and that access to /influences from these traces need not be restricted to later, higher-order processing stages [17]. In the case of semantically incongruent pairs, no such enhanced trace is established relative to that established under unisensory conditions (and by extension no behavioral effect is observed). That is, the visual stimulus activates one object representation and the sound an altogether other. However, since multiple objects are routinely treated in parallel (i.e., visual scenes seldom include solitary objects; e.g., Ref. [26]) and since multiple objects can be simultaneously encoded via distributed neuronal representations [25], no diminution in performance is observed relative to the visual-only condition. In the case of images paired with tones, we would propose that the association of a single sound with multiple visual objects over trials effectively leads to the introduction of 'noise' into the establishment of distinct multisensory perceptual/memory traces for individual objects. That is, the same sound has produced an interactive response with several different object representations. Consequently, the fidelity of perceptual/memory traces for these objects is diminished and comparatively impaired performance is obtained. Finally, that no behavioral effects were observed when images were paired with vibrations (save for during initial stimulus presentations, discussed below) would suggest that the present results might be selective for the auditory and visual modalities; though further investigations are required to bolster this proposition.

It is similarly noteworthy that the present study dissociated effects observed on initial stimulus presentations from those observed on stimulus repetitions. That is, behavioral effects of multisensory experiences on encoding and retrieval appear to be dissociable. This impacts the plausibility of several alternative explanations of our results. In the case of Experiment 1, reaction times were significantly slower for the initial presentations of both auditory–visual and also somatosensory–visual multisensory stimulus pairs versus unisensory visual presentations. By contrast, it was only the repetitions of images that had been initially paired with sounds whose accuracy was significantly affected. In the case of Experiment 2, performance was indistinguishable for the initial presentations of semantically congruent and incongruent auditory–visual multisensory pairs, and reaction times to both were slower than those to initial presentations of visual stimuli. Yet, accuracy on stimulus repetitions was significantly improved only for the

case of images that had been part of a semantically congruent pair. Thus, it would appear that any differences in performance on the initial presentations of stimuli cannot readily account for performance differences on the repeated presentations of just the visual components. One implication is that multisensory influences in the determination of a visual stimulus' novelty are distinct from those in the determination of a visual stimulus' repetition, at least in the case of the present continuous recognition task.

More generally, multisensory information of *all* varieties during initial presentations led to slowed reaction times (relative to unisensory stimuli) in the absence of a significant alteration in the accuracy of novelty discrimination. This interference effect from multisensory conditions may be a consequence of the present task being restricted to the visual modality, which would at least partially explain differences between our results and those of others examining semantic congruence in multisensory object discrimination that required fuller processing of both visual and auditory signals [6,16]. By contrast, it was only performance with repeated images (in particular the accuracy of repetition discrimination) that was significantly affected by the episodic or semantic attributes of the initial multisensory experience. It is thus clear that the effects observed with stimulus repetitions do not simply reflect a general, long-lasting effect of multisensory interactions between auditory–visual or somatosensory–visual events.

An account of the present results in terms of selective attention (i.e., subjects focusing their attention to the auditory or somatosensory channel) or novel context detection [23,32] can therefore be discounted. Such accounts would have predicted faster and/or more accurate performance on initial multisensory presentations, particularly since the mere presence of non-visual information would have been a sufficient cue to indicate a novel image presentation. That is, on the basis of selectively attending to either audition or touch, subjects would have been able to more accurately and rapidly indicate an image's initial presentation (for multisensory versus unisensory trials). Such a pattern was not observed in either experiment. Additionally, in the case of Experiment 2 such accounts would have predicted equivalent performance for both the V+$_c$ and V+$_i$ conditions, since they are both equally "novel". Rather, this was not the case. A similar argument applies to an explanation in terms of general alerting, wherein multisensory events would have been predicted to produce the fastest and/or most accurate behavior. Rather, the pattern of reaction times on initial stimulus presentations fits well with results suggesting that events in an unexpected modality can lead to slowed reaction times on a detection paradigm [29]. However, this variety of selective attention still would not account for the performance pattern observed with repeated image presentations, particularly those of Experiment 2. In addition, effects of general arousal and fatigue cannot readily account for the present results. As detailed in Materials and methods, the different stimulus conditions were approximately homoge-

nously distributed throughout blocks of trials. Thus, even if subjects were more engaged in the task during the beginning of a block of trials, this would have applied equally to all stimulus conditions.

Further investigations are clearly required to more fully detail the (neurophysiological) basis of the observed effects. For example, it will be important to determine whether such effects are observed when the meaningless sound varies across trials (i.e., different pure tones for each visual object) or with non-objects (i.e., abstract designs or patterns). Likewise, it will be useful to determine whether the observed effects can similarly be elicited when the repeated images are either different exemplars of the same object or novel orientations (i.e., when the semantic attribute remains constant, but the physical features change), in order to better ascertain the precise representation accessed during retrieval. Nonetheless, irrespective of the direction of the change in performance accuracy, the present data suggest that unisensory percepts might trigger the multisensory representations associated with them, which can be formed without explicit study or active classification. This hypothesis will be the focus of future electrophysiological and hemodynamic imaging studies within our laboratory that examine the possibility put forth in prior investigations [10,17] that perceptual/memory traces for multisensory auditory–visual events involve distinct cortical representations relative to their unisensory counterparts. In particular, the present psychophysical data further suggest that the specificity of these representations may additionally depend on aspects such as semantic content and semantic congruency between the stimuli of multisensory pairs.

Acknowledgments

References

[1] A. Amedi, G. Jacobson, T. Hendler, R. Malach, E. Zohary, Convergence of visual and tactile shape processing in the human lateral occipital complex, Cereb. Cortex 12 (2002) 1202–1212.

[2] M.S. Beauchamp, K.E. Lee, B.D. Argall, A. Martin, Integration of auditory and visual information about objects in superior temporal sulcus, Neuron 41 (2004) 809–823.

[3] G.A. Calvert, Crossmodal processing in the human brain: insights from functional neuroimaging studies, Cereb. Cortex 11 (2001) 1110–1123.

[4] G.A. Calvert, R. Campbell, M.J. Brammer, Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex, Curr. Biol. 10 (2000) 649–657.

[5] A. Fort, C. Delpuech, J. Pernier, M.H. Giard, Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans, Cereb. Cortex 12 (2002) 1031–1039.

[6] A. Fort, C. Delpuech, J. Pernier, M.H. Giard, Early auditory–visual interaction in human cortex during nonredundant target identification, Cogn. Brain Res. 14 (2002) 20–30.

[7] J.J. Foxe, I.A. Morocz, M.M. Murray, B.A. Higgins, D.C. Javitt, C.E. Schroeder, Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping, Cogn. Brain Res. 10 (2000) 77–83.

[8] J.J. Foxe, G.R. Wylie, A. Martinez, C.E. Schroeder, D.C. Javitt, D. Guilfoyle, W. Ritter, M.M. Murray, Auditory–somatosensory multisensory processing in auditory association cortex: an fMRI study, J. Neurophysiol. 88 (2002) 540–543.

[9] M.H. Giard, F. Peronnet, Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study, J. Cogn. Neurosci. 11 (1999) 473–490.

[10] J.R. Gibson, J.H.R. Maunsell, Sensory modality specificity of neural activity related to memory in visual cortex, J. Neurophysiol. 78 (1997) 1263–1275.

[11] J.V. Haxby, M.I. Gobbini, M.L. Furey, A. Ishai, J.L. Schouten, P. Pietrini, Distributed and overlapping representations of faces and objects in ventral temporal cortex, Science 293 (2001) 2425–2430.

[12] P.J. Laurienti, M.T. Wallace, J.A. Maldjian, C.M. Susi, B.E. Stein, J.H. Burdette, Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices, Hum. Brain Mapp. 19 (2003) 213–223.

[13] P.J. Laurienti, R.A. Kraft, J.A. Maldjian, J.H. Burdette, M.T. Wallace, Semantic congruence is a critical factor in multisensory behavioral performance, Exp. Brain Res. 158 (2004) 405–414.

[14] R. Malach, J.B. Reppas, R.R. Benson, K.K. Kwong, H. Jiang, W.A. Kennedy, P.J. Ledden, T.J. Brady, B.R. Rosen, R.B. Tootell, Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 8135–8139.

[15] S. Molholm, W. Ritter, M.M. Murray, D.C. Javitt, C.E. Schroeder, J.J. Foxe, Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study, Cogn. Brain Res. 14 (2002) 115–128.

[16] S. Molholm, W. Ritter, D.C. Javitt, J.J. Foxe, Multisensory visual–auditory object recognition in humans: a high-density electrical mapping study, Cereb. Cortex 14 (2004) 452–465.

[17] M.M. Murray, C.M. Michel, R. Grave de Peralta Menendez, S. Ortigue, D. Brunet, S.L. Gonzalez Andino, A. Schnider, Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging, NeuroImage 21 (2004) 125–135.

[18] M.M. Murray, S. Molholm, C.M. Michel, D.J. Heslenfeld, W. Ritter, D.C. Javitt, C.E. Schroeder, J.J. Foxe, Grabbing your ear: rapid auditory–somatosensory multisensory interactions in low-level sensory cortices are not constraitoned by stimulus alignment, Cereb. Cortex (2004) (doi:10.1093/cercor/bhh197).

[19] L. Nyberg, R. Habib, A.R. McIntosh, E. Tulving, Reactivation of encoding-related brain activity during memory retrieval, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 11120–11124.

[20] R.C. Oldfield, The assessment and analysis of handedness: The Edinburgh Inventory, Neuropsychologia 9 (1971) 97–113.

[21] P. Pietrini, M.L. Furey, E. Ricciardi, M.I. Gobbini, W.H. Wu, L. Cohen, M. Guazzelli, J.V. Haxby, Beyond sensory images: object-based representation in the human ventral pathway, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 5658–5663.

[22] G. Pourtois G, B. de Gelder, Semantic factors influence multisensory pairing: a transcranial magnetic stimulation study, NeuroReport 13 (2002) 1567–1573.

[23] C. Ranganath, G. Rainer, Neural mechanisms for detecting and remembering novel events, Nat. Rev., Neurosci. 4 (2003) 193–202.

[24] T. Raij, K. Uutela, R. Hari, Audiovisual integration of letters in the human brain, Neuron 28 (2000) 617–625.

[25] E.T. Rolls, A. Treves, M.J. Tovee, The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex, Exp. Brain Res. 114 (1997) 149–162.

[26] G.A. Rousselet, M. Fabre-Thorpe, S.J. Thorpe, Parallel processing in high-level categorization of natural images, Nat. Neurosci. 5 (2002) 629–630.

[27] D.L. Schacter, Implicit memory: a new frontier for cognitive neuro-science, in: M. Gazzaniga (Ed.), The Cognitive Neurosciences, MIT Press, Cambridge, 1995, pp. 815–824.

[28] J.G. Snodgrass, M. Vanderwart, A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity, J. Exp. Psychol. Hum. Learn. 6 (1980) 174–215.

[29] C. Spence, M.E.R. Nicholls, J. Driver, The cost of expecting events in the wrong sensory modality, Percept. Psychophys. 63 (2001) 330–336.

[30] B.E. Stein, M.A. Meredith (Eds.), The Merging of the Senses, MIT Press, Cambridge, 1993.

[31] M.R. Stoesz, M. Zhang, V.D. Weisser, S.C. Prather, H. Mao, K. Sathian, Neural networks active during tactile form perception: common and differential activity during macrospatial and microspatial tasks, Int. J. Psychophysiol. 50 (2003) 41–49.

[32] D. Tsivilis, L.J. Otten, M.D. Rugg, Context effects on the neural correlates of recognition memory: an electrophysiological study, Neuron 31 (2001) 497–505.

[33] M.E. Wheeler, S.E. Petersen, R.L. Buckner, Memory's echo: vivid remembering reactivates sensory-specific cortex, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 11125–11129.