

# **Multisensory contributions to object recognition and memory across the lifespan**

Pawel J. Matusz<sup>1,2,3</sup>, Mark T. Wallace<sup>3-7</sup>, Micah M. Murray<sup>1,3,8,9\*</sup>

<sup>1</sup>The LINE (Laboratory for Investigative Neurophysiology), Department of Radiology, University Hospital Center and University of Lausanne, 1011 Lausanne, Switzerland

<sup>2</sup>Information Systems Institute at the University of Applied Sciences Western Switzerland (HES-SO Valais), 3960 Sierre, Switzerland

<sup>3</sup>Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>4</sup>Department of Psychology, Vanderbilt University, Nashville, TN, USA

<sup>5</sup>Kennedy Center for Research on Human Development, Vanderbilt University, Nashville, TN, USA

<sup>6</sup>Department of Psychiatry, Vanderbilt University, Nashville, TN, USA

<sup>7</sup>Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA

<sup>8</sup>Department of Ophthalmology, University of Lausanne and Fondation Asile des Aveugles, 1003 Lausanne, Switzerland

<sup>9</sup>The EEG Brain Mapping Core, Center for Biomedical Imaging (CIBM), University Hospital Center and University of Lausanne, 1011 Lausanne, Switzerland

\*Address correspondence to:

**Prof. Micah Murray**

Department of Radiology, University Hospital Center

Rue du Bugnon 46, RC07.04.021

1011 Lausanne Switzerland

Tel: +41 21 314 1321; Email: micah.murray@chuv.ch

Contact details:

Dr. Pawel Matusz (pawel.matusz@gmail.com)

Prof. Mark Wallace (mark.wallace@vanderbilt.edu)

## **Abstract**

Everyday environments, such as classrooms or the High Street, typically stimulate multiple senses at once (i.e., are multisensory). Evidence from the last 40 years has characterized how multisensory perception unfolds and manifests neurophysiologically. However, a large majority of this research has focused on the effects of multisensory processes on instantaneous perception. This omission is important inasmuch as learning in multisensory settings is a rule rather than exception. Further, it is increasingly recognized that object representations are inherently multisensory in nature. This chapter reviews efforts to understand when and how multisensory information improves learning and memory, particularly in the context of single-trial encoding that emulates real-world settings. This chapter bridges multisensory findings and the more traditional, unisensory literature on object representations, and provides a tentative roadmap for more scientifically-based utilization of multisensory processes towards improving two crucial domains of everyday life, i.e., education and rehabilitation of sensory and cognitive functions.

### **Key words:**

Multisensory, learning, memory, object, auditory, visual

## **Introduction**

We are all familiar with this situation, typical for conferences and other networking-oriented professional meetings: You are a novice (e.g., a first-year graduate student in neuroscience, psychology etc.) and you decide to go to the pre-conference reception. Naturally, as you don't know anyone there, you arrive to the event with another young colleague from your lab, Casey. You nervously hold onto your drink and hover around one end of the table with snacks. You comment on how busy the place is, list the presentations you're looking forward to hearing, and discuss the places you'd like to visit while in the conference city. Suddenly someone comments "Apparently they have a fantastic museum of modern art here..". You and Casey turn around and you both instantly recognize an internationally renowned neuroscientist, Professor Alex Broderick. You are star-struck, but your friend Casey quickly recovers and thanks the Professor for advice as the museum wasn't on your list. Casey then introduces themselves, gesticulates at you to do the same - which you do, hesitatingly and quietly - and goes on saying which laboratory you two work on and the project you two are working on. Professor Broderick congratulates you two on an interesting research program, stating that they quickly pursued it around 20 years ago, during their graduate studies, but then stopped. A typical silence ensues, awkward smiles are exchanged and Professor Broderick excuses themselves saying they have to join their colleagues. Two days later, you and Casey bump into Professor Broderick again during the conference breakfast. Whose name do you think Professor Broderick remembers more easily: yours or Casey's?

This is just one of numerous examples that one can provide to demonstrate a fundamental principle regarding information processing and learning in the real world: They occur in contexts where information stimulating multiple senses at once is commonplace. Over the last decades, many fundamental mechanisms and principles have been revealed with respect to how memory-related processes support our abilities to perceive and interact with objects and individuals in the outside world<sup>1-5</sup>. However, these discoveries were based on unisensory, typically visual or auditory, research. This leaves open the question of the extent to which this knowledge generalizes to everyday environments, which, among possessing other important attributes, such as unpredictability, or noisiness of stimulation, are inherently multisensory. More recent research in this area demonstrated that processing stimulus attributes, at least in the case of naturalistic objects (e.g. alphanumeric symbols, identities) activates the same regions in the brain regardless

of the modality of stimulation (visual, auditory, tactile etc.), consistent with the brain inherently representing objects in a multisensory fashion<sup>6-8</sup>. However, even these studies are limited to the extent that they did not directly measure the extent of improvements (or impairments, depending on the task context) elicited in object recognition by multisensory relative to unisensory information.

In this chapter, we first review the interactions between multisensory processes and the traditional processes involved in object recognition as well as learning. We then focus on one such line of systematic investigation, that is, the processes governing the efficacy of single multisensory experiences in influencing memory for both visual and auditory objects. We summarize the main findings emerging from this research, and situate them within the broader context of literature on learning in multisensory contexts. We then identify the underlying mechanisms and conditions sufficient for multisensory-induced memory improvements (at least in some contexts), and the implications of these processes for information processing in healthy individuals across the lifespan as well as in atypical and clinical populations.

## **1. Multisensory contributions to object recognition**

What do we know about the interplay between multisensory and memory functions from the point of view how they can interact to influence brain and behavioral responses in everyday situations? Objects in the real world are typically complex and familiar, at least in terms of their semantic categories: from voices and faces through animate and inanimate objects (animals and tools, respectively) to alphanumeric, symbolic objects such as digits, letters and words (see an equally broad definition of object in other reviews<sup>9</sup>). As such, the extent to which our brains can detect, extract and benefit from redundancies in object-defining attributes within multisensory stimuli provides us with an important advantage during perceiving and interacting with objects in everyday situations. Early studies in the area have confirmed that long-term memory-dependent processes based on semantic congruence do improve perception<sup>10</sup>. Since then, multisensory processes associated with different categories of naturalistic objects have been demonstrated to influence object recognition, selective attention, memory and other cognitive functions (for detailed reviews see <sup>11-15</sup>). As such, semantic congruence seems to be an important principle governing multisensory integrative processes<sup>10</sup>, complementing in real-world environments the “classic” principles, focused on the spatial and temporal coincidence of signals across the

senses<sup>16</sup>. Behavioral benefits of such memory-based multisensory processes are typically accompanied by and/or are directly related to the activity of a network centered around the superior temporal cortex and inferior parietal sulcus (e.g. <sup>17–23</sup>). The dominating viewpoint is that these brain areas are themselves involved in and are the locus of the integration of object features into unified representations (see also <sup>24,25</sup>). Others posit that brain regions such as the superior temporal sulcus serve as a relay of unisensory information to other brain areas where these are finally integrated into a consolidated object representation<sup>26,27</sup>. For example, some views proposed that multisensory representations of object-related information are typically located in the visual cortices, which is taken to indicate the predominance of vision in object processes (e.g. <sup>28,29</sup>; but see <sup>30</sup>). Notably, other brain areas are also implicated, e.g. the planum temporale (speech/script: e.g.<sup>20</sup>), and lateral-occipital cortex (object recognition involving touch; reviewed in <sup>11</sup>). Frontal cortices (inferior and dorsolateral prefrontal areas) are typically engaged only by incongruent and/or unfamiliar audiovisual associations (for a comprehensive review, see <sup>31</sup>). We would emphasize that these propositions are not mutually exclusive, and multiple circuits and varieties of representations are likely to co-exist.

However, in the large majority of these studies, multisensory information was central to the task, i.e., participants were advised to use information across multiple senses to perform a given task. This leaves open a crucial question as to whether multisensory processes can influence perception and behavior with objects when the multisensory information is not central to the performed task. In one study on this topic, we have demonstrated that peripheral audiovisual distractors interfere with an attention-demanding task such as visual search, and do so to the same extent when the search task is easy or difficult. We demonstrated that these findings generalize across both simpler (color-defined objects) and more complex (letters) stimuli<sup>32</sup>. Notably, in these studies, targets and distractors always shared their features, suggesting the potential dependence of these effects on the goals of the observer. Furthermore, in real-world settings, where stimuli are dynamic, the detection of semantic multisensory congruence and consequent behavioral facilitation might be more dependent on the available attentional resources. In a setup with multiple visual speakers and a single voice, Alsius and Soto-Faraco<sup>33</sup> showed that detection of audiovisual face/voice congruence is dependent on the number of simultaneously presented faces, indicating the importance of available attentional resources (see also <sup>34</sup>). Similarly, the McGurk illusion (i.e. perceiving a novel auditory syllable from mismatching auditory and visual

syllables<sup>35</sup>) elicited by a task-irrelevant stimulus is reduced when attention is directed away, towards a concurrent attention-demanding task<sup>36,37</sup>(see <sup>38</sup> for absence of event-related potential (ERP) indices of the McGurk illusion in such contexts). To summarize, these findings provide evidence that multisensory processes can influence object recognition, even in naturalistic, cluttered settings. At the same time, these influences seem to be at least partly contingent on available attentional resources and/or goals of the individual. This idea is supported by the relative late latency with which semantic memory-dependent multisensory processes engage the brain (>100ms post-stimulus<sup>19,28,39-43</sup>) and their strong dependence on the task<sup>44-46</sup>. These effects have important clinical implications in terms of utility of such processes in supporting recovery of neurocognitive functions during rehabilitation. An area that has been relatively less researched is the extent to which memory-dependent multisensory processes can influence learning.

## **2. Multisensory learning as the norm rather than exception**

Circumstances under which multisensory memory traces impact subsequent unisensory retrieval is an issue directly related to the extent to which multisensory processes can be utilized to support learning in real-world settings, as well as in rehabilitation. However, these circumstances remain largely unresolved. This question falls into the broader research framework focusing on the general differences in learning across unisensory (visual, auditory) and multisensory (audiovisual) settings (e.g. <sup>47-49</sup>) (**Figure 1A**). Research involving a wide variety of stimuli has consistently demonstrated that learning in multisensory settings is more effective and efficient than in comparable unisensory settings (reviewed in <sup>48</sup>). For example, during coherent motion detection and discrimination, perceptual learning involving auditory–visual stimuli is more effective than visual training<sup>50</sup>. Individuals undergoing audiovisual training, compared to those undergoing purely visual training, learned faster not only overall, across the whole training involving 10 sessions, but this advantage was already visible within the first training session. These and other studies clearly demonstrate that the brain’s perceptual skills and cognitive functions are particularly attuned to multisensory processes. To the extent that multisensory attuning is a general property of brain functions, fundamental memory processes, such as encoding, storage and retrieval of information would all be facilitated in the context of multisensory information, whereas unisensory information is typically suboptimal, to the extent

that the computational brain architecture in place is not utilized to its full extent under unisensory conditions. In another study involving visual motion discrimination<sup>51</sup>, benefits of audio-visual training over purely visual training were found exclusively in a group that trained with congruent multisensory information (auditory and visual stimuli were moving in the same direction) but not in the incongruent condition group (two types of stimuli moving in opposite directions). While research on perceptual learning has provided important insights into the efficacy and potential circumstances promoting the benefits elicited by multisensory processes in learning, typically simplified and artificial stimuli were employed, thus leaving unclear whether these findings generalize to settings involving more naturalistic objects.

Predominantly two types of paradigms have been utilized to study the efficacy of multisensory learning on the ability to recognize unisensory (typically visual) objects. In the first paradigm, effortful and extended multisensory training preceded unisensory object recognition, with the two clearly separated into two sessions<sup>52,53</sup>. Participants were required to explicitly remember the unisensory versus multisensory context in which a stimulus appeared during training. At the brain level, visual words presented previously with matching sounds activated auditory cortices<sup>52</sup> (see <sup>53</sup> for similar findings involving images of naturalistic objects). These results were taken as supporting the so-called theory of reintegration<sup>54</sup>. This theory postulates that networks active during encoding are re-activated during retrieval. That is, presentation of a single element of a consolidated memory suffices to (re)activate the representation of the whole experience. However, it is noteworthy that in these studies, stimuli learned in a multisensory context were remembered less well than those learned in unisensory, visual contexts. Other findings in this area are more in line behavioral benefits of multisensory learning on memory. For example, in a study by von Kriegstein and Giraud<sup>55</sup>, participants learned to associate semantically congruent multisensory pairings (faces and voices) as well as arbitrarily multisensory pairings (voices and written names, and ringtones paired with cell phones or names). Subsequently, purely auditory voices were better recognized when they were initially paired with faces rather than written names, and the face-voice associations elicited enhanced functional coupling between the anterior temporal and fusiform brain regions involved in processing voices and faces, respectively. Sounds (ringtones) from arbitrary pairings showed no similar results.

The other type of paradigm provides a more consistent picture on the circumstances facilitating multisensory memory improvements (**Figure 1B**). The task requires the participants

to indicate whether the presented image (or sound, in the case of the auditory version of the task), such as that of a cow, has been presented for the first time or is a repeated stimulus (“old/new” task). Upon its initial presentation, the image is either presented alone (visual only trials) or together with a matching sound, such a “moo” for an image of a cow. This paradigm, used extensively by our group over the past ~15 years<sup>56</sup>, has a number of distinctive features that distinguish it from the first type of paradigm, while at the same time, in our opinion, bringing it closer to the settings in which multisensory processes exert their effect on learning and memory in everyday situations. First and perhaps most crucially, encoding and retrieval trials are randomly intermixed with each other within every block of trials. Second, the encoding and retrieval are separated only by a short interval of time (up to 1 minute, see below). Third, due to its focus on the episode (i.e., seeing an image or sound for the first time or subsequently) rather than on the image, the memory processes engaged by objects can be studied, without the potential confounds from focusing attention in a top-down fashion on the object identity/ category. Fourth and relatedly, the multisensory information, similarly to the experiment of von Kriegsten and Giraud<sup>55</sup>, is irrelevant to the task itself, which allows for more rigorous investigation, unconfounded by top-down attentional processes, of the effects of distinct multisensory processes.

To more closely emulate information processing in naturalistic environments, we manipulated the type of senses engaged, their task-relevance, and the relationship between the two cross-modal stimuli, as well as a variety of other factors that could determine the efficacy of multisensory memories<sup>57</sup>. The mounting evidence from our group and, more recently, other laboratories, has provided novel insights into the behavioral and brain mechanisms guiding memory and information processing in everyday situations. One take-home message from this research, which we summarize in the following section, is that a single exposure to multisensory pairings suffices to improve subsequent recognition (i.e. memory) of a unisensory stimulus and that these benefits generalize across vision and hearing<sup>25,30,56,58–62</sup>.

### **3. From multisensory learning to unisensory object memory**

#### *3.1. When is multisensory memory better than memory based on unisensory experiences?*



In our “old/new” continuous recognition task, the improvements from multisensory contexts on object memory are visible in discrimination accuracy, with no comparable benefits found on response speed. For example, when an image is presented initially (and thus encoded) in a multisensory context, its discrimination as ‘old’ vs. ‘new’ is more accurate, with these benefits found across all participants when the stimuli across two senses are semantically congruent. Across experiments, the magnitude of multisensory benefits imparted on memory by semantic congruence ranged between 2.5% and 9% compared to performance for purely visual, or auditory trials (**Figure 1C**). In experiments where the task design permitted the calculation of more direct measures of sensory processes as distinct from decision bias (i.e.,  $d'^{63}$ ), the benefits rose to 12% improvement<sup>62</sup>. Statistically, these multisensory benefits, reported until now in 6 published studies in >100 participants, include effects ranging from small to large ( $\eta^2_p = 0.14-0.63$ ; for similar size of effects in studies involving setups with separate exposure and recall, see <sup>64-68</sup>). The semantic congruence-induced benefits seem robust against a variety of factors that can be considered typical for everyday situations. For example, the multisensory benefits were reported in an fMRI study<sup>58</sup> where the usual lag between the initial and repeated presentations (5 seconds) increased ten-fold to allow for inter-trial intervals long enough to accommodate the requirements for the acquisition of fMRI data. Of particular note is that the multisensory, audiovisual benefits were found despite the ubiquitous scanner noise.

We urge the reader to recognize the importance of these findings to applications in everyday situations: Benefits for episodic memory from long-term multisensory associations of semantic features of naturalistic objects transpire without the explicit will of the individual. That is, it is not necessary for the individual to focus explicitly on the multisensory nature of the object for the benefits to be present. It seems that the crucial factor here is the established and preserved multisensory representation of these features. In the case of pairings of attributes that do not match the same object, in our paradigm, multisensory contexts are detrimental or at most highly variable in terms of their effect on object memory. Perhaps least surprisingly, semantically *incongruent* initial multisensory contexts (e.g. image of a cow with a barking sound) impair object recognition, relative to unisensory (both visual and auditory) contexts, with these impairments being of similar magnitudes to the respective multisensory benefits (4–16.5% accuracy decrease). Multisensory contexts that involve arbitrary, non-semantically related pairings, such as an image of a cow with a simple “pip”, overall also impair recognition, but these

decrements are less strong (3–4% accuracy decrease). This can be explained by the fact that roughly half the participants benefit from such meaningless multisensory contexts, while the other half are impaired by them, and this proportion is similar irrespective of whether the task is visual or auditory<sup>25</sup>. In the case of multisensory contexts that involve arbitrary, non-semantically related pairings, the benefits (when observed) ranged between 0.5–7% and 2.5–10% performance improvement for visual and auditory, respectively. We discuss these differences and the potential mechanisms underlying the differential multisensory-induced memory effects in Section 3.2.

The demonstration of benefits from multisensory contexts involving semantically congruent pairings for object memory provides important advances to our understanding of how brain functional organization as well as how such crucial cognitive functions as memory and semantic knowledge likely operate in naturalistic, multisensory settings. These novel insights are treated in detail elsewhere (see <sup>69</sup> for a comprehensive discussion). We succinctly summarize them here in Section 4, where we discuss potential applications of our findings to developmental and clinical domains.

### *3.2. What do multisensory benefits on memory depend on and how/why do they vary across individuals?*

To summarize the main characteristics of the multisensory benefits in object memory emerging from the work of our group and others, these benefits occur even following a single exposure to a multisensory context. These memory effects remain present for approximately one minute, are robust to the intervening presentation of several other test items, and are most uniformly present for multisensory pairings that are semantically congruent. There are several factors that influence both the presence and strength of multisensory benefits.

Semantic congruence elicits benefits for both visual and auditory object memory. There is a continued interest in how sensory and more higher-level cognitive processes differ between hearing and vision, the two senses so critically determining our abilities to interact with the environment<sup>70–75</sup>. Consequently, research is increasingly focusing on how the task-relevant sense influences multisensory processing in a bottom-up manner (e.g. <sup>76–78</sup>). In one of our studies<sup>61</sup> we have compared the strength of multisensory benefits in visual and auditory tasks within the same individuals, as evidence generally points to memory for auditory objects being weaker than that for visual objects<sup>43,79,80</sup>: Consistent with previous research, memory for sounds was generally

weaker than for images (67% vs. 92% accuracy, respectively). However, as expected, auditory memory benefits from semantically congruent pairings were four-fold larger than the visual benefits (8.8% vs. 2.2% accuracy improvement). As the same individuals took part in the two tasks, which were quite similar to each other, these results are in line with the principle of “inverse effectiveness”, i.e. multisensory benefits are often stronger in contexts where the inputs are weakly effective<sup>81-84</sup> extending this principle from the originally studied context of instantaneous perception to the context of memory function. .

While semantic congruence and the specific task- or goal-relevant sense are factors determining the efficacy of multisensory benefits in object memory, there seem to exist relatively important inter-individual differences in the benefits exhibited in our paradigm within healthy adults and across populations. In a study from our laboratory, Thelen et al.<sup>25</sup> systematically analyzed the bimodal distribution of benefits versus impairments from initial multisensory meaningless contexts. We found this bimodal distribution in two separate samples that each performed either the visual or the auditory version of the old/new task. As we recorded EEG activity and analyzed ERPs within the electrical neuroimaging framework in these as well as other studies (See Section 3.3), we were able to shed some light on the potential differences in the two subpopulations (i.e. those benefitting or not from multisensory contexts). When we analyzed responses (separately for participants in the visual and those in the auditory task) to the initial multisensory meaningless stimuli, we have found that multisensory benefits versus impairments were associated with enhanced versus reduced strength of brain response to the multisensory stimuli, respectively. Additionally, despite differences in latency, the same brain area, that is inferior parietal cortices or IPC, differentiated between those improving and impaired across both visual and auditory memory tasks. Notably, the two groups did not show any differences in processing unisensory visual or auditory information. The two groups also did not differ in terms of their performance metrics (accuracy and response speed), suggesting that the differences did not emerge from general distractibility. At the same time, the relative long-latency of the differences at the brain level (>150ms post-stimulus<sup>25</sup>) suggests the involvement of selective attention processes (e.g., see <sup>85,86</sup> for evidence of audiovisual modulation of unisensory responses at these latencies). Also the localization of the origin of these differences to IPC, a known area involved in multisensory processes as well as top-down attentional control (e.g., <sup>22</sup>), suggests the differences may lie in the way multisensory processes influence selective attention processes

linked to encoding of the object information. One line of current efforts in the lab is to determine whether inter-individual differences, such as those demonstrated here, originate because some individuals have a higher propensity to integrate multisensory information, irrespective of the stimulus type, stimulus combination, or even task. As will be discussed in Section 4.1, such differences do seem to be present in the population, and also emerge relatively early in life.

### 3.3. *What are the cognitive and brain mechanisms governing multisensory benefits in memory?*

Our paradigm provides an access point for a particular example of memory processes. Our paradigm focuses on episodic memory (are you seeing this object for the first time or was it shown earlier?). As the task used naturalistic objects (tools, animals), the involvement of semantic (multisensory) object memory is expected. Finally, the multisensory processes that we are investigating are those activated outside of the individual's attentional focus and goals, as the task is always unisensory (e.g. <sup>42,87</sup> for studies of similar effects of task-irrelevant multisensory processes on selective visual attention). In this section we will first discuss results related to the brain responses and mechanisms accompanying the discussed behavioral benefits, and will then contrast these with the mechanisms for multisensory benefits in memory proposed by studies involving paradigms that facilitate effortful and explicit, rather than incidental and implicit, encoding of multisensory stimuli.

#### 3.3.1. *Brain correlates of implicit multisensory benefits in memory*

Most of our research on the brain mechanisms governing the observed multisensory benefits pertains to visual memory. Besides the study on inter-individual differences in adults<sup>88</sup>, all of our analyses focused on the brain responses to *repeated* stimuli. Across studies involving ERPs and fMRI, we found consistently that the lateral occipital cortex (LOC) responds more strongly to repeated presentations of images of naturalistic objects that initially appeared with semantically congruent sounds, relative to images always presented alone. Using ERPs,<sup>56</sup> we demonstrated two different topographies, indicating that statistically distinct brain networks are activated in response to the two types of repeated images, as early as in the first 60–135ms post-stimulus (with later differences found as well; at ~210-260ms and 318-390ms). Using source

localization techniques, we revealed these effects to be driven by enhanced responses within the LOC for images previously seen in a multisensory versus visual-only context.

We have also determined the brain loci of differences in impaired recognition for images that previously presented in meaningless multisensory contexts. Images were paired with one and the same tone<sup>59</sup> or, in later studies, with a distinct tone (with tones modulated in their spectral composition, amplitude envelope and waveform type<sup>30</sup>). ERP differences underlying the behavioral impairments were observed as early as 100ms post-stimulus, and as reviewed above, were driven by changes in ERP topography and changes in the underlying configurations of brain sources. Importantly, these effects were yet again localized to a small cluster within the LOC (right), as well as a larger cluster in the posterior superior temporal sulcus (pSTS). Notably, the LOC activity was now *weaker* for multisensory contexts, while STS activations were stronger, for previously multisensory compared to visual conditions. There were also topographic differences in responses at 270-310ms, with these differences now localized to the right middle temporal cortex; the strength of the source estimation response of this area was directly related to the magnitude of memory impairment. Thus, the exact brain areas activated during a visual memory task are determined not by the mere presence of multisensory contexts, but rather the “sign” of their effect on visual object memory.

Consistent with the marked differences in the extent of behavioral benefits for visual and auditory memory<sup>61</sup>, quite different networks of brain areas as well as mechanisms seem to orchestrate the multisensory benefits in auditory versus visual memory. In a study where EEG was recorded from participants performing the old/new task on sounds<sup>62</sup>, the ERP differences associated with previous multisensory semantically congruent contexts on auditory memory were found at 35-85ms post-stimulus. Notably, right IPC, right superior temporal cortices (STC) as well as the right inferior occipital cortex and left frontal cortex supported multisensory-induced benefits in auditory memory. Crucially, right IPC and right STC – the two areas whose activity modulated to be consistent with the pattern of observed behavioral benefits – showed *suppressed* responses to previously multisensory semantically congruent sounds compared to sounds just presented alone, despite the former eliciting behavioral benefits. This direction of brain responses suggests potential involvement of a response suppression mechanism, proposed to govern short-term learning within auditory cortices<sup>89,90</sup>.

### *3.3.2. Multisensory representations of objects in the brain*

Collectively, the results discussed in Section 3.3.1 bear important implications for our understanding of the way in which naturalistic objects are represented in the brain and how these representations are accessed. A consistent finding emerging across our fMRI and EEG studies is that the representations of task-relevant objects were affected early during brain processing by whether previous object presentations involved multisensory contexts or not. That is, networks responsible for the processing of unisensory stimuli have access to multisensory memory representations early on in sensory–cognitive processing. Notably, using source estimation techniques, we demonstrated that this access is reflected by brain activity within nominally unisensory object-recognition brain areas (accompanied by IPC activity in a task involving memory for sounds).

Our proposal is that these early ERP modulations reflect rapid reactivation of distinct multisensory (audio-visual) and visual or auditory object representations affected in the course of encoding during initial stimulus presentation. Several lines of evidence support this idea. First, it is now well-established that unisensory objection-recognition regions - LOC in the case of vision and STC in the case of hearing - exhibit auditory–visual convergence (e.g., reviewed in <sup>15,31,91</sup>). Second, multisensory object representations are present in these areas and are distinguishable from their unisensory counterparts. Studies recording from microelectrodes in monkey posterior inferotemporal (IT) cortex (LOC is believed to be the human homologue of IT), as well as visual area V4, show selective delay-period responses on a delayed match-to-sample task for specific multisensory and unisensory pairings (e.g., <sup>92–95</sup>; see also <sup>96–98</sup>). The IT and V4 neurons were selectively responsive to unisensory stimuli as a function of the learned association, i.e., whether a given visual stimulus appeared with another visual stimulus or rather an auditory stimulus. Notably, these neurons were selectively responsive to a given learned association<sup>92</sup>. While we recognize that our findings can be influenced by the initial multisensory experiences impacting unisensory representations (which may not be mutually exclusive with our proposal above), these single-cell recordings support the idea that there are distinct representations of unisensory and multisensory associations within patches of the IT cortex (see <sup>99</sup> for findings of “patchy”, uni-/multisensory organization of areas bordering between multisensory and unisensory areas of STC).

Our work extends this body of knowledge in several important ways. First, the multisensory representations can be accessed in a fashion largely independent of the goals of the observer, to the extent that only one sensory modality was ever important to perform the task in our paradigm, while the sensory signals from the task-irrelevant modality would not be expected consciously to provide an advantage in the task. Second, these multisensory representations are established or accessed within the cortices of the sense relevant to the task. Third, these representations and/or their activation (i.e. memory traces) can be accessed after a time delay. In other words, object categorization based on past experiences, at least at early brain processing stages, is supported by processes within the task-relevant cortices that likely operate on multisensory representations.

These processes are unlikely to be similar to those engaged by the effortful encoding paradigms utilized by early studies, which provided discrepant findings on the benefits of semantically congruent multisensory processes in unisensory memory (e.g. <sup>52,53,100</sup>). Across these studies, the areas activated during memory encoding and retrieval closely overlapped. The findings from these studies were regarded as evidence for the “reintegration” account <sup>54</sup> that proposes that consolidated memory leads to the re-activation of both the task-relevant cortices (here, visual) as well as the task-irrelevant cortices (here, auditory) despite the presence of only task-relevant stimuli. However, these findings need to be qualified. For one, in the study of Nyberg et al. <sup>52</sup>, the absence of activations in other than in the auditory cortices can be due the fact that the brain areas activated during the encoding stage in this study served as regions of interest for analyses of brain responses at retrieval. Second, given the tasks explicitly required the participants to recall if a given word was learned with a sound, the activation of auditory cortices is consistent with the participants utilizing mental imagery to aid their memory recall. Notwithstanding, those paradigms, involving effortful encoding and recall, and those utilized by us, focusing on the implicit activation and influence of multisensory processes on continuous encoding/retrieval are likely to rely on different types of object representations. These paradigmatic differences could help to reconcile our findings with those proposing the critical role of the medial temporal cortices (especially perirhinal cortex) in governing the binding of semantic multisensory features into coherent object representations<sup>26,97</sup>. This notion is based predominantly on lesion studies, showing that disconnection of the perirhinal cortex results in impaired performance in the delayed match-to-sample task, in line with both encoding and

retrieval relying on the integrity of this particular area. What characterizes all of these discussed studies is the likely (yet uncontrolled) role in individuals' abilities to attend to and encode into memory the cross-modal feature pairings. In contrast, in our task, encoding was focused solely on the one, task-relevant sense. As such, these other studies might be building or engaging much richer representations than those targeted by the continuous recognition task. Repetition priming may be another mechanism at play within a continuous recognition task (see, in vision, e.g.,<sup>101</sup>; in hearing, e.g.<sup>90,102</sup>). However, we would contend that repetition priming alone could not account fully for our effects (cf.<sup>56</sup>). Instead, we reiterate, our findings are likely to be driven by multisensory representations of naturalistic objects that reside in the early cortices of the task-relevant sense that can be activated by task-irrelevant but semantically congruent stimuli, with this activation improving memory for the unisensory task-relevant counterparts of these objects during repeated presentations.

There are several lines of evidence against the multisensory benefits we have observed being driven predominantly by the initial multisensory experiences impacting unisensory representations. First, there were no accuracy differences when initial trials were unisensory and or multisensory, when all multisensory conditions (including semantically incongruent and congruent) were considered, that would indicate the presence of multisensory perception benefits. In fact, reaction times were consistently and significantly *slower* for multisensory than unisensory initial trials, suggesting some performance *costs* of initial multisensory presentations, despite the later accuracy improvements during unisensory recognition. The same pattern was observed in both visual and auditory tasks (cf. Figure 2 in<sup>103</sup>). Second, while the initial-presentation responses did not modulate reliably according to the manipulated conditions, these manipulations were highly effective in influencing both behavioral and brain responses to repeated presentations. Third, the topographic ERP differences and the differential modulation across visual<sup>56,104</sup> and auditory cortices<sup>105</sup> in response to repeated-presentation stimuli as a function of the initial multisensory contexts goes against a generic, increased top-down (memory-/goal-driven) attention and/or salience mechanism influencing the processing of the initial-presentation stimuli. Lastly, the study of Thelen et al.<sup>106</sup> has clearly demonstrated that the extent to which one benefitted versus was impaired by the initial meaningless multisensory contexts was predicted only by brain responses to the multisensory, not unisensory, initial stimuli. If perceptual processing was the driver of the memory enhancements, one would have predicted an overall



stronger responses to both multisensory and unisensory stimuli in the individuals exhibiting multisensory memory benefits versus impairments, yet no such general group differences were found.

#### **4. Broader implications: multisensory processes scaffold cognition across the lifespan**

We now first succinctly summarize the theoretical implications of our findings for models of multisensory processing as well as those of memory. We then focus on their potential practical applications for supporting development, education and well-being across the lifespan within the healthy population as well as their rehabilitative potential in atypical and clinical populations.

##### ***4.1. Theoretical implications of the interplay between multisensory processes and memory functions***

First, we demonstrate that the products of multisensory processes persist over time. This research fits with and extends the larger body of research focused on learning in multisensory contexts, based on congruencies across features from simple object categories<sup>48,107,108</sup>. This work points also to the importance of an individual's sensory experience, both long- and short-term, in influencing responses to unisensory and multisensory objects; a topic that we have treated in detail elsewhere<sup>109,110</sup>. This framework is consistent with other research aiming to clarify the interplay of stimulus-driven and top-down attentional control processes that jointly shape memory performance<sup>111</sup>.

Second, our findings challenge some of the most fundamental principles proposed to govern memory functions that have been derived from research based on purely visual stimulation. Traditional research suggests that memory performance is maximal when we retrieve information in similar contexts to those in which we have encoded it<sup>2,112</sup>. These principles may not generalize beyond visual contexts, to naturalistic contexts where notions such as conceptual novelty vs. physical familiarity come into play<sup>113</sup>. When considered together with the implicit nature of the multisensory benefits that we have observed, multisensory processes based on the detection of semantic congruence and thus based on the activation of long-term memory associations, might be particularly ubiquitous in their influences in everyday environments. Furthermore, the observed benefits are likely specific to multisensory processes, rather than any

specific object-related feature (e.g. visual) redundancy. Effects of multisensory versus unisensory redundancy are confirmed by research, across both humans and non-human animals, focused on perception<sup>114,115</sup> as well as memory<sup>64,92</sup>.

Lastly, our findings bear important implications for models of functional brain organization, by providing independent evidence for the inherently multisensory nature of object representations<sup>8,109,110</sup>. Moreover, our findings would suggest that simultaneity may be a sufficient condition for reaping multisensory benefits on learning and memory with objects in the real world (at least in the case of semantically congruent information). The majority of models of multisensory processing is based on simple stimuli and their spatio-temporal co-occurrence (cf.,<sup>14,110,116,117</sup> for reviews on the role of audiovisual simultaneity detection in modulating instantaneous perception and selective attention). Our findings suggest that in everyday life the efficacy of these processes to benefit behavior might be limited (but see<sup>16,87,118,119</sup>). To better understand the importance of multisensory processes in supporting cognitive functions in everyday environments, research in our laboratories for some time has been focused on understanding how multisensory processes influence cognitive functions in populations other than healthy typical adults.

#### ***4.2. Outlook: The importance of multisensory processes in public health***

One significant line of active research by our group focuses on the idea that a person's capacity to integrate multisensory information, such as during a simple detection task, may scale up directly to the extent to which this person utilizes multisensory experiences to facilitate object recognition and memory. That is, does one's ability to benefit from multisensory contexts in a memory task rely on a more general capacity to integrate multisensory signals, including even simple beeps and flashes presented simultaneously at the same location? One shortcoming of our extant work is that all responses were related to different components within the same paradigm (i.e. initial vs. repeated exposures during a continuous recognition task). It thus remains unclear whether links between multisensory processes are still found when measured using two or multiple tasks (each with their own stimulus set, goals, and attentional demands).

The extent to which multisensory integrative capacity maps onto specific behavioral metrics and brain mechanisms is equally unclear. Thus far, in the continuous recognition paradigm, we have reported a link between brain activity at one point in time and behavior at a

subsequent time point on the same task. We are now enriching those findings by studying relationship between multisensory processes and other metrics of behavior. We have recently explored directly the scaling of multisensory benefits across separate laboratory tasks of detection and recognition memory as well as the links of such benefits with clinical metrics of working memory and fluid intelligence (Denervaud et al., under review). In schoolchildren, like the adults described above, we observed that the magnitude of multisensory benefits on a simple detection task positively predicted the magnitude of benefits of multisensory encoding on the continuous recognition task we have described throughout this chapter. In addition, such multisensory benefits also predicted working memory scores on the ascending digits tasks and fluid intelligence scores as measured using Raven's Progressive Matrices. Our findings show that the scaffolding that low-level multisensory processes provide for higher-order memory and cognition is already established during childhood. One consequence is that typical models of cognitive development will surely need to better incorporate the role of multisensory processes; with a likely impact on education practices. They might also open exciting opportunities to facilitate early learning through multisensory programs. More generally, these data suggest that simpler and more resource-effective sensory-based methods can provide direct insights into the integrity of cognitive functions in schoolchildren.

We have likewise applied a similar approach in aging. Behavior on a simple multisensory detection task can predict memory performance measured with a standardized questionnaire indexing memory function (the Mini-Mental State Examination)<sup>120</sup>. Specifically we have demonstrated that an index combining the extent of an older person's sensory preference for auditory or visual stimuli (i.e., sensory dominance) and the extent of their multisensory benefits, both of which are derived from the same audio-visual detection task, can accurately diagnose a person as belonging to the healthy elderly versus mild-cognitive impairment group. Crucially, our task requires no specialist or trained personnel, is fast (<5 minutes), and requires no calibration of stimuli (sensory dominance and multisensory gain were linearly correlated suggesting that alternations in stimulus effectiveness would have comparable effects on both metrics). Nonetheless, it provided diagnostic values similar to more traditional neuropsychological tests that were designed specifically for such assessments and which require both financial and human infrastructure to administer and interpret.

## 5. *Conclusion*

We hope the reader has gained a sense for how the common understanding of learning and memory must be expanded into a multisensory framework<sup>109</sup>. Multisensory processes are not limited to instantaneous perceptions, though certainly important for such. Rather, they also extend over time and can impact what is typically viewed as a strictly unisensory perception. Similarly, the manner in which we now conceived of mental representations of objects, and how these representations are established and maintained, has been expanded to incorporate multisensory processes and cross-modal plasticity. In these regards, a multisensory framework can facilitate attempts to link laboratory-based research with more realistic and ethological settings and applications, including but not limited to two crucial domains of everyday life, i.e., education and rehabilitation of sensory and cognitive functions

## **Acknowledgements**

PJM receives support from the Swiss National Science Foundation (PZ00P1\_174150) and Pierre Mercier Foundation. MTW receives support from the National Institutes of Health (grants CA183492, DC014114, HD083211, MH109225), from the Simons Foundation Autism Research Initiative and from the Wallace Foundation. MMM receives support from the Swiss National Science Foundation (grants 320030-149982 and 320030-169206 as well as National Centre of Competence in Research project “SYNAPSY, The Synaptic Bases of Mental Disease” [project 51AU40-125759]) and a grantor advised by Carigest S.A. PJM and MMM are both supported by the Fondation Asile des Aveugles.

## References

1. Fuster JM. Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *J Neurophysiol.* 1973;36(1):61-78. doi:10.1152/jn.1973.36.1.61
2. Baddeley A, Eysenck MW, Anderson MC. *Memory.* Psychology Press; 2009.
3. Gazzaley A, Nobre AC. Top-down modulation: bridging selective attention and working memory. *Trends Cogn Sci.* 2012;16(2):129-135. doi:10.1016/j.tics.2011.11.014
4. Rugg MD, Thompson-Schill SL. Moving Forward With fMRI Data. *Perspect Psychol Sci.* 2013;8(1):84-87. doi:10.1177/1745691612469030
5. Summerfield C, de Lange FP. Expectation in perceptual decision making: neural and computational mechanisms. *Nat Rev Neurosci.* 2014;15(11):745-756. doi:10.1038/nrn3838
6. Mahon BZ, Anzellotti S, Schwarzbach J, Zampini M, Caramazza A. Category-Specific Organization in the Human Brain Does Not Require Visual Experience. *Neuron.* 2009;63(3):397-405. doi:10.1016/j.neuron.2009.07.012
7. Striem-Amit E, Cohen L, Dehaene S, Amedi A. Reading with Sounds: Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind. *Neuron.* 2012;76(3):640-652. doi:10.1016/j.neuron.2012.08.026
8. Reich L, Maidenbaum S, Amedi A. The brain as a flexible task machine: implications for visual rehabilitation using noninvasive vs. invasive approaches. *Curr Opin Neurol.* 2012;25(1):86-95. doi:10.1097/WCO.0b013e32834ed723
9. Amedi A, Von Kriegstein K, Van Atteveldt NM, Beauchamp MS, Naumer MJ. Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res.* 2005;166(3-4):559-571. doi:10.1007/s00221-005-2396-5
10. Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT. Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res.* 2004;158(4):405-414. doi:10.1007/s00221-004-1913-2
11. Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ. Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res.* 2005;166(3-4):559-571. doi:10.1007/s00221-005-2396-5
12. Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory? *Trends Cogn Sci.* 2006;10(6):278-285.

[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=16713325](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16713325).

13. Naumer MJ, Doehrmann O, Müller NG, Muckli L, Kaiser J, Hein G. Cortical plasticity of audio-visual object representations. *Cereb Cortex*. 2009;19(7):1641-1653. doi:10.1093/cercor/bhn200
14. Murray MM, Thelen A, Thut G, Romei V, Martuzzi R, Matusz PJ. The multisensory function of the human primary visual cortex. *Neuropsychologia*. 2016;83:161-169. doi:10.1016/j.neuropsychologia.2015.08.011
15. ten Oever S, Romei V, van Atteveldt N, Soto-Faraco S, Murray MM, Matusz PJ. The COGs (context, object, and goals) in multisensory processing. *Exp Brain Res*. 2016;234(5):1307-1323. doi:10.1007/s00221-016-4590-z
16. Matusz PJ, Dikker S, Huth AG, Perrodin C. Are We Ready for Real-world Neuroscience? *J Cogn Neurosci*. June 2018:1-13. doi:10.1162/jocn\_e\_01276
17. Calvert G a, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol*. 2000;10(11):649-657. <http://www.ncbi.nlm.nih.gov/pubmed/10837246>.
18. Calvert G a. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex*. 2001;11(12):1110-1123. <http://www.ncbi.nlm.nih.gov/pubmed/11709482>.
19. Raji T, Uutela K, Hari R. Audiovisual integration of letters in the human brain. *Neuron*. 2000;28(2):617-625. <http://www.ncbi.nlm.nih.gov/pubmed/11144369>.
20. Van Atteveldt N, Formisano E, Goebel R, Blomert L. Integration of letters and speech sounds in the human brain. *Neuron*. 2004;43(2):271-282. doi:10.1016/j.neuron.2004.06.025
21. Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*. 2004;41(5):809-823. <http://www.ncbi.nlm.nih.gov/pubmed/15813999>.
22. Werner S, Noppeney U. Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J Neurosci*. 2010;30(7):2662-2675. doi:10.1523/JNEUROSCI.5091-09.2010
23. Powers AR, Hevey M a, Wallace MT. Neural correlates of multisensory perceptual

- learning. *J Neurosci*. 2012;32(18):6263-6274. doi:10.1523/JNEUROSCI.6138-11.2012
24. Tanabe HC, Honda M, Sadato N. Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. *J Neurosci*. 2005;25(27):6409-6418. doi:10.1523/JNEUROSCI.0636-05.2005
  25. Thelen A, Matusz PJ, Murray MM. Multisensory context portends object memory. *Curr Biol*. 2014;24(16):320030. doi:10.1016/j.cub.2014.06.040
  26. Taylor KI, Moss HE, Stamatakis E a, Tyler LK. Binding crossmodal object features in perirhinal cortex. *Proc Natl Acad Sci U S A*. 2006;103(21):8239-8244. doi:10.1073/pnas.0509704103
  27. Hocking J, Price CJ. The role of the posterior superior temporal sulcus in audiovisual processing. *Cereb Cortex*. 2008;18(10):2439-2449. doi:10.1093/cercor/bhn007
  28. Molholm S. Multisensory Visual-Auditory Object Recognition in Humans: a High-density Electrical Mapping Study. *Cereb Cortex*. 2004;14(4):452-465. doi:10.1093/cercor/bhh007
  29. Diaconescu AO, Hasher L, McIntosh AR. Visual dominance and multisensory integration changes with age. *Neuroimage*. 2013;65:152-166. doi:10.1016/j.neuroimage.2012.09.057
  30. Thelen A, Cappe C, Murray MM. Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *Neuroimage*. 2012;62(3):1478-1488. doi:10.1016/j.neuroimage.2012.05.027
  31. Doehrmann O, Naumer MJ. Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Res*. 2008;1242:136-150. doi:10.1016/j.brainres.2008.03.071
  32. Matusz PJ, Broadbent H, Ferrari J, Forrest B, Merkley R, Scerif G. Multi-modal distraction: Insights from children's limited attention. *Cognition*. 2015;136:156-165. doi:10.1016/j.cognition.2014.11.031
  33. Alsius A, Soto-Faraco S. Searching for audiovisual correspondence in multiple speaker scenarios. *Exp brain Res*. 2011;213(2-3):175-183. doi:10.1007/s00221-011-2624-0
  34. Morís Fernández L, Visser M, Ventura-Campos N, Ávila C, Soto-Faraco S. Top-down attention regulates the neural expression of audiovisual integration. *Neuroimage*. 2015;119:272-285. doi:10.1016/j.neuroimage.2015.06.052
  35. McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature*. 264(5588):746-748. <http://www.ncbi.nlm.nih.gov/pubmed/1012311>.



36. Alsius A, Navarra J, Campbell R, Soto-Faraco S. Audiovisual integration of speech falters under high attention demands. *Curr Biol.* 2005;15(9):839-843.  
doi:10.1016/j.cub.2005.03.046
37. Alsius A, Navarra J, Soto-Faraco S. Attention to touch weakens audiovisual speech integration. *Exp brain Res.* 2007;183(3):399-404. doi:10.1007/s00221-007-1110-1
38. Alsius A, Möttönen R, Sams ME, Soto-Faraco S, Tiippana K. Effect of attentional load on audiovisual speech perception: evidence from ERPs. *Front Psychol.* 2014;5:727.  
doi:10.3389/fpsyg.2014.00727
39. Senkowski D, Saint-Amour D, Kelly SP, Foxe JJ. Multisensory processing of naturalistic objects in motion: a high-density electrical mapping and source estimation study. *Neuroimage.* 2007;36(3):877-888. doi:10.1016/j.neuroimage.2007.01.053
40. Wassenhove V Van, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech Neurophysiological Basis of Multisensory Integration. 2004.
41. Stekelenburg JJ, Vroomen J. Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci.* 2007;19(12):1964-1973.  
doi:10.1162/jocn.2007.19.12.1964
42. Diaconescu AO, Alain C, McIntosh AR. The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *J Neurophysiol.* 2011;106(6):2896-2909.  
doi:10.1152/jn.00303.2011
43. Yuval-Greenberg S, Deouell LY. What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *J Neurosci.* 2007;27(5):1090-1096. doi:10.1523/JNEUROSCI.4828-06.2007
44. van Atteveldt NM, Formisano E, Goebel R, Blomert L. Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage.* 2007;36(4):1345-1360. doi:10.1016/j.neuroimage.2007.03.065
45. van Atteveldt NM, Peterson BS, Schroeder CE. Contextual control of audiovisual integration in low-level sensory cortices. *Hum Brain Mapp.* 2014;35(5):2394-2411.  
doi:10.1002/hbm.22336
46. Stevenson R a, Wallace MT. Multisensory temporal integration: task and stimulus dependencies. *Exp brain Res.* 2013;227(2):249-261. doi:10.1007/s00221-013-3507-3
47. Murray MM, Sperdin HF. Single-trial multisensory learning and memory retrieval. :1-30.

48. Shams L, Seitz AR. Benefits of multisensory learning. *Trends Cogn Sci.* 2008;12(11):411-417. doi:10.1016/j.tics.2008.07.006
49. Shams L, Kim R. Cross-Modal Facilitation of Unisensory Learning. 2011:475-482.
50. Seitz AR, Kim R, Shams L. Sound facilitates visual learning. *Curr Biol.* 2006;16(14):1422-1427. doi:10.1016/j.cub.2006.05.048
51. Kim RS, Seitz AR, Shams L. Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS One.* 2008;3(1):e1532. doi:10.1371/journal.pone.0001532
52. Nyberg L, Habib R, McIntosh a R, Tulving E. Reactivation of encoding-related brain activity during memory retrieval. *Proc Natl Acad Sci U S A.* 2000;97(20):11120-11124. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=27158&tool=pmcentrez&rendertype=abstract>.
53. Wheeler ME, Petersen SE, Buckner RL. Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proc Natl Acad Sci.* 2000;97(20):11125-11129. doi:10.1073/pnas.97.20.11125
54. Hamilton W. *Lectures on Metaphysics and Logic.* Boston: Gould and Lincoln; 1859.
55. von Kriegstein K, Giraud A-L. Implicit multisensory associations influence voice recognition. *PLoS Biol.* 2006;4(10):e326. doi:10.1371/journal.pbio.0040326
56. Murray MM, Michel CM, Grave de Peralta R, et al. Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage.* 2004;21(1):125-135. doi:10.1016/j.neuroimage.2003.09.035
57. Matusz PJ, Wallace MT, Murray MM. A multisensory perspective on object memory. *Neuropsychologia.* 2017;105(February):243-252. doi:10.1016/j.neuropsychologia.2017.04.008
58. Murray MM, Foxe JJ, Wylie GR. The brain uses single-trial multisensory memories to discriminate without awareness. *Neuroimage.* 2005;27(2):473-478. doi:10.1016/j.neuroimage.2005.04.016
59. Lehmann S, Murray MM. The role of multisensory memories in unisensory object discrimination. *Brain Res Cogn Brain Res.* 2005;24(2):326-334. doi:10.1016/j.cogbrainres.2005.02.005
60. Murray MM, Thelen A. The Efficacy of Single-Trial Multisensory Memories. *Multisens Res.* 2013;26(5):483-502. doi:10.1163/22134808-00002426

61. Thelen A, Talsma D, Murray MM. Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition*. 2015;138:148-160.  
doi:10.1016/j.cognition.2015.02.003
62. Matusz PJ, Thelen A, Amrein S, Geiser E, Anken J, Murray MM. The role of auditory cortices in the retrieval of single-trial auditory-visual object memories. *Eur J Neurosci*. 2015;41(5). doi:10.1111/ejn.12804
63. Macmillan NA, Creelman CD. *Detection Theory: A User's Guide*. 2nd ed. Lawrence Erlbaum Associates Publishers; 2005.
64. Heikkilä J, Alho K, Hyvönen H, Tiippana K. Audiovisual semantic congruency during encoding enhances memory performance. *Exp Psychol*. 2015;62(2):123-130.  
doi:10.1027/1618-3169/a000279
65. Heikkilä J, Tiippana K. School-aged children can benefit from audiovisual semantic congruency during memory encoding. *Exp brain Res*. June 2015. doi:10.1007/s00221-015-4341-6
66. Naghavi HR, Eriksson J, Larsson A, Nyberg L. Cortical regions underlying successful encoding of semantically congruent and incongruent associations between common auditory and visual objects. *Neurosci Lett*. 2011;505(2):191-195.  
doi:10.1016/j.neulet.2011.10.022
67. Ueno D, Masumoto K, Sutani K, Iwaki S. Latency of modality-specific reactivation of auditory and visual information during episodic memory retrieval. *Neuroreport*. 2015;26(6):303-308. doi:10.1097/WNR.0000000000000325
68. Moran ZD, Bachman P, Pham P, Cho SH, Cannon TD, Shams L. Multisensory encoding improves auditory recognition. *Multisens Res*. 2013;26(6):581-592. doi:10.1027/1618-3169/a000279
69. Matusz PJ, Wallace MT, Murray MM. A multisensory perspective on object memory. *Neuropsychologia*. 2017;105. doi:10.1016/j.neuropsychologia.2017.04.008
70. Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S. Rapid brain discrimination of sounds of objects. *J Neurosci*. 2006;26(4):1293-1302.  
doi:10.1523/JNEUROSCI.4511-05.2006
71. De Lucia M, Camen C, Clarke S, Murray MM. The role of actions in auditory object discrimination. *Neuroimage*. 2009;48(2):475-485. doi:10.1016/j.neuroimage.2009.06.041

72. De Lucia M, Cocchi L, Martuzzi R, Meuli RA, Clarke S, Murray MM. Perceptual and semantic contributions to repetition priming of environmental sounds. *Cereb Cortex*. 2010;20(7):1676-1684. doi:10.1093/cercor/bhp230
73. Meo R De, Matusz PJ, Knebel J-F, Murray MM, Thompson WR, Clarke S. What makes medical students better listeners? *Curr Biol*. 2016;26(13):R519-R520. doi:10.1016/j.cub.2016.05.024
74. Matusz PJ, Retsa C, Murray MM. The context-contingent nature of cross-modal activations of the visual cortex. *Neuroimage*. 2016;125. doi:10.1016/j.neuroimage.2015.11.016
75. Retsa C, Matusz PJ, Schnupp JWH, Murray MM. What's what in auditory cortices? *Neuroimage*. 2018;176. doi:10.1016/j.neuroimage.2018.04.028
76. Romei V, Murray MM, Cappe C, Thut G. Preperceptual and Stimulus-Selective Enhancement of Low-Level Human Visual Cortex Excitability by Sounds. *Curr Biol*. 2009;19(21):1799-1805. doi:10.1016/j.cub.2009.09.027
77. Romei V, Murray M, Cappe C, Thut G. The contributions of sensory dominance and attentional bias to cross-modal enhancement of visual cortex excitability. *J Cogn Neurosci*. 2013;25(7):1122-1135. doi:10.1162/jocn\_a\_00367
78. Schmid C, Büchel C, Rose M. The neural basis of visual dominance in the context of audio-visual object processing. *Neuroimage*. 2011;55(1):304-311. doi:10.1016/j.neuroimage.2010.11.051
79. Cohen M a, Horowitz TS, Wolfe JM. Auditory recognition memory is inferior to visual recognition memory. *Proc Natl Acad Sci U S A*. 2009;106(14):6008-6010. doi:10.1073/pnas.0811884106
80. Yuval-Greenberg S, Deouell LY. The dog's meow: asymmetrical interaction in cross-modal object recognition. *Exp brain Res*. 2009;193(4):603-614. doi:10.1007/s00221-008-1664-6
81. Stein BE, Meredith MA. *The Merging of the Senses*. MIT Press; 1993.
82. Wallace MT, Ramachandran R, Stein BE. A revised view of sensory cortical parcellation. *Proc Natl Acad Sci U S A*. 2004;101(7):2167-2172. doi:10.1073/pnas.0305697101
83. James TW, Stevenson RA, Kim S. Inverse Effectiveness and BOLD fMRI. 2011:207-222.
84. Holmes NP. The law of inverse effectiveness in neurons and behaviour: multisensory

- integration versus normal variability. *Neuropsychologia*. 2007;45(14):3340-3345.  
doi:10.1016/j.neuropsychologia.2007.05.025
85. Matusz PJ, Eimer M. Top-down control of audiovisual search by bimodal search templates. *Psychophysiology*. 2013;50:996-1009. doi:10.1111/psyp.12086
  86. Matusz PJ, Turoman N, Tivadar RI, Retsa C, Murray MM. Brain and cognitive mechanisms of top-down attentional control in a multisensory world: Benefits of electrical neuroimaging. *J Cogn Neurosci*. 2018.
  87. Matusz PJ, Eimer M. Multisensory enhancement of attentional capture in visual search. *Psychon Bull Rev*. 2011;18(5):904-909. doi:10.3758/s13423-011-0131-8
  88. Thelen A, Matusz PJ, Murray MM. Multisensory context portends object memory. *Curr Biol*. 2014;24(16):1-10. doi:10.1016/j.cub.2014.06.040
  89. Bergerbest D, Ghahremani DG, Gabrieli JDE. Neural correlates of auditory repetition priming: reduced fMRI activation in the auditory cortex. *J Cogn Neurosci*. 2004;16(6):966-977. doi:10.1162/0898929041502760
  90. Murray MM, Camen C, Spierer L, Clarke S. Plasticity in representations of environmental sounds revealed by electrical neuroimaging. *Neuroimage*. 2008;39(2):847-856. doi:10.1016/j.neuroimage.2007.09.002
  91. Murray MM, Thelen A, Thut G, Romei V, Martuzzi R, Matusz PJ. The multisensory function of the human primary visual cortex. *Neuropsychologia*. 2016;83(1):161-169. doi:10.1016/j.neuropsychologia.2015.08.011
  92. Gibson JR, Maunsell JH. Sensory modality specificity of neural activity related to memory in visual cortex. *J Neurophysiol*. 1997;78(3):1263-1275.  
<http://www.ncbi.nlm.nih.gov/pubmed/9310418>.
  93. Colombo M, Gross CG. Responses of inferior temporal cortex and hippocampal neurons during delayed matching to sample in monkeys (*Macaca fascicularis*). *Behav Neurosci*. 1994;108(3):443-455. <http://www.ncbi.nlm.nih.gov/pubmed/7917038>.
  94. Haenny PE, Maunsell JH, Schiller PH. State dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. *Exp brain Res*. 1988;69(2):245-259.  
<http://www.ncbi.nlm.nih.gov/pubmed/3345806>.
  95. Maunsell JH, Sclar G, Nealey TA, DePriest DD. Extraretinal representations in area V4 in the macaque monkey. *Vis Neurosci*. 1991;7(6):561-573.

- <http://www.ncbi.nlm.nih.gov/pubmed/1772806>.
96. Goulet S, Murray EA. Neural substrates of crossmodal association memory in monkeys: the amygdala versus the anterior rhinal cortex. *Behav Neurosci*. 2001;115(2):271-284. <http://www.ncbi.nlm.nih.gov/pubmed/11345954>.
  97. Murray EA, Bussey TJ. Perceptual – mnemonic functions of the perirhinal cortex. 1999;3(4):142-151.
  98. Murray EA, Bussey TJ, Saksida LM. Visual perception and memory: a new view of medial temporal lobe function in primates and rodents. *Annu Rev Neurosci*. 2007;30:99-122. doi:10.1146/annurev.neuro.29.051605.113046
  99. Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci*. 2004;7(11):1190-1192. doi:10.1038/nn1333
  100. Von Kriegstein K, Giraud AL. Implicit multisensory associations influence voice recognition. *PLoS Biol*. 2006;4(10):1809-1820. doi:10.1371/journal.pbio.0040326
  101. Doniger GM, Foxe JJ, Schroeder CE, Murray MM, Higgins B a, Javitt DC. Visual perceptual learning in human object recognition areas: a repetition priming study using high-density electrical mapping. *Neuroimage*. 2001;13(2):305-313. doi:10.1006/nimg.2000.0684
  102. De Lucia M, Cocchi L, Martuzzi R, Meuli R a, Clarke S, Murray MM. Perceptual and semantic contributions to repetition priming of environmental sounds. *Cereb Cortex*. 2010;20(7):1676-1684. doi:10.1093/cercor/bhp230
  103. Thelen A, Murray MM. The efficacy of single-trial multisensory memories. *Multisens Res*. 2013;26(5):483-502. doi:10.1163/22134808-00002426
  104. Thelen A, Cappe C, Murray MM. Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *Neuroimage*. 2012;62(3):1478-1488. doi:10.1016/j.neuroimage.2012.05.027
  105. Matusz PJ, Thelen A, Amrein S, Geiser E, Anken J, Murray MM. The role of auditory cortices in the retrieval of single-trial auditory-visual object memories. *Eur J Neurosci*. 2015;41(5):699-708. doi:10.1111/ejn.12804
  106. Thelen A, Matusz PJ, Murray MM. Multisensory context portends object memory. *Curr Biol*. 2014;24(16):R734-R735. doi:10.1016/j.cub.2014.06.040

107. Powers AR, Hillock AR, Wallace MT. Perceptual training narrows the temporal window of multisensory binding. *J Neurosci*. 2009;29(39):12265-12274.  
doi:10.1523/JNEUROSCI.3501-09.2009
108. Powers AR, Hevey MA, Wallace MT. Neural correlates of multisensory perceptual learning. *J Neurosci*. 2012;32(18):6263-6274. doi:10.1523/JNEUROSCI.6138-11.2012
109. Murray MM, Lewkowicz DJ, Amedi A, Wallace MT. Multisensory Processes: A Balancing Act across the Lifespan. *Trends Neurosci*. 2016;39(8):567-579.  
doi:10.1016/j.tins.2016.05.003
110. ten Oever S, Romei V, van Atteveldt N, Soto-Faraco S, Murray MM, Matusz PJ. The COGs (context, object, and goals) in multisensory processing. *Exp Brain Res*. 2016;234(5). doi:10.1007/s00221-016-4590-z
111. Matusz PJ, Traczyk J, Sobkow A, Strelau J. Individual differences in emotional reactivity moderate the strength of the relationship between attentional and implicit-memory biases towards threat-related stimuli. *J Cogn Psychol*. 2015;27(6):715-724.  
doi:10.1080/20445911.2015.1027210
112. Smith SM, Vela E. Environmental context-dependent memory: a review and meta-analysis. *Psychon Bull Rev*. 2001;8(2):203-220.  
<http://www.ncbi.nlm.nih.gov/pubmed/11495110>.
113. Reggev N, Bein O, Maril A. Distinct Neural Suppression and Encoding Effects for Conceptual Novelty and Familiarity. *J Cogn Neurosci*. 2016;28(10):1455-1470.  
doi:10.1162/jocn\_a\_00994
114. Gingras G, Rowland B a, Stein BE. The differing impact of multisensory and unisensory integration on behavior. *J Neurosci*. 2009;29(15):4897-4902.  
doi:10.1523/JNEUROSCI.4120-08.2009
115. Chen YC, Spence C. The Crossmodal Facilitation of Visual Object Representations By Sound: Evidence From the Backward Masking Paradigm. *J Exp Psychol Hum Percept Perform*. 2011;37(6):1784-1802. doi:10.1037/a0025638
116. De Meo R, Murray MM, Clarke S, Matusz PJ. Top-down control and early multisensory processes: Chicken vs. egg. *Front Integr Neurosci*. 2015;9(MAR).  
doi:10.3389/fnint.2015.00017
117. Tang X, Wu J, Shen Y. The interactions of multisensory integration with endogenous and

exogenous attention. *Neurosci Biobehav Rev.* 2016;61:208-224.

doi:10.1016/j.neubiorev.2015.11.002

118. Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J. Pip and Pop: Nonspatial Auditory Signals Improve Spatial Visual Search. *J Exp Psychol Hum Percept Perform.* 2008;34(5):1053-1065. doi:10.1037/0096-1523.34.5.1053
119. Van der Burg E, Talsma D, Olivers CNL, Hickey C, Theeuwes J. Early multisensory interactions affect the competition among multiple visual objects. *Neuroimage.* 2011;55(3):1208-1218. doi:10.1016/j.neuroimage.2010.12.068
120. Murray MM, Eardley AF, Edginton T, Oyekan R, Smyth E, Matusz PJ. Sensory dominance and multisensory integration as screening tools in aging. *Sci Rep.* 2018;8(1):8901. doi:10.1038/s41598-018-27288-2



## Figure caption

**Figure 1. A.** A general schematic of how the impact of multisensory encoding on later unisensory recognition may be investigated. **B.** A schematic of a continuous recognition task requiring participants to indicate whether the image was presented for the first or repeated time. Whether or not the image was presented with a sound was task-irrelevant. **C.** Summary of behavioral findings. Accuracy for the various repeated presentations are displayed. Lines with circular markers refer to studies where the task was performed in the visual modality, while lines with square markers refer to studies where the task was performed in the auditory modality. Across studies, it can be seen that stimuli that had been initially presented in a semantically congruent multisensory context result in higher accuracy than stimuli that had only been experienced in a unisensory context. Other had-been multisensory contexts generally result in no difference or even performance impairment relative to the unisensory context.