Title: Single-trial multisensory memories affect later auditory and visual object discrimination.

Authors: Thelen A, Talsma D, Murray MM

Journal: Cognition

Year: 2015 May

Issue: 138

Pages: 148-60

DOI: 10.1016/j.cognition.2015.02.003

Single-trial multisensory memories affect later auditory and visual object discrimination

Antonia Thelen[1, 5], Durk Talsma[2] & Micah M. Murray[1, 3 – 5]

[1]The Laboratory for Investigative Neurophysiology (The LINE), Department of Clinical Neurosciences and Department of Radiology, Vaudois University Hospital Center and University of Lausanne, Lausanne, Switzerland

[2] Department of Experimental Psychology, Ghent University, Ghent, Belgium

[3]Electroencephalography Brain Mapping Core, Center for Biomedical Imaging of Lausanne and Geneva, Switzerland

[4]The Center for Neuroscience Research (CRN), Department of Clinical Neurosciences, Vaudois University Hospital Center and University of Lausanne, Switzerland

[5]Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

*Address correspondence to:
Antonia Thelen
Department of Clinical Neurosciences
CHUV, BH07.081
Rue du Bugnon 46
1011 Lausanne, Switzerland
Thelen.Antonia@gmail.com

Abstract

Multisensory memory traces established via single-trial exposures can impact subsequent visual object recognition. This impact appears to depend on the meaningfulness of the initial multisensory pairing, implying that multisensory exposures establish distinct object representations that are accessible during later unisensory processing. Multisensory contexts may be particularly effective in influencing auditory discrimination, given the purportedly inferior recognition memory in this sensory modality. The possibility of this generalization and the equivalence of effects when memory discrimination was being performed in the visual versus auditory modality were at the focus of this study. First, we demonstrate that visual object discrimination is affected by the context of prior multisensory encounters, replicating and extending previous findings by controlling for the probability of multisensory contexts during initial as well as repeated object presentations. Second, we provide the first evidence that single-trial multisensory memories impact subsequent auditory object discrimination. Auditory object discrimination was enhanced when initial presentations entailed semantically congruent multisensory pairs and was impaired after semantically incongruent multisensory encounters, compared to sounds that had been encountered only in a unisensory manner. Third, the impact of single-trial multisensory memories upon unisensory object discrimination was greater when the task was performed in the auditory vs. visual modality. Fourth, there was no evidence for correlation between effects of past multisensory experiences on visual and auditory processing, suggestive of largely independent object processing mechanisms between modalities. We discuss these findings in terms of the conceptual short term memory (CSTM) model and predictive coding. Our results suggest differential recruitment and modulation of conceptual memory networks according to the sensory task at hand.

Keywords: *multisensory, auditory, visual, object recognition, implicit, memory*

Introduction

A substantial body of work suggests that multisensory interactions can already occur at early latencies and within primary or near-primary cortices (reviewed in Murray, et al., 2012; van Atteveldt et al., 2014). Moreover, these interactions have been correlated with behavior (Cappe, Thelen, Romei, Thut, & Murray, 2012; Romei, Murray, Merabet, & Thut, 2007; van den Brink et al., 2014; Van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011; Thelen, Matusz, & Murray 2014). Cappe et al. (2012) found that increases in neuronal response strength at early latencies were positively correlated with multisensory gains in a motion discrimination task. Similarly, Romei and colleagues (2007) found correlations between multisensory events and the impact of a TMS pulse delivered over the occipital pole on auditory detection response speed. In another study, van der Burg et al (2011) showed auditory facilitation effects in a visual search task modulating activity within parieto-occipital cortices. Following up on the latter results, van den Brink et al. (2014) found that this facilitation was predicted by the strength of anatomical connections between sub-cortical and cortical auditory structures.

While these and similar data reveal much about the instantaneous interactions between the senses, other studies have focused on how multisensory interactions taking place at one point in time have an impact on subsequent unisensory processing. For example, a large number of studies have investigated how unisensory stimulus discrimination and perceptual learning are affected by prior multisensory experiences (Shams & Seitz, 2008; Shams, Wozny, Kim, & Seitz, 2011; Gottfried, Smith, Rugg, & Dolan, 2004; Nyberg, Habib, McIntosh, & Tulving, 2000; von Kriegstein & Giraud, 2006; Wheeler, Petersen, & Buckner, 2000). Likewise, Meylan and Murray (2007) showed that occipital cortical activation, due to the processing of visual stimuli was significantly attenuated when these stimuli were preceded by a multisensory stimulus. Our group has therefore specifically focused on how multisensory contexts may exert their influences in a more implicit manner and via single-trial exposures (Lehmann & Murray, 2005; Murray, Foxe, & Wylie, 2005; Murray et al., 2004; Thelen, Cappe, & Murray, 2012; Murray & Sperdin, 2010; Thelen & Murray, 2013; Thelen, Matusz, & Murray, 2014). These studies show that visual object recognition is improved when the initial multisensory context had been semantically congruent and can be impaired if this context was either semantically incongruent or meaningless, when compared to recognition of visual stimuli only encountered in a unisensory visual context. More generally, these "single-trial" memories (i.e. memories that form after a single, initial pairing of a semantically congruent image and sound) of multisensory object associations are formed incidentally (i.e. parenthetically) and despite many intervening stimuli, are distinguishable from encoding processes, and promote distinct object representations that manifest as differentiable brain networks whose activity is correlated with recognition performance (Thelen & Murray, 2013).

Despite these advances in our understanding of multisensory memory and its impact on visual recognition, it is still not clear whether or not auditory object discrimination also benefits from (single-trial) multisensory memories. Some research would emphatically contend that auditory memory is grossly inferior to visual memory (Cohen, Horwitz, & Wolfe, 2009). Memory performance in a recognition task was impaired for sounds that had been paired with a corresponding image during the preceding study phase, as well as when the stimuli for the task were either speech stimuli or clips of music, which were considered to be richer in their content. The only situation wherein recognition memory for sounds was better than that for images was when the images were highly degraded. In terms of a putative explanation, Cohen et al. went so far as to suggest the following: "*…auditory memory might be fundamentally different/smaller than visual memory. We might simply lack the capacity to remember more than a few auditory objects, however memorable, when they are presented one after another in rapid succession.*" (p. 6010 of Cohen et al, 2009).

By this account, benefits of multisensory contexts on subsequent unisensory auditory discrimination may not be expected. If true, this would dramatically curtail potential applications of this paradigm to remediation or training situations; a central issue for the development of multisensory rehabilitation strategies across the lifespan (White-Traut et al., 2013; Johansson, 2012). By contrast, an alternative interpretation of the results of Cohen et al. (2009) may be warranted. This is based on an extension of the principle of inverse effectiveness (Stein & Meredith, 1993; Altieri, Stevenson, Wallace, & Wenger, 2013; Stevenson et al., 2014). This interpretation would instead suggest that greater benefits would be observed in the sensory modality wherein information is less effective in eliciting a given behavior. If memory is generally less efficient in the auditory modality, then relatively greater gains from multisensory contexts would be expected. In accordance, Yuval-Greenberg and Deouell (2009) observed that visual information has a greater impact on auditory object identification than vice-versa. Likewise, selective delay-period activity on a delayed match-to-sample task was observed in intracellular recordings from monkey infero-temporal neurons not only when the animal performed a visual-to-visual task, but also when it performed either a visual-to-auditory or auditory-to-visual task (Gibson & Maunsell, 1997). This kind of neural response provides an indication that memory representations can be formed across the senses, and can also be activated by input from either sense alone. Likewise, functional imaging in humans is increasingly documenting the involvement of visual cortices in the categorical processing of sounds either via predictive coding (Vetter et al., 2014) or multisensory learning (von Kriegstein & Giraud, 2006; see also Schall, Kiebel, Maess, & von Kriegstein, 2013; Sheffert & Olson, 2004).

It thus remains to be established 1) if auditory object discrimination is affected by single-trial multisensory memories and if so whether this is to the same degree as that observed in the visual modality, and 2) if there is a systematic relationship between memory performance in the visual and auditory modalities. Given these outstanding issues, the present study assessed the efficacy of multisensory exposures on auditory object discrimination during the completion of a continuous recognition task requiring the discrimination of initial from repeated sound object presentations. On the one hand, establishing such an effect will reveal whether or not auditory object processing has access to (and potentially benefits from) visual object representations, even when such information is task-irrelevant and occurred during initial object encoding. On the other hand and given the preponderance of auditory functional deficits following stroke (e.g. Griffiths, 2002), determining the ability of multisensory learning contexts to improve auditory memory functions in an incidental manner confers potential clinical applicability. By having the same set of participants also perform the task in the visual modality, we were able to compare the relative impact of single-trial and task-irrelevant multisensory contexts on subsequent unisensory memory functions (see also Cohen et al., 2009). This would reveal potential coupling and/or independence between the senses in terms of memory functions and by extension potential common resources.

Material and Methods

Participants

The experiment included 26 adults (6 men) aged 17 - 41 years (mean age±SD = 26±6.16 years). 24 subjects were right-handed, according to the Edinburgh Inventory (Oldfield 1971). No subject had a history of neurological or psychiatric illness, and all subjects had normal or corrected-to-normal vision and reported normal hearing. Subjects were either undergraduate students enrolled in psychology at the University of Lausanne (N=13), who received course credit in exchange or were unpaid volunteers (N=13). The study was conducted in accordance with the Declaration of Helsinki, and all subjects provided their informed consent to participate in the study. The experimental procedures were approved by the Ethics Committee of the Vaudois University Hospital Center and University of Lausanne.

Task

Subjects performed a continuous recognition task, which required them to discriminate whether an item had been presented for the first or second time during a block of trials. Task-relevant items were either line drawings or sounds of environmental objects. The image and sound discrimination tasks were presented in separate experimental sessions and the stimuli themselves were pseudo-randomized within a block of trials. The participants were instructed to perform as quickly and as accurately as possible. Furthermore, each object (irrespective of whether it was initially presented in a unisensory or multisensory context) was only repeated once throughout each experimental block (see Figure 1 for a schematic representation of the paradigm).

In both recognition tasks, half of the initial presentations were auditory-visual multisensory pairings, which were semantically congruent (24 initial presentations per block), incongruent (24 initial presentations per block) or meaningless (24 initial presentations per block), while the other half were unisensory presentations (72 initial presentations per block) (see Table 1). The design of the experiment was as follows. First, the overall probability of unisensory versus multisensory presentations was the same over all trials (P(multisensory)=P(unisensory)=0.5). Further, the probability of unisensory and multisensory presentations was equal for initial and repeated presentations. Consequently, whether an object was presented in a unisensory or multisensory manner was not predictive of whether it was an initial or a repeated condition. This aspect addresses a potential shortcoming of the paradigm used in our prior studies (see Thelen & Murray, 2013 for discussion).

Upon repetition half of the stimuli were identical to the initial presentation (36 trails of repeated unisensory stimuli; 12 previously unisensory presentations, which were repeated in a congruent, incongruent or meaningless multisensory context, respectively). Of the remaining stimuli, half of the previously multisensory stimuli were presented in a unisensory manner (12 trails for each previous encounter context). The remaining initially unisensory stimuli were paired with either a meaningful congruent, incongruent or meaningless sound (or image) where each variety of pairing was equally probable (12 trails for each previous encounter context) (see Table 1).

Stimuli

The line drawings were taken from a standardized set (Snodgrass & Vanderwart, 1980) or obtained from an online library (dgl.microsoft.com), and included a mix of living and non-living stimuli (see Appendix 1). Additionally, we created a series of not obviously meaningful (scrambled) pictures from the above line drawings with an in-house MATLAB script (www.mathworks.com). All pictures had the same dimensions (585 x 585 pixels), and were divided in 5 x 5 squares (117 x 117 pixels). Within each of these squares pixels were randomized, leading to the creation of meaningless and unrecognizable clouds of dots (see Figure 1b). This procedure ensured that differences found between meaningful and meaningless visual object processing were in fact due to object discrimination per se, rather than to differences due to low-level visual features (Knebel, Toepel, Hudry, le Coutre, & Murray, 2008).

The auditory objects were taken from a library of 500ms-duration sounds that have been extensively used by our laboratory and that have been evaluated with regard to their acoustics, psychoacoustics as well as brain responses as a function of semantic category. Briefly, these stimuli are readily recognized and are highly familiar (cf. Table 1 in Murray et al., 2006; see also De Lucia, Clarke, & Murray, 2010; De Lucia, Tzovara, Bernasconi, Spierer, & Murray, 2012; De Lucia, Cocchi, et al., 2010; Spierer et al., 2011; Murray et al., 2009) (see Appendix 2). Meaningless sounds were created with Adobe Audition 1.0 and were either pure tones or modulated sounds. Tones differed in their spectral composition, ranging from 100Hz to 4700Hz, and sounds were modulated in terms of amplitude envelopes and/or waveform types (triangular or sinusoid). All sounds, irrespective of whether they were meaningful or meaningless, were 500ms duration (10ms rise/fall, in order to avoid clicks; 16bit mono; 44100Hz digitization).

All stimuli were presented synchronously for 500ms, followed by a randomized inter-stimulus interval (ISI) ranging from 900 to 1500ms, and subjects had to respond within this 2s window. The mean (±SD) number of trials between the initial and the repeated presentation of the target object (either visual or auditory, respectively) was 9±4 intervening stimuli for all presentation conditions. Also, the distribution of old and new target stimuli throughout the length of the blocks was controlled, so as to avoid fatigue and response-decision bias. This type of bias refers to subjects being able to calculate predictive probabilities about the upcoming stimuli and responses, which could lead to faster reaction times and/or a drop in attention. The experiment took place in a sound-attenuated chamber, where subjects were seated centrally in front of a 20" computer monitor (HP LP2065), and located ~ 140 cm away from them (visual angle of objects ~ 4°). The auditory stimuli were presented over insert earphones (Etymotic model: ER4S), and the volume was adjusted to a comfortable level (~62dB). The stimuli were presented and controlled by E-Prime 2.0, and all behavioral data were recorded in conjunction with a serial response box (Psychology Software Tools, Inc.; www.pstnet.com).

Subjects performed both the auditory and the visual task on different days, separated by one week. The order of task completion was counterbalanced across subjects. The same stimuli were used in both sensory tasks (144 stimuli per experimental block, adding up to 288 trials per block), in order to directly compare performance accuracy across modalities to representations of the same objects.

We likewise directly tested for systematic performance differences as a function of the modality in which the task was first completed. There was no evidence for a systematic effect of task order, though there was evidence of an interaction between Task Order and Multisensory Context in terms of recognition accuracy upon repeated trials ($F_{(2, 23)}$=6.585; p=0.005; η_p^2=0.869). Post-hoc unpaired t-tests revealed that this interaction stemmed from a between-group difference for visual objects that had been presented with a congruent sound upon initial encounter (V+c Auditory vs. Visual first: 4.86 ± 1.6% vs. 0.2 ± 1.5%; $t_{(24)}$=2.148; p=0.042). No other significant differences were found. Although prior research focusing on cross-sensory semantic priming has shown cross sensory modality priming effects between vision and audition in object identification tasks (Schneider, Debener, Oostenveld, & Engel, 2008; Schneider, Engel, & Debener, 2008; Senkowski, Schneider, Tandler, & Engel, 2009; for an example of haptic to visual priming see Schneider, Lorenz, Senkowski, & Engel, 2011), these results were characterized by faster RTs for congruent auditory to visual and visual to auditory priming pairs as compared to incongruent pairings.

Data Analysis

Accuracy and RT data were computed for each condition for every subject over all blocks of trials. Subjects completed two visual blocks. Only trials where subjects responded within a 150-1500ms post-stimulus onset window were considered for the computation of accuracy rates. Similarly, only RT data of correct response trials were considered in the analyses. A pilot study indicated that subjects performed with a greater inter-block variability in the auditory task. Thus, to ensure that the task was understood and could be performed at a reliable level of accuracy, subjects completed three auditory blocks. Because there was no evidence for a learning effect across the auditory blocks, all three blocks were collapsed in the analyses. In order to directly compare performance between the visual and the auditory tasks, we computed the gain/cost index for each subject and for each condition. This index was calculated as the accuracy/RT difference for repeated presentations of repeated unisensory presentations. This resulted in a comparable measure of the impact of multisensory memory traces on subsequent auditory and visual object discrimination, avoiding the caveat of introducing differences due to general task-related performance differences. Gain/cost indices were calculated for all types of unisensory repetitions of prior multisensory contexts.

The general nomenclature for experimental conditions used throughout the remainder of the manuscript is the following. Unisensory repetitions of previously visual and auditory unisensory presentations are V- and A-, respectively. Unisensory repetitions of visual and auditory objects that had been initially presented in a multisensory context are V+ and A+, respectively. Moreover, we use the following subscripts to specify the nature of the prior multisensory context: *c* for semantically congruent pairings; *i* for semantically incongruent pairings; and *m* for otherwise meaningless pairings. Although the original design included multisensory repetitions of either previously unisensory or multisensory presentations, we here focus on the impact of multisensory memories upon subsequent unisensory retrieval.

Statistical Analyses

Analyses of the data were directed at responding to three specific research questions. First and in order to directly compare the impact of multisensory memory traces upon subsequent visual and auditory object discrimination, we submitted gain/cost indices from both sensory modalities to a 2x3 within-subject analysis of variance (ANOVA). One aspect of this analysis is that it addresses the proposal from Cohen et al. (2009) concerning the generally impoverished memory for sounds. Second, gain/cost indices were evaluated within each sensory modality (after first observing a significant interaction in the above 2x3 ANOVA). On the one hand this analysis would directly assess if auditory object discrimination is affected by prior single-trial multisensory contexts. On the other hand, this analysis allows for situating the present study with respect to our prior works (Lehmann & Murray, 2005; Murray et al., 2004; Murray, Foxe, & Wylie, 2005; Thelen, Cappe, & Murray, 2012). Third, we assessed correlations across sensory modalities in which the task was performed as well as correlations between performance on initial and repeated presentations. This would provide insights regarding common multisensory memory processes as well as carry-over effects from encoding to retrieval.

Data were analyzed with ANOVA. Post-hoc t-tests were then performed in the event of significant effects/interactions. Correction for multiple comparisons was done according to the Holm-Bonferroni method (Holm, 1979). Because we had a strong a priori hypothesis regarding the directionality of the effects due to previous investigations (Lehmann & Murray, 2005; Murray et al., 2004, 2005; Thelen et al., 2012) we applied one-tailed statistics to test for specific differences between multisensory pairings for the visual task modality. By contrast, 2-tailed statistical thresholds were used in the analysis of the auditory task modality. Lastly, we compared values to a zero matrix to determine if a given gain/cost significantly differed from zero.

Results

Gain/Cost Indices

The gain/cost index describes the relative percentage of accuracy enhancement or impairment for objects initially encountered in a multisensory vs. unisensory context, independently of general sensory modality related differences. These values were entered into a 2x3 repeated-measures ANOVA. There was no main effect of Task Modality (overall gain/cost ± s.e.m.: visual blocks = -1.44 ± 1.01%; vs. auditory blocks = -1.63 ± 1.37%; $F_{(1,25)}=0.021$; $p=0.885$; $\eta_p^2=0.001$), indicative of similar magnitudes of impacts of task-irrelevant stimuli on unisensory object discrimination. There was a main effect of Multisensory Context

($F_{(2,24)}$=40.507; p<0.001; η_p^2=0.771) and a significant interaction between the factors Task Modality and Multisensory Context ($F_{(2,24)}$=11.548; p<0.001; η_p^2=0.490)..

Given this interaction, additional ANOVAs were conducted. The task-specific one-way ANOVA on the gain/cost indices for the visual task revealed a significant effect of Multisensory Pairing ($F_{(2,24)}$=12.504; p<0.001; η_p^2=0.510) (Figure 2a). Post-hoc 1-tailed t-tests revealed that subjects showed a positive gain index for previously congruent presentations, compared to previously incongruent and meaningless presentations (V+c vs. V+i = 2.35 ± 1.16% vs. -3.9 ± 1.61%; $t_{(25)}$=4.555; p<0.001; V+c vs. V+m = 2.35 ± 1.16% vs. -2.77 ± 1.35; $t_{(25)}$=3.192; p=0.008). Gain/cost indices for previously incongruent and previously meaningless presentations did not reliably differ (V+i vs. V+m = -3.9 ± 1.61% vs. -2.77 ± 1.35; $t_{(25)}$=0.6; p=0.557).

The one-way ANOVA on the gain/cost indices from the auditory task revealed a significant effect of Multisensory Pairing ($F_{(2,24)}$=32.252; p<0.001; η_p^2=0.729) (Figure 2b). Post-hoc 2-tailed t-tests showed that previously congruent presentations led to a positive gain index and differed from both previously incongruent and previously meaningless presentations (A+c vs. A+i = 6.35 ± 1.95% vs. -11.15 ± 1.78%; $t_{(25)}$=8.054; p<0.001; A+c vs. A+m = 6.35 ± 1.95% vs. -0.09 ± 1.44; $t_{(25)}$=3.882; p=0.001). Moreover, indices for the A+i and A+m conditions also differed significantly (A+i vs. A+m = -11.15 ± 1.78% vs. -0.09 ± 1.44; $t_{(25)}$=-6.454; p<0.001).

In order to ensure that these gain/cost indices significantly differed from zero, we entered the gain/cost indices into independent one-tailed t-tests vs. a zero matrix. This analysis showed that gain/cost indices differed significantly from zero for all conditions in the visual task (V+c = 2.35 ± 1.16%, $t_{(25)}$=2.03, p=0.027; V+i = -3.9 ± 1.61%, $t_{(25)}$=-2.419, p=0.012; V+m = -2.77 ± 1.35% , $t_{(25)}$=-2.057, p=0.025), suggesting that visual object discrimination is generally affected by single-trial multisensory encounters (albeit in different directions). When the task was performed in the auditory modality, all gain/cost indices differed from zero except the A+m condition (A+c = 6.35 ± 1.95% , $t_{(25)}$=3.244, p=0.002; A+i = -11.15 ± 1.78% , $t_{(25)}$=-6.257, p<0.001; A+m = -0.09 ± 1.44% , $t_{(25)}$=-0.065, p=0.949).

After having investigated gain/cost indices for discrimination accuracy, we submitted the gain/cost indices of RTs into the same type of analyses (results are not shown). The modality-specific one-way ANOVAs as well as the 2x3 repeated-measures ANOVA between modalities did not reveal any significant effects, demonstrating that single-trial multisensory memories do not reliably impact the response speed of subsequent unisensory object discrimination.

Correlation Analysis

We tested whether there was a direct carry-over effect between initial encoding differences (V vs. AV/ A vs. VA) and differences between subjects' performance upon repeated trials (V- vs. V+/ A- vs. A+). Table 2b lists the correlation coefficients between the difference in response speed upon initial presentation and the difference in accurate discrimination upon repeated presentations. Generally, there was no evidence for such a carry-over effect[i].

Finally, we assessed whether subjects' performance in one modality was correlated with the performance accuracy in the other modality (A vs. V). The results suggest that there was no linear relationship either between initial and repeated presentations within a sensory modality or between modalities. Rather, response accuracies were only significantly correlated within modalities and only within presentation type (initial vs. repeated). More precisely, we found significant correlations ($-0.39>r_{(26)}>+0.39$) between response accuracy within initially presented visual objects in different Encounter Contexts (i.e. visual-only, or paired with either a congruent, incongruent or meaningless auditory sound) (V vs. AV: $r_{(26)}=0.67$; $p<0.001$;).

General memory effect

In order to address the hypothesis put forth by Cohen and colleagues (2009), stating that auditory memory is grossly inferior to visual memory, we entered the raw accuracy and RTs data into two separate 2x4 (Task Modality by Encounter Context) repeated measures ANOVAs (Table 2a).

The analyses on the raw accuracy data, revealed significant main effects of Task Modality ($F_{(1,25)}=74.268$; $p<0.001$; $\eta_p^2=1$) and Encounter Context ($F_{(1,25)}=26.22$; $p<0.001$; $\eta_p^2=1$). Furthermore, we also found a significant Task Modality by Encounter Context interaction ($F_{(3,23)}=7.424$; $p=0.001$; $\eta_p^2=0.966$). These findings concur with the findings of Cohen and colleagues (2009), suggesting that auditory memory for objects is generally inferior to visual object memory (Cohen, Horowitz, & Wolfe, 2009).

Similarly, in terms of RTs, we proceeded to a 2x4 (Task Modality by Encounter Context) repeated measures ANOVA, to test for sensory modality related differences (Yuval-Greenberg & Deouell, 2007, 2009). The analyses revealed a significant main effect of Task Modality

($F_{(1,25)}$=188.274; $p<0.001$; η_p^2=0.883) and of Encounter Context ($F_{(3,23)}$=3.037; $p=0.05$; η_p^2=0.284) (Table 2a), thus mirroring the findings on the raw accuracy data.

Discussion

The present study demonstrated that the discrimination of objects presented in an auditory manner is affected by prior, single-trial multisensory experiences. In what follows we discuss results of the auditory recognition task in light of our prior and present findings in the visual modality with a particular focus on the potential inter-independence of multisensory influences on visual and auditory object discrimination. Further, since similar patterns of performance were observed for unisensory visual and auditory object discrimination, we discuss the potential involvement of common memory processes, proposing how the present findings are compatible with a more general auditory-visual object association framework.

The primary finding of this study is that auditory object discrimination is differentially affected by prior multisensory contexts (Figure 2b). More precisely, recognition was enhanced for sounds presented with a congruent image upon initial encounter and impaired for sounds that had been presented with an incongruent image upon initial presentation. This was compared to sounds presented with a meaningless image in terms of gain/cost indices. The present data extend our previous findings concerning the visual modality to the auditory modality, namely that a single encounter with an auditory-visual pairing is sufficient to incidentally impact subsequent auditory object discrimination. Our work therefore constitutes a partial replication of the work of von Kriegstein and Giraud (2006), who investigated whether (auditory) speaker recognition could benefit from multisensory learning and whether benefits were linked to feature redundancy between the senses. They postulated that auditory object recognition can benefit from multisensory learning only when the sensory features carry information about one-and-the-same object (e.g. voice-face pairing of an individual). Interestingly, von Kriegstein and Giraud (2006) failed to find any impact of initial, arbitrary auditory-visual couplings upon subsequent auditory recognition (in terms of discrimination accuracy; see Table 2a). This discrepancy may be linked to the type of stimuli that were presented to subjects. While in our study, sounds belonged to a multitude of object categories (spanning living and man-made objects), von Kriegstein and Giraud (2006) investigated a rather particular sound category, namely speaker recognition, which entail unique voice-face associations. Recognition of these pairs was contrasted with voice-name associations, which are arbitrary in nature (many people carry the same name, but have a unique voice). Alternatively, the present study employed a battery of category-representative stimuli in both sensory modalities. Thus, while von Kriegstein and Giraud (2006) presented unique auditory-visual pairs, we presented pairs that are linked at a more general semantic level of object association. Consequently, the discrepancy between these findings could be explained in the light of prior evidence suggesting specialized processing mechanisms for faces and speech, which differ from other object processing mechanisms (O'Mahony & Newell, 2012; von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005). Consequently, the findings of von Kriegstein and Giraud may be limited to particular object category (voice and faces) and not readily generalizable to other categories.

The results of our visual recognition task showed that recognition was enhanced for images that had been paired with a congruent sound upon their initial encounter, whereas it was impaired for images that had been paired with an incongruent or a meaningless sound upon their initial encounter (Figure 2a). Consequently, we replicated our previous findings in visual object discrimination being incidentally affected by past multisensory encounters (Murray et al. 2004; 2005; Thelen et al. 2012). This further emphasizes that single-trial multisensory memories have a robust impact upon subsequent unisensory object discrimination. Additionally, the current study addressed some paradigmatic shortcomings in our prior work. Most importantly, we fully counterbalanced the probability of multisensory vs. unisensory events over initial and repeated conditions. In other words, whether an object was presented in a unisensory or multisensory manner was not predictive of whether it was an initial or repeated presentation. Moreover, by intermixing initial unisensory and multisensory presentations, we could directly address the question of whether attentional capture by the task-irrelevant modality could explain the impact upon unisensory recognition by increasing the salience of these stimuli with respect to unisensory presentations (Donohue, Todisco, & Woldorff, 2013; Kiss & Eimer, 2011; McDonald, Stormer, Martinez, Feng, & Hillyard, 2013; Van der Burg, et al. 2008). The initial semantic pairing of the auditory and visual stimuli had a significant effect on subsequent recall of the unisensory stimulus. Because this semantic pairing had such a strong influence, we exclude the possibility that attentional capture played a role (e.g. by merely enhancing the saliency of stimuli) (Zimmer, Roberts, Harshbarger, & Woldorff, 2010), suggesting instead the involvement of a perceptual memory mechanism (Brunel, Goldstone, Vallet, Riou, & Versace, 2013; Brunel, Labeye, Lesourd, & Versace, 2009).

Additionally, the specific multisensory pairings were manipulated on a trial-by-trial basis, rather than presenting specific pairings in blocks, again preventing that participants could predict anything about the upcoming trial. It has been argued that the magnitude of the congruency effect (i.e. faster reaction times (RTs) and higher accuracy upon congruent auditory-visual trials as compared to incongruent trials) is highly context-dependent (Botvinick, Cohen, & Carter, 2004; Egner, 2007; Lindsay & Jacoby, 1994; Sarmiento, Shore, Milliken, & Sanabria, 2012). These studies have argued that the magnitude of the interference depends on the proportion of congruent vs. incongruent trials within a block. More precisely, they have shown that interference effects were observed when 25% of trials within a block were incongruent presentations (vs. 75% congruent trials), but not when the majority of trials were incongruent (i.e. 75% incongruent vs. 25% congruent trials). King and colleagues (2012) have argued that frequent, task-irrelevant stimuli can lead to an enhanced conflict resolution, thus diminishing the interference effect (King, Korb, & Egner, 2012). This interference resolution is thought to occur in an automatic fashion, and to bypass participants' awareness. Although the findings in support of this conclusion were always relative to simultaneously presented multisensory pairs, they can still be related to the present study. In fact, if such congruity effects impact the encoding of initial object presentations in the present study, this could be reflected in the ambiguity of the response given upon subsequent retrieval of auditory and visual objects. In a prior study Lehmann and Murray (2005) failed to observe the impact of prior incongruent multisensory pairings upon subsequent visual object recognition (Lehmann & Murray, 2005). In the light of the aforementioned findings, this can be explained, by the relatively high percentage of incongruent (25%) vs. congruent (25%) trials in their design (vs. 50% of unisensory presentations). It could be argued that incongruent pairings occurred too frequently throughout the block, thus engaging conflict resolution mechanisms. Contrariwise, in the present design, incongruent pairings occurred on 8.3% of the initial presentation trials, which might have led to no or very little recruitment of such conflict resolution mechanisms. Consequently, the eventual engagement of such context-dependent conflict resolution mechanisms, which could have been differentially involved in our past studies, can likely be excluded.

The major findings here (and in our previous work) are largely in accordance with the conceptual short-term memory (CSTM) model proposed by Potter and Intraub (Intraub, 1980, 1984; Potter, 1976). This model is based on the "momentary identification hypothesis", which states that during rapid presentation of visual objects, images are momentarily understood, but immediately forgotten upon presentation of the following event. In a more recent study, Crouzet et al. (2014) have shown that object identification is impaired by object-substitution masking. Object-substitution masking has been studied in the context of visual perception, to elucidate the interplay between feedforward and feedback processing (see Breitmeyer & Ogmen, 2000 for a review). Commonly, this type of masking is achieved by presenting dots surrounding a target object within a briefly presented search array. Upon disappearance of the search array, these dots remain visible on the screen, interfering with reentrant information from higher-order to lower-order visual areas. Crouzet et al. (2014), asked subjects to saccade towards the side of the screen containing a target vs. a distractor object. While accuracy was generally high when the search array was followed by a blank screen, it dropped considerably in the object substitution masking condition, where four dots placed around the target and distractor remained on the screen for 300ms after the search array disappeared. Similarly, Donk and van Zoest (2008) have reported transient, short-lived saliency effects to occur in a visual search task. More precisely, these authors reported that brief presentations of search arrays led to high responses accuracies when subjects were asked to indicate the location of a highly salient singleton within a search array (Donk & van Zoest, 2008, Experiment 2). When the search array was presented for longer duration, subject's performance dropped significantly. Similarly, Joubert et al. (2008), showed that incongruent context/object pairings lead to a decrease in accuracy in a rapid animal vs. non-animal categorization task (Joubert, Fize, Rousselet, & Fabre-Thorpe, 2008). Furthermore, these authors showed that such a decrease in object categorization accuracy occurred independently from object saliency, and that the impact of context processing influenced object processing during early perceptual stages. Taken together, these studies provide evidence that visual object related information is accessed immediately after onset of presentation, but rapidly deteriorated by subsequent visual information interfering with the maintenance of object information in perceptual/working memory (Crouzet, Overgaard, & Busch, 2014; Donk & van Zoest, 2008; Joubert et al., 2008).

Although the above-mentioned studies focused on sequential visual-only presentations, their conclusions can likely be generalized to the simultaneous auditory-visual presentations we used in our study. In fact, Intraub (1980, 1984) proposed that short presentations in rapid succession interfere with sensory/memory trace formation when attention is shifted from one image to the next. Here, we couple auditory and visual objects, which are most likely processed by independent sensory short term memory processes, as suggested by the lack of explicit correlations between modalities. Consequently, interference effects upon subsequent unisensory retrieval were strongest for objects that had been paired with a semantically incongruent stimulus upon initial encounters. Additionally, the CSTM model can also explain the recognition enhancement observed for objects that had been paired with a semantically congruent stimulus. If switching attention between modalities still entails processing of the same object, this would lead to a further enhancement (rather than interference through incongruent sensory information) of the conceptual representation in either of the senses, facilitating subsequent retrieval processes.

Further support for this hypothesis comes from a recent EEG study on visual working memory capacity (Diamantopoulou, Poom, Klaver, & Talsma, 2011). This study examined the impact of stimulus distinctiveness upon visual object recognition. More precisely, subjects performed a delayed match-to-sample task of either discrete (different shapes and colors) or continuous (a set of ellipses which varied across the shape and color dimension in a continuous manner) geometrical forms. Visual working memory capacity was increased for discrete stimuli as compared to continuous stimuli. The authors hypothesized that this difference could be linked to whether or not subjects could verbalize the stimuli during the memorization period. In other words, while subjects could easily associate distinct labels to stimuli in the discrete condition, this was more difficult for stimuli varying within the same shape and color category. These findings can be related to the present ones, when considering the impact of recruiting semantic concepts from long-term memory representations. In the case of congruent auditory-visual pairings, both modalities access the same concept within long-term memory networks, reinforcing the object representation and, most probably, leading to internal verbalization of the object (see also Chen & Spence, 2011 for a putative cognitive model). The activation of such higher-order object processing networks could have led to enhanced recognition accuracy upon subsequent unisensory retrieval. Contrariwise, the presentation of an incongruent auditory-visual pair would have led to the internal verbalization of two distinct concepts, leading to recognition accuracy impairment upon subsequent unisensory presentations. In the case of initial pairings of meaningful sounds with meaningless visual objects, subjects would not associate a label to the concurrent visual stimulus, thus not interfering with encoding processes of the auditory object.

While unisensory object discrimination is similarly affected by prior multisensory contexts (that is, discrimination is improved by prior semantically congruent contexts and impaired by prior semantically incongruent and meaningless contexts), we also observed some notable distinctions between the sensory modalities in which the task was performed. First, effects in one modality did not correlate with those in the other. While we are reluctant to over-interpret a null result, it would nonetheless suggest that visual and auditory object processing mechanisms operate in relative independence, as has been previously proposed by psychophysical findings (Goll, Crutch, & Warren, 2010; Murray, De Santis, Thut, & Wylie, 2009). Support for this partial segregation of processing mechanisms between sensory modalities also comes from studies of attentional mechanisms. Talsma et al. (2006) investigated how attending to visual, auditory or auditory-visual objects affected the processing of a rapid stream of letters that was presented concurrently with the objects This was done by recoding steady-state visual evoked potentials that were evoked by the letter streams. The amplitudes of these potentials were significantly decreased when subjects had to pay attention to concurrent visual and auditory-visual stimuli, compared to when subjects attended the auditory objects. This result suggests that attending to the visual objects competes with the processing of the letter stream, whereas attending to the auditory objects evokes no such competition. The authors thus concluded that attentional modulations of auditory and visual neural processes occurred in relative independence. Consequently, rather than solely involving a general object recognition/memory and/or attentional process, it seems as though single-trial multisensory memories affect sensory-specific memory trace formation and retrieval processes.

Second, consideration of the raw accuracy rates Table 2a would indicate that performance was generally worse when subjects had to make auditory discriminations than visual ones. This result is consistent with Cohen et al.'s (2009) proposal. In fact, these authors have proposed that auditory memory capacity might be generally lower than visual memory. Alternatively, the generally lower accuracy rates for auditory objects observed in Cohen et al. (2009) and the present study could stem from the specific presentation context, i.e. the multisensory pairing. In fact, Welch and Warren (1980) proposed that vision to be the more efficient, and thus reliable, sensory modality when processing objects (Welch & Warren, 1980). In accordance with this modality appropriateness hypothesis, the present data would suggest that the co-occurrence of unisensory and multisensory trials within an experimental block could have given rise to higher interference from the visual information on auditory object processing, thus leading to generally lower accuracy rates.

Furthermore, we analyzed the gain/cost indices as a function of task order (see Materials and Methods). Results of this analysis showed that subjects who performed the auditory task one week prior to completing the visual task showed greater gains for repeated image presentations that had been paired with a congruent auditory stimulus upon initial encounters. Similarly, Hecht et al. (2009) suggest that visual stimuli show greater facilitation/priming effects following congruent vs. incongruent auditory-visual exposure (Hecht, Reiner, & Karni, 2009). In accordance to the findings of Vetter et al. (2014), both Hecht et al.'s findings and ours suggest that auditory-visual priming effects might be strongly intertwined with predictive coding effects during initial auditory-visual presentations, and that these effects affect visual more than auditory object processing. Additional research is clearly required to examine the importance of subjects performing the task in a single sensory modality. That is, here subjects explicitly attended to only the task-relevant modality. It remains unclear if similar effects would be observed had subjects been confronted with unisensory stimuli in either sensory modality within the same block of trials; a topic of ongoing research in our group.

Third, interference from the semantically incongruent task-irrelevant stimuli was greater for subsequent auditory recognition as compared to visual object discrimination (Figures 2a and 2b). Interestingly, this specific effect was observed in the absence of a main effect of Task Modality, but was described by a Task Modality by Multisensory Encounter Context interaction (see result section). Thus, the lack of a task specific difference in terms of gain/cost indices, along with generally higher recognition accuracy in the visual task compared to the auditory task, might be explained in the light of the assumption that vision is the more appropriate and thus dominant sense in object processes at least under the conditions used here (but see Suied & Viaud-Delmon, 2009; Welch & Warren, 1980; Yuval-Greenberg & Deouell, 2007, 2009). In other words, this specific difference in magnitude of impact between sensory modalities, cannot merely be explained by the Principle of Inverse Effectiveness (Stein & Meredith, 1993). If tis would have been the case, we would expect to see a general amplification of the magnitude of the observed effects within the auditory task, irrespective of the initial encounter context. Rather, the underlying mechanism is thought to be the high spatial sampling rate of the visual system, which relays the less ambiguous information very rapidly, whereas the auditory system necessitates information to unfold over time in order to unambiguously identify an object. Thus, presenting a semantically incongruent task-irrelevant object when subjects discriminate auditory objects led to greater interference upon formation of the sensory/memory trace and, consequently a more ambiguous retrieval of the latter upon subsequent encounters. In contrast, during the visual task subjects do not rely upon audition to unambiguously discriminate objects. Moreover, visual dominance effects can explain why auditory object processing is less prone to interference from prior co-exposure to meaningless visual stimuli; the hypothesis being that the visual system rapidly identifies the objects as not conveying relevant object-related information. Consequently, object discrimination resources between the sensory systems are less likely to compete.

Likewise, these results suggest that predictive coding mechanisms might differ in their magnitude between the auditory and visual object processes. The more robust impact of visual information on subsequent auditory object recognition suggests that visual information can lead to category specific predictive activations within auditory object processing areas, similar to what has been reported by Vetter et al. (2014) for auditory information. Such a mechanism is reflected in the greater gain/cost indices observed for auditory as compared to visual object recognition. More precisely, when auditory objects had been presented in a congruent or incongruent pairing upon initial encounters, the gain/cost indices were significantly larger than in the visual modality. In fact, if such predictive coding mechanisms are involved during the initial presentation in the present study, auditory object processing is facilitated when visual information is congruent, leading to more robust memory trace formation. Similarly, if incongruent visual information is forwarded to auditory object sensitive cortices, the resulting activation patterns would interfere with the processing of the auditory object, and ultimately with object memory trace encoding.

Conclusions

Taken together, the present study shows that memory traces formed after single-trial multisensory encounters impact subsequent auditory object discrimination. To our knowledge this is the first demonstration of such effects. Moreover, we demonstrate there to be generally similar effects of prior multisensory contexts on both auditory and visual object discrimination in the same group of participants. This was the case even though raw performance was generally poorer in the auditory than visual modality. This suggests that both modalities can benefit from past task-irrelevant multisensory experiences, despite their likely being general underlying differences in the efficacy of memory processes within each sensory modality. .

Acknowledgements

We thank Dr. Pawel Matusz for comments and discussion during the revision of this manuscript. Financial support was provided by the Swiss National Science Foundation (grants 151771 to AT and 149982 to MMM). Additional support was provided by the Swiss Brain League (2014 Research Prize to MMM).

References:

Altieri, N., Stevenson, R. A., Wallace, M. T., & Wenger, M. J. (2013). Learning to Associate Auditory and Visual Stimuli: Behavioral and Neural Mechanisms. Brain Topography.

Botvinick, M. M., Cohen, J. D., and Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends in cognitive sciences, 8(12), 539-546.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: a comparison, review, and update. Perception & Psychophysics, 62(8), 1572–95.

Brunel, L., Goldstone, R. L., Vallet, G., Riou, B., & Versace, R. (2013). When seeing a dog activates the bark: multisensory generalization and distinctiveness effects. Experimental Psychology, 60(2), 100–12.

Brunel, L., Labeye, E., Lesourd, M., & Versace, R. (2009). The sensory nature of episodic memory: sensory priming effects due to memory trace activation. Journal of Experimental Psychology. Learning, Memory, and Cognition, 35(4), 1081–8.

Cappe, C., Thelen, A., Romei, V., Thut, G., and Murray, M. M. (2012). Looming signals reveal synergistic principles of multisensory interactions. Journal of Neuroscience, 32, 1171-82.

Chen, Y. C. and Spence, C. (2011). Crossmodal Semantic Priming by Naturalistic Sounds and Spoken Words Enhances Visual Sensitivity. Journal of Experimental Psychology Human Perception and Performance, 37(5), 1554–1568.

Cohen, M. A., Horowitz, T. S., and Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. Proc Natl Acad Sci USA 106(14), 6008-6010

Crouzet, S. M., Overgaard, M., & Busch, N. A. (2014). The fastest saccadic responses escape visual masking. PloS One, 9(2), e87418.

De Lucia, M., Clarke, S., & Murray, M. M. (2010). A temporal hierarchy for conspecific vocalization discrimination in humans. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 30(33), 11210–21.

De Lucia, M., Cocchi, L., Martuzzi, R., Meuli, R. a, Clarke, S., & Murray, M. M. (2010). Perceptual and semantic contributions to repetition priming of environmental sounds. Cerebral Cortex (New York, N.Y. : 1991), 20(7), 1676–84.

De Lucia, M., Tzovara, A., Bernasconi, F., Spierer, L., & Murray, M. M. (2012). Auditory perceptual decision-making based on semantic categorization of environmental sounds. NeuroImage, 60(3), 1704–15.

Diamantopoulou, S., Poom, L., Klaver, P., and Talsma, D. (2011). Visual working memory capacity and stimulus categories: a behavioral and electrophysiological investigation. Experimental brain research, 209(4), 501–13.

Donk, M., & van Zoest, W. (2008). Effects of salience are short-lived. Psychological Science, 19(7), 733–9.

Donohue, S. E., Todisco, A. E., and Woldorff, M. G. (2013). The rapid distraction of attentional resources toward the source of incongruent stimulus input during multisensory conflict. Journal of cognitive neuroscience, 25(4), 623-635

Egner, T. (2007). Congruency sequence effects and cognitive control. Cognitive, Affective, & Behavioral Neuroscience, 7(4), 380–390.

Gibson, J. R., & Maunsell, J. H. (1997). Sensory modality specificity of neural activity related to memory in visual cortex. Journal of Neurophysiology, 78(3), 1263–75.

Goll, J. C., Crutch, S. J., & Warren, J. D. (2010). Central auditory disorders: toward a neuropsychology of auditory objects. Current Opinion in Neurology, 23(6), 617–27.

Gottfried, J. A., Smith, A. P., Rugg, M. D., and Dolan, R. J. (2004). Remembrance of odors past: human olfactory cortex in cross-modal recognition memory. Neuron 42(4), 687-695.

Griffiths, T. D. (2002). Central auditory pathologies. Brit Med Bull, 63, 107-120.

Hecht, D., Reiner, M., & Karni, A. (2009). Repetition priming for multisensory stimuli: task-irrelevant and task-relevant stimuli are associated if semantically related but with no advantage over uni-sensory stimuli. Brain Research, 1251, 236–44.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. Journal of experimental psychology: Human learning and memory, 6(1), 1-12.

Intraub, H. (1984). Conceptual masking: the effects of subsequent visual events on memory for pictures. Journal of experimental psychology: Learning, memory, and cognition, 10(1), 115-125.

Johansson, B. B. (2012). Multisensory stimulation in stroke rehabilitation. Frontiers in Human Neuroscience, 6(April), 60. doi:10.3389/fnhum.2012.00060

Joubert, O. R., Fize, D., Rousselet, G. a, & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. Journal of Vision, 8(13), 11.1–18.

King, J. A., Korb, F. M., and Egner, T. (2012). Priming of control: implicit contextual cuing of top-down attentional set. Journal of neuroscience, 32(24), 8192-8200.

Kiss, M. and Eimer, M. (2011). Faster target selection in preview visual search depends on luminance onsets: behavioral and electrophysiological evidence. Attention, perception & psychophysics, 73(6), 1637-1642.

Knebel, J.-F., Toepel, U., Hudry, J., Le Coutre, J., and Murray, M. M. (2008). Generating controlled image sets in cognitive neuroscience research. Brain topography, 20(4), 284–9.

Lehmann, S. and Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. Brain research. Cognitive brain research, 24(2), 326–34.

Lindsay, D. S. and Jacoby, L. L. (1994). Stroop process dissociations: the relationship between facilitation and interference. Journal of experimental psychology: Human perception and performance, 20(2), 219-234.

McDonald, J. J., Stormer, V. S., Martinez, A., Feng, W., and Hillyard, S. A. (2013). Salient sounds activate human visual cortex automatically. Journal of neuroscience, 33(21), 9194-9201.

Meylan, R. V. and Murray, M. M. (2007). Auditory-visual multisensory interactions attenuate subsequent visual responses in humans. NeuroImage, 35(1), 244–54.

Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., and Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. NeuroImage, 21(1), 125–135.

Murray, M. M., Foxe, J. J., and Wylie, G. R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. NeuroImage, 27(2), 473–8.

Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 26(4), 1293–302.

Murray, M. M., De Santis, L., Thut, G., and Wylie, G. R. (2009). The costs of crossing paths and switching tasks between audition and vision. Brain and cognition, 69(1), 47-55.

Murray, M. M. and Sperdin, H. F. (2010). Single-trial multisensory learning and memory retrieval. In Kaiser, J. & Naumer, M. J. (eds.) Multisensory object perception in the primate brain.

Murray, M. M., Cappe, C., Romei, V., Martuzzi, R., and Thut, G. (2012). Auditory-visual multisensory interactions in humans: a synthesis of findings from behavior, ERPs, fMRI, and TMS, in: The New Handbook of Multisensory Processes, B.E. Stein (Ed.), MIT Press, Cambridge, MA, USA. 223-238.

Naghavi, H. R., Eriksson, J., Larsson, A., and Nyberg, L. (2011). Cortical regions underlying successful encoding of semantically congruent and incongruent associations between common auditory and visual objects. Neuroscience letters, 505(2), 191-195.

Nyberg, L., Habib, R., McIntosh, A. R., and Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. Proc Natl Acad Sci USA 97(20), 11120-11124.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9(1), 97-113.

O'Mahony, C. and Newell, F. N. (2012). Integration of faces and voices, but not faces and names, in person recognition. British journal of psychology, 103(1), 73-82.

Potter, M. C. (1976). Short-term conceptual memory for pictures. Journal of experimental psychology: Human learning and memory, 2(5), 509-522.

Romei, V., Murray, M. M., Merabet, L. B., and Thut, G. (2007). Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions. Journal of Neuroscience, 27, 11465-72.

Sarmiento, B. R., Shore, D. I., Milliken, B., and Sanabria, D. (2012). Audiovisual interactions depend on context of congruency. Attention, perception & psychophysics, 74(3), 563-574.

Schall, S., Kiebel, S. J., Maess, B., & von Kriegstein, K. (2013). Early auditory sensory processing of voices is facilitated by visual mechanisms. NeuroImage, 77, 237–45.

Schneider, T. R., Debener, S., Oostenveld, R., & Engel, A. K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *NeuroImage*, *42*(3), 1244–54.

Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory Identification of Natural Objects in a Two-Way Crossmodal Priming Paradigm. *Experimental Psychology (formerly "Zeitschrift Für Experimentelle Psychologie")*, *55*(2), 121–132.

Schneider, T. R., Lorenz, S., Senkowski, D., & Engel, A. K. (2011). Gamma-band activity as a signature for cross-modal priming of auditory object recognition by active haptic exploration. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(7), 2502–10.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends in Neurosciences*, *31*(8), 401–9.

Senkowski, D., Schneider, T. R., Tandler, F., & Engel, A. K. (2009). Gamma-band activity reflects multisensory matching in working memory. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *198*(2-3), 363–72.

Shams, L. and Seitz, A. R. (2008). Benefits of multisensory learning. [Review]. Trends in cognitive sciences, 12(11), 411-417.

Shams, L., Wozny, D. R., Kim, R., and Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. Frontiers in psychology, 2, 264.

Sheffert, S. M., & Olson, E. (2004). Audiovisual speech facilitates voice learning. Perception & Psychophysics, 66(2), 352–62.

Snodgrass, J. G. and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychol Hum Learn 6(2), 174-215.

Spierer, L., De Lucia, M., Bernasconi, F., Grivel, J., Bourquin, N. M.-P., Clarke, S., & Murray, M. M. (2011). Learning-induced plasticity in human audition: Objects, time, and space. *Hearing Research*, *271*(1), 88–102.

Stein, B. E., & Meredith, M. A. (1993). *The merging of Senses*. (M. : M. P. Cambridge, Ed.).

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., … Wallace, M. T. (2014). Identifying and Quantifying Multisensory Integration: A Tutorial Review. Brain Topography.

Suied, C., & Viaud-Delmon, I. (2009). Auditory-visual object recognition time suggests specific processing for animal sounds. PloS One, 4(4), e5256.

Talsma, D., Doty, T. J., Strowd, R., and Woldorff, M. G. (2006). Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. Psychophysiology, 43(6), 541–9.

Thelen, A., Cappe, C., and Murray, M. M. (2012). Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. NeuroImage, 62(3), 1478–1488.

Thelen, A. and Murray, M. M. (2013). The efficacy of single-trial multisensory memories. Multisensory Research;

Thelen, A., Matusz, P. J., & Murray, M. M. (2014). Multisensory context portends object memory. *Current Biology*, *24*(16), R734–R735.

Van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. Neuron, 81(6), 1240–53.

Van den Brink, R. L., Cohen, M. X., Van der Burg, E., Talsma, D., Vissers, M. E., and Slagter, H. a. (2013). Subcortical, Modality-Specific Pathways Contribute to Multisensory Processing in Humans. Cerebral cortex.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., and Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. Journal of experimental psychology. Human perception and performance, 34(5), 1053–65.

Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., and Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. NeuroImage, 55(3), 1208–18.

van der Linden, M., van Turennout, M., and Indefrey, P. (2010). Formation of category representations in superior temporal sulcus. Journal of Cognitive Neuroscience 22, 1270-1282.

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. Current Biology : CB, 24(11), 1256–62.

von Kriegstein, K., Kleinschmidt, A., Sterzer, P., and Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. Journal of cognitive neuroscience, 17(3), 367-376.

von Kriegstein, K. and Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. PLoS Biol 4(10), e326.

Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. Psychological bulletin, 88(3), 638-667.

Wheeler, M. E., Petersen, S. E., and Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. Proc Natl Acad Sci USA 97(20), 11125-11129.

White-Traut, R., Norr, K. F., Fabiyi, C., Rankin, K. M., Li, Z., & Liu, L. (2013). Mother-infant interaction improves with a developmental intervention for mother-preterm infant dyads. Infant Behavior & Development, 36(4), 694–706.

Yuval-Greenberg, S., and Deouell, L. Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. The Journal of neuroscience, 27(5), 1090–6.

Yuval-Greenberg, S., and Deouell, L. Y. (2009). The dog's meow: asymmetrical interaction in cross-modal object recognition. Experimental brain research, 193(4), 603–14.

Zimmer, U., Roberts, K. C., Harshbarger, T. B., & Woldorff, M. G. (2010). Multisensory conflict modulates the spread of visual attention across a multisensory object. NeuroImage, 52(2), 606–16.

Figure and Table Legends:

Figure. 1. Schematic representation of the paradigm. The middle row indicates the task-relevant stimulus stream, while the upper row indicates the task-irrelevant stimuli. Context labels are shown beneath the time line. (V-/A- are unisensory repetitions of previous unisensory object presentations; V+/A+ are unisensory repetitions of previous multisensory object presentations; c = congruent; i = incongruent; m = meaningless) **a.** Illustration of the visual task. **b.** Illustration of the auditory task.

(Image width: 1.5 columns)

Figure. 2. Behavioral data from the visual and the auditory tasks (mean ± s.e.m.). **a.** and **b.** show the performance gain/cost in percentage (V+/A+ minus V-/A-), for the visual and auditory tasks, respectively. Significant effects are marked with an asterix either between conditions (above the bar graphs) or compared to a zero-matrix (on the error bar).

(Image width: 2 columns)

Table 1. Illustration of the probability of unisensory (light gray boxes) and multisensory stimuli (dark gray boxes) over trials within a block. The color code denotes initial encounter contexts (unisensory=black; congruent=green; incongruent=red; meaningless=blue). Only stimulus presentations discussed here are color coded in the repeated presentations. Note that a specific object was repeated only once throughout a block of trials (repetitions here are only for illustrative purpose). **a.** Visual blocks. **b.** Auditory blocks.

(Image width: 1.5 columns)

Table 2. a. Reaction times ± s.e.m. for the visual and the auditory tasks**. b. and c.** Correlation coefficient matrix between the reaction times upon initial encounters and discrimination accuracy upon repeated presentations in the visual and auditory tasks. Gray boxes indicate significant correlation coefficients.

(Image width: 1 column)

Appendix 1. A full list of objects used in the Visual task is shown.

(Image width: 2 columns)

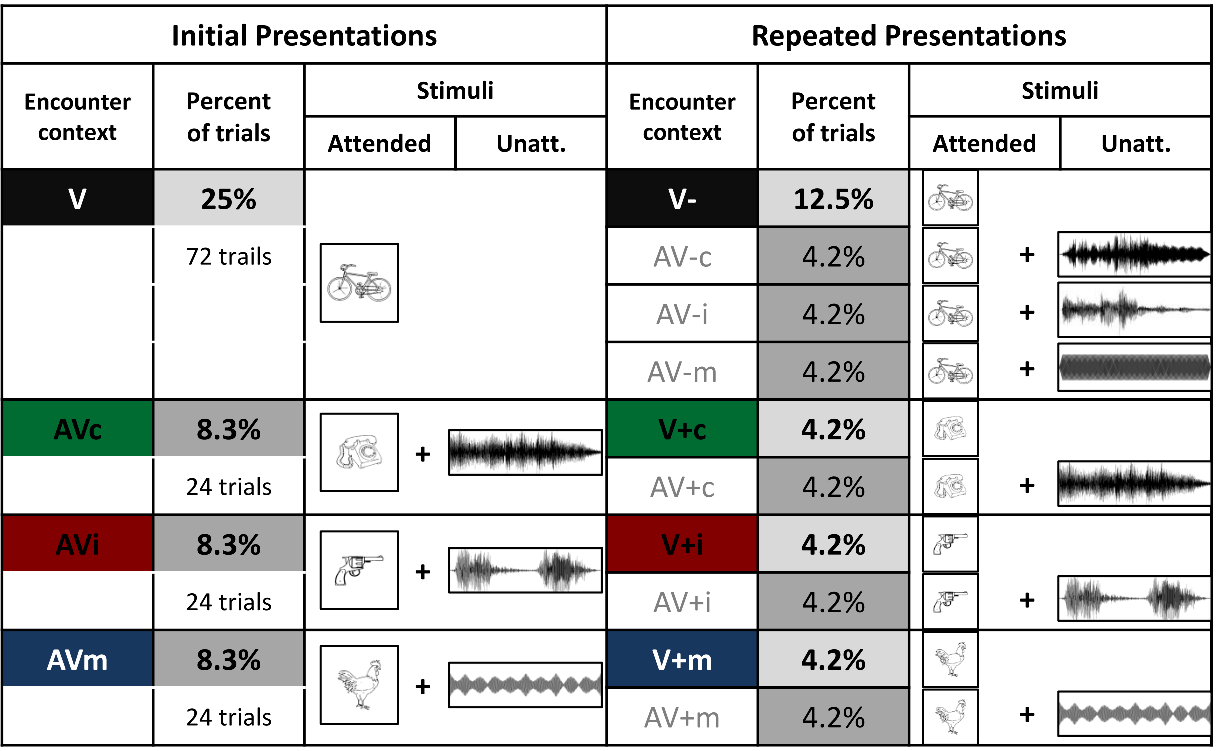
Appendix 2. A full list of objects used in the Auditory task is shown.

(Image width: 2 columns)

[i] We found that the relative slowing of response speed upon multisensory, incongruent, initial presentations in the auditory task was correlated with the retrieval accuracy for these auditory objects upon unisensory repetition ($r_{(26)}$=0.437; $t_{(24)}$=2.38; p=0.026). Also, RTs in the same initially incongruent encounter context correlated with accurate discrimination of objects that had been paired with a meaningless image upon initial encounter ($r_{(26)}$=0.564; $t_{(24)}$=3.35; p=0.003)(gray boxes in Table 2b)

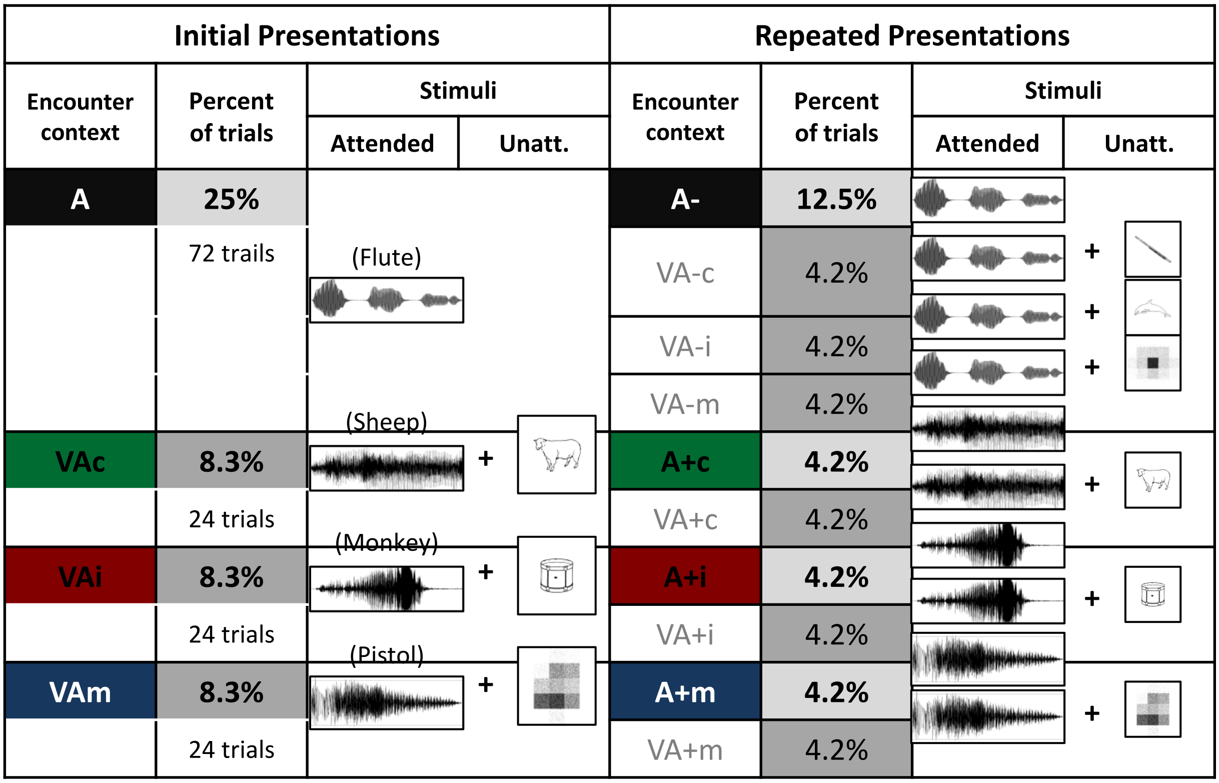
Table(s)

a.



Initial Presentations				Repeated Presentations			
Encounter context	Percent of trials	Stimuli		Encounter context	Percent of trials	Stimuli	
		Attended	Unatt.			Attended	Unatt.
V	25%			V-	12.5%		
	72 trails			AV-c	4.2%		+
				AV-i	4.2%		+
				AV-m	4.2%		+
AVc	8.3%	+		V+c	4.2%		
	24 trials			AV+c	4.2%		+
AVi	8.3%	+		V+i	4.2%		
	24 trials			AV+i	4.2%		+
AVm	8.3%	+		V+m	4.2%		
	24 trials			AV+m	4.2%		+

b.



Initial Presentations				Repeated Presentations			
Encounter context	Percent of trials	Stimuli		Encounter context	Percent of trials	Stimuli	
		Attended	Unatt.			Attended	Unatt.
A	25%			A-	12.5%		
	72 trails	(Flute)		VA-c	4.2%		+
				VA-i	4.2%		+
				VA-m	4.2%		+
VAc	8.3%	(Sheep) +		A+c	4.2%		
	24 trials			VA+c	4.2%		+
VAi	8.3%	(Monkey) +		A+i	4.2%		
	24 trials			VA+i	4.2%		+
VAm	8.3%	(Pistol) +		A+m	4.2%		
	24 trials			VA+m	4.2%		+

Table 1

a.

Conditions		Accuracy [%]	S.E.M	RTs [ms]	S.E.M
Visual Task	V-	93	1.0	765	16
	V+c	95	1.4	768	17
	V+i	89	2.0	768	16
	V+m	90	1.7	774	17
Auditory Task	A-	68	2.9	948	11
	A+c	74	3.0	948	11
	A+i	57	3.2	967	13
	A+m	68	2.6	955	13

b.

RTs Initial Presentation		V+c	V+i	V+m
	AVc	0.18	0.07	-0.08
	AVi	0.29	0.14	-0.09
	AVm	0.30	0.17	-0.03

Accuracy Repeated Presentation

c.

RTs Initial Presentation		A+c	A+i	A+m
	VAc	-0.11	0.03	-0.01
	VAi	0.22	**0.44**	**0.56**
	VAm	0.03	-0.05	0.21

Accuracy Repeated Presentation

Table 2

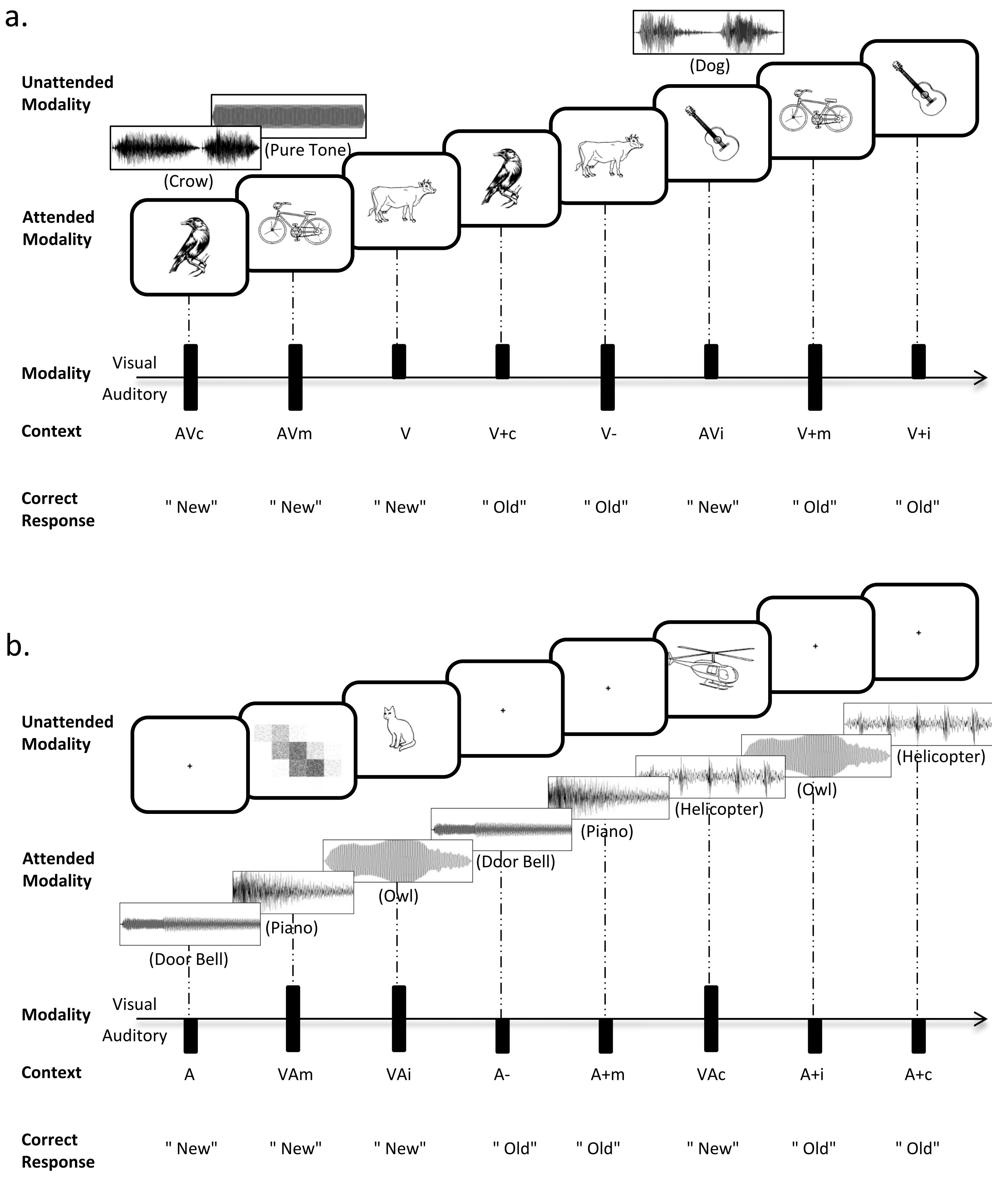

a.
Unattended Modality
Attended Modality
(Dog)
(Pure Tone)
(Crow)
Modality
Visual
Auditory
Context
AVc
AVm
V
V+c
V-
AVi
V+m
V+i
Correct Response
" New"
" New"
" New"
" Old"
" Old"
" New"
" Old"
" Old"
b.
Unattended Modality
Attended Modality
(Helicopter)
(Owl)
(Helicopter)
(Piano)
(Door Bell)
(Owl)
(Piano)
(Door Bell)
Modality
Visual
Auditory
Context
A
VAm
VAi
A-
A+m
VAc
A+i
A+c
Correct Response
" New"
" New"
" New"
" Old"
" Old"
" New"
" Old"
" Old"

Figure 1

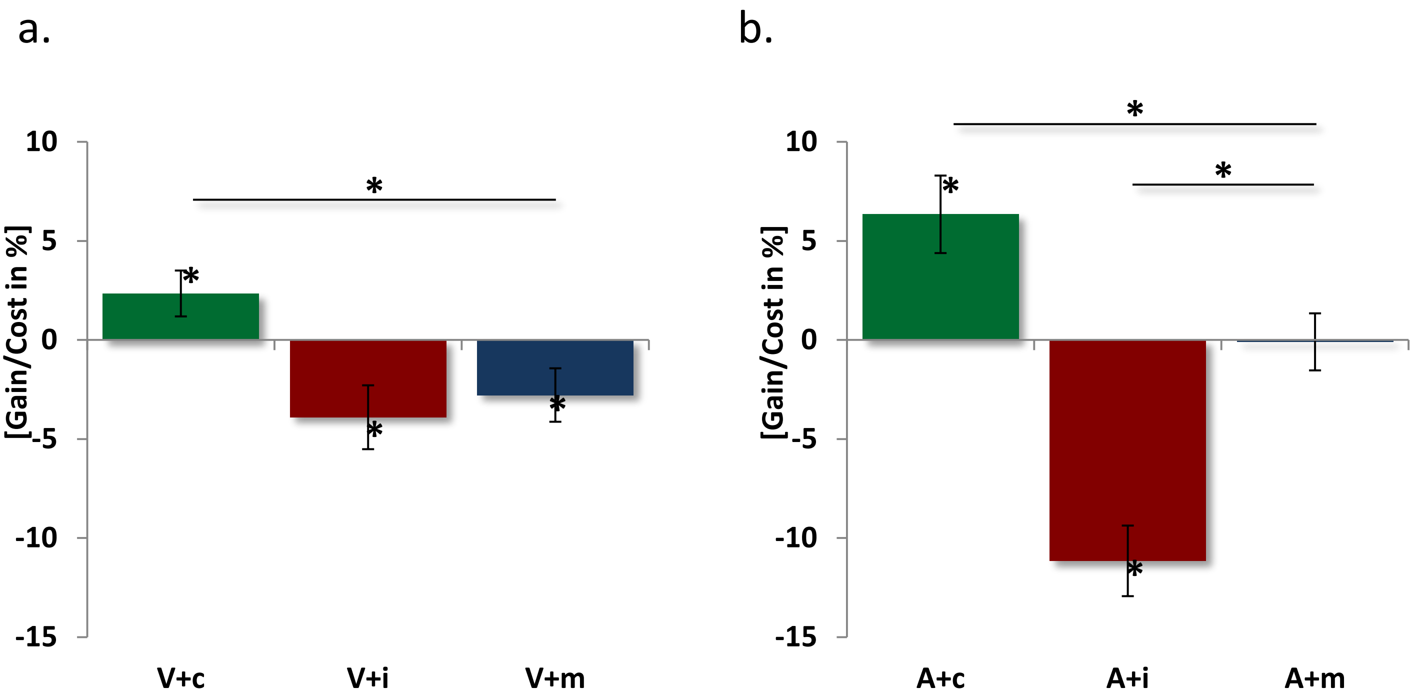
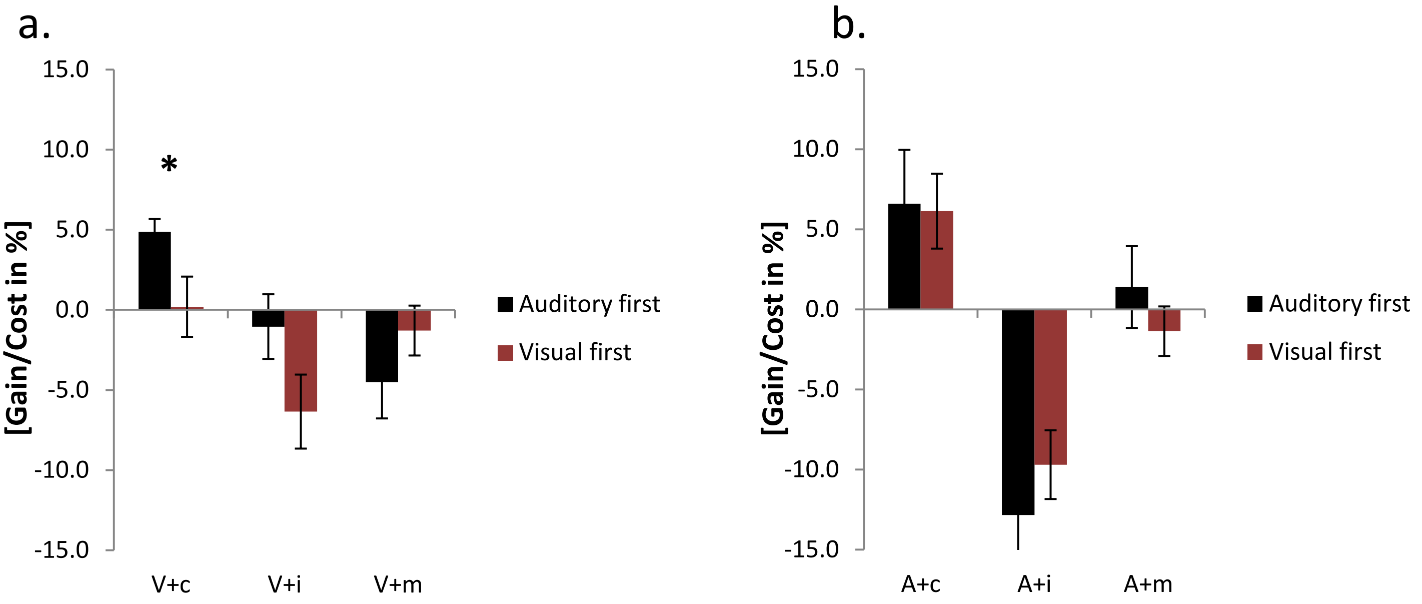

a.
10
5
0
-5
-10
-15
[Gain/Cost in %]
*
*
*
*
V+c
V+i
V+m
b.
*
*
10
5
0
-5
-10
-15
[Gain/Cost in %]
*
*
A+c
A+i
A+m

Figure 2



Reply Figure 1