UNIL | Université de Lausanne

Unicentre
CH-1015 Lausanne
http://serval.unil.ch

Year : 2014

MULTISENSORY INTEGRATION AND PERCEPTUAL ENHANCEMENT

Thelen Antonia

Thelen Antonia, 2014, MULTISENSORY INTEGRATION AND PERCEPTUAL ENHANCEMENT

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive.
http://serval.unil.ch

Droits d’auteur
L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l’éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright
The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Département de Neurosciences Cliniques du CHUV

MULTISENSORY INTEGRATION AND PERCEPTUAL ENHANCEMENT

Thèse de doctorat en Neurosciences

présentée à la

Faculté de Biologie et de Médecine
de l'Université de Lausanne

par

Antonia Thelen

Neuroscientifique, diplômée de l'Université de Genève, Suisse

Jury

Prof. Philippe Moreillon, Président
Prof. Micah Murray, Directeur
Prof. Mark Wallace, Expert
Dr. PhD Kirsten Taylor, Expert

Thèse n° 120

Lausanne 2014

Programme doctoral interuniversitaire en Neurosciences
des Universités de Lausanne et Genève

UNIL | Université de Lausanne

LEMANIC NEURO SCIENCE DOCTORAL SCHOOL

UNIVERSITÉ DE GENÈVE

Unil
UNIL | Université de Lausanne
Faculté de biologie
et de médecine

LEMANIC NEURO SCIENCE
DOCTORAL SCHOOL

Programme doctoral interuniversitaire en Neurosciences
des Universités de Lausanne et Genève

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président	Monsieur Prof. Philippe **Moreillon**
Directeur de thèse	Monsieur Prof. Micah **Murray**
Co-directeur de thèse	
Experts	Monsieur Prof. Mark **Wallace**
	Madame Dr Kirsten **Taylor**

le Conseil de Faculté autorise l'impression de la thèse de

Madame Antonia Thelen

Master en Neurosciences de l'Université de Genève

intitulée

MULTISENSORY INTEGRATION AND PERCEPTUAL ENHANCEMENT

Lausanne, le 7 février 2014

pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Philippe Moreillon

Acknowledgements

A thesis is never achieved by a student on his/her own. The secret ingredients to a happy PhD lie within our social and working environment.

First, an expert in the field willing to guide and promote the student during his/her first steps in the academic world must be found. I would like to thank Micah for being an excellent guide, mentor and doctor-father. He has supported and suffered me through these years, pushing and pulling me when I needed assistance (be it be day or electronically by night). Also, I was lucky to be supported by such an incredible mentor, who promoted me within the scientific community, introducing me to his colleagues (for example Durk Talsma, a fellow alienist and helpful scientific collaborator) and paving my way into the academic world.

Second, a student must be taught his/her métier and learn about the pitfalls of faulty data acquisition. I want to thank Céline Cappe, who taught me how to acquire and log data during the initial phases of my stay within the lab. Further, I would like to thank the physics & maths squad, who promptly identified my lack in proper education in programming, signal processing and bio-physical models of electrical source estimations. Throughout the years, Marzia De Lucia, Athina Tzovara, Jean-François Knebel, Cristian Carmeli, Jéremy Lefebvre and Férath Kérif have assisted me with their programming skills and their willingness to convey knowledge about physical and mathematical phenomena to a psychology student. Similarly, Giovanni Battistella and Domenica Bueti have repeatedly confronted me to the miraculous world of magnetic resonance imaging techniques.

Third, in order to cope with stress and delusions of the scientific world, a student necessitates a group of people surrounding him/her who reminds them of the existence of a world outside of the lab. This was brilliantly done by Giovanni Battistella and Sébastien Tourbier who have always been excellent bench neighbors, knowing when it was time to log off the computer and exchange the mouse against a refreshing drink. At these occasions, our jolly gatherings would be crashed by others I haven't mentioned yet, and who have all contributed to making my stay in the lab a very pleasant 4 and a half years. Thank you to: Ulrike (who has always known how to organize a good party!), Marie-Laure, Jacques and Jeff; Carlos, Meri, Matthias, Gabri, Ruud, the Jérôme's, Hélène and Andrew; the EPFL squad, especially Tobias; the Aquarium folks and Léna, and all the others I have not mentioned!!

Forth, and most importantly, a student must be as lucky as to find an understanding "significant other" as well as a family with expertise in cheerleading and moral support to rely on when times are dark, and stress levels are high. I found this support in Simone Coppo and my family and thank them all for not dumping me in the streets when I was a bothersome lunatic.

Abstract (*English*)

We perceive our environment through multiple sensory channels. Nonetheless, research has traditionally focused on the investigation of sensory processing within single modalities. Thus, investigating how our brain integrates multisensory information is of crucial importance for understanding how organisms cope with a constantly changing and dynamic environment. During my thesis I have investigated how multisensory events impact our perception and brain responses, either when auditory-visual stimuli were presented simultaneously or how multisensory events at one point in time impact later unisensory processing.

In "Looming signals reveal synergistic principles of multisensory integration" (Cappe, Thelen et al., 2012) we investigated the neuronal substrates involved in motion detection in depth under multisensory vs. unisensory conditions. We have shown that congruent auditory-visual looming (i.e. approaching) signals are preferentially integrated by the brain. Further, we show that early effects under these conditions are relevant for behavior, effectively speeding up responses to these combined stimulus presentations.

In "Electrical neuroimaging of memory discrimination based on single-trial multisensory learning" (Thelen et al., 2012), we investigated the behavioral impact of single encounters with meaningless auditory-visual object parings upon subsequent visual object recognition. In addition to showing that these encounters lead to impaired recognition accuracy upon repeated visual presentations, we have shown that the brain discriminates images as soon as ~100ms post-stimulus onset according to the initial encounter context.

In "Single-trial multisensory memories affect later visual and auditory object recognition" (Thelen et al., in review) we have addressed whether auditory object recognition is affected by single-trial multisensory memories, and whether recognition accuracy of sounds was similarly affected by the initial encounter context as visual objects. We found that this is in fact the case. We propose that a common underlying brain network is differentially involved during encoding and retrieval of images and sounds based on our behavioral findings.

Abstract (*French*)

Nous percevons l’environnement qui nous entoure à l'aide de plusieurs organes sensoriels. Antérieurement, la recherche sur la perception s'est focalisée sur l'étude des systèmes sensoriels indépendamment les uns des autres. Cependant, l'étude des processus cérébraux qui soutiennent l'intégration de l’information multisensorielle est d’une importance cruciale pour comprendre comment notre cerveau travail en réponse à un monde dynamique en perpétuel changement. Pendant ma thèse, j'ai ainsi étudié comment des évènements multisensoriels impactent notre perception immédiate et/ou ultérieure et comment ils sont traités par notre cerveau.

Dans l'étude " Looming signals reveal synergistic principles of multisensory integration" (Cappe, Thelen et al., 2012), nous nous sommes intéressés aux processus neuronaux impliqués dans la détection de mouvements à l’aide de l’utilisation de stimuli audio-visuels seuls ou combinés. Nos résultats ont montré que notre cerveau intègre de manière préférentielle des stimuli audio-visuels combinés s’approchant de l'observateur. De plus, nous avons montré que des effets précoces, observés au niveau de la réponse cérébrale, influencent notre comportement, en accélérant la détection de ces stimuli.

Dans l'étude "Electrical neuroimaging of memory discrimination based on single-trial multisensory learning" (Thelen et al., 2012), nous nous sommes intéressés à l’impact qu’a la présentation d’un stimulus audio-visuel sur l’exactitude de reconnaissance d'une image. Nous avons étudié comment la présentation d'une combinaison audio-visuelle sans signification, impacte, au niveau comportementale et cérébral, sur la reconnaissance ultérieure de l’image. Les résultats ont montré que l’exactitude de la reconnaissance d’images, présentées dans le passé, avec un son sans signification, est inférieure à celle obtenue dans le cas d’images présentées seules. De plus, notre cerveau différencie ces deux types de stimuli très tôt dans le traitement d'images.

Dans l'étude "Single-trial multisensory memories affect later visual and auditory object recognition" (Thelen et al., in review), nous nous sommes posés la question si l’exactitude de la reconnaissance de sons était affectée de manière semblable par la présentation d'évènements multisensoriels passés. Ceci a été vérifié par nos résultats. Nous avons proposé que cette similitude puisse être expliquée par le recrutement différentiel d'un réseau neuronal commun.

Table of Contents

1 Introduction or Why study multisensory integration?

Research on the mechanisms of perception has classically focused on single sensory systems. Nonetheless, our perception of the world is fundamentally multisensory. The brain continuously integrates physical information (light, sound, touch, odors, and taste) which is captured by unique sensory organs and translated into a coherent percept of the outside world.

Although multisensory information is generally beneficial for perception, conveying redundant and/or complementary information about our environment; it can also give rise to illusory perceptions. For example, in a noisy environment, such as a crowded party, congruent information gathered by lip reading can improve the comprehension of the auditory stream of a conversation (Bernstein, Auer, & Takayanagi, 2004). On the other hand, incongruent streams of visual and auditory information can lead to illusory or wrong percepts. Classical examples of these illusions are the 'Ventriloquist Effect', where the sounds are mis-located towards simultaneously and correlated, but spatially disparate visual events (Bertelson, 1999) and the 'McGurk Effect' (McGurk & MacDonald, 1976), where the perception of speech sounds is affected by concurrent lip reading, leading to faulty speech perception. In addition, discordant multisensory information can also impact the perception of more rudimentary stimuli. Shams and colleagues have reported the illusory percept of a second visual flash when coupling the rapid presentation of two auditory beeps to the presentation of a single flash (Shams, Kamitani, & Shimojo, 2002; Shams, Kamitani, & Shimojo, 2000). On the other hand also the contrary is true, i.e. that the rapid presentation of two visual flashes coupled to the presentation of either none, or a single auditory stimulus leads to the perception of a single 'fused' visual event (Bertelson & Radeau, 1981).

As mentioned, multisensory events can also enhance/facilitate perception and behavior. Imagine you are an ornithologist, walking through a forest. Hearing a birds’ song can help to visually locate the animal in the thicket of leaves. Similarly, the roar of an approaching car informs about the direction to which to orientate your gaze, effectively speeding up the detection of the car and determining its’ approaching speed (danger evaluation). For instance, several studies have shown faster and/or more accurate responses to occur under multisensory conditions as compared either unisensory condition presented on their own (Otto, Dassy, & Mamassian, 2013; Raab, 1962). Further, the speeded responses observed under multisensory conditions exceed the probability summation of responding to either of its unisensory components alone (Miller, 1986, 1991). More precisely, the 'redundant signal effect' (RSE), i.e. faster responses to simultaneously presented multisensory stimuli, has been shown to violate the 'separate-decision' or 'race model'. The race model assumes that independent perceptual codes race on independent channels to elicit a response (Raab, 1962). Such speeded responses are observed in a variety of tasks, such as simple stimulus detection and localization (Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008). Van der Burg and colleagues have investigated the neuronal mechanisms involved in the so called “pip and pop” effect. This effect refers to speeded detection of a visual target reported under multisensory conditions as compared to unisensory conditions, even when the sound does not convey spatial information. In terms of accuracy, stimulus discrimination has been shown to be enhanced, when presenting auditory-visual pairs, as compared to presenting either unisensory component alone (Giard & Peronnet, 1999). In their study the authors asked subjects to discriminate between two objects composed of a visual and an auditory feature. The results showed that object discrimination was significantly faster and more accurate under multisensory as compared to unisensory conditions.

We live in dynamic world, where objects which we encounter in a multisensory context at one point time might be presented to us only in a unisensory way at a later point in time (for example in a picture). Several studies have investigated how multisensory information can impact later unisensory perception. For example, meeting someone in person can facilitate later speaker identification on the phone. Thus, in addition to studies investigating how multisensory cues interact and impact behavioral and/or brain responses when sensory cues in two or more different modalities are simultaneously presented, others have focused on how unisensory stimulus processing and perceptual learning are affected by prior multisensory experiences (Gottfried, Smith, Rugg, & Dolan, 2004; Meylan & Murray, 2007; Nyberg, Habib, McIntosh, & Tulving, 2000; Shams & Seitz, 2008; Shams, Wozny, Kim, & Seitz, 2011; van der Linden, van Turennout, & Indefrey, 2010; von Kriegstein & Giraud, 2006; Wheeler, Petersen, & Buckner, 2000). Generally, these studies have shown that exposure to multisensory events can affect subsequent unisensory processing over a variety of stimulus features presented, ranging from rudimentary stimulus sets (Giard & Peronnet, 1999; Meylan & Murray, 2007) to complex and ethologic stimulus sets (van der Linden, van Turennout, & Fernández, 2011; von Kriegstein & Giraud, 2006). Further, these studies have shown effects to occur both immediately following the multisensory event (Meylan & Murray, 2007) as well as up to more than a day (Wheeler et al., 2000; Zangenehpour & Zatorre, 2010). Such findings, suggest that multisensory memory traces might have an important advantage over unisensory training protocols for memory rehabilitation.

Multisensory integration has also been studied in the framework of sensory substitution devices (SSDs) (Haigh, Brown, Meijer, & Proulx, 2013; Proulx, Brown, Pasqualotto, & Meijer, 2012; Striem-Amit, Cohen, Dehaene, & Amedi, 2012). The most known sensory substitution device is the vOICe (Oh! I see!) (http://www.seeingwithsound.com; Amedi et al., 2007; Bach-y-rita & Kercel, 2003; Merabet et al., 2009) although other devices such as the Prosthesis Substituting Vision for Audition (PSVA) and "the vibe" should also be mentioned (Auvray, Hanneton, Lenay, & O'Regan, 2005; Capelle, Trullemans, Arno, & Veraart, 1998). Other SSDs have focused on tactile-to-visual substitution (TVSS, Bach-y-Rita, Collins, Saunders, White, & Scadden, 1969). Generally, these systems convert information from the deficient sensory modality (vision), into soundscapes or tactile stimulations (fully functional sensory modality)(see Figure 1). The advantage of such devices is that they are non-invasive, and their use is reasonably easy-to-learn (for a review see: Reich, Maidenbaum, & Amedi, 2012).

Another application of multisensory research has been reported in training protocols aiming at diminishing auditory-visual integration deficits observed in clinical populations (for example in dyslexia: Powers, Hillock, & Wallace, 2009; Temple et al., 2003). Such research has shown that multisensory processing is susceptible to training-induced plasticity. Thus, the investigation of the efficacy of training protocols and the identification of the underlying neuronal mechanisms involved has been the focus of recent studies (see: Powers, Hevey, & Wallace, 2012).

Overall, understanding the neuronal substrates involved and the organizational principles governing multisensory integration, as well as their susceptibility to training-induced plasticity is of crucial importance for understanding how these interactions shape the perception of the world in healthy subjects and how we can improve deficient sensory processing in clinical populations.

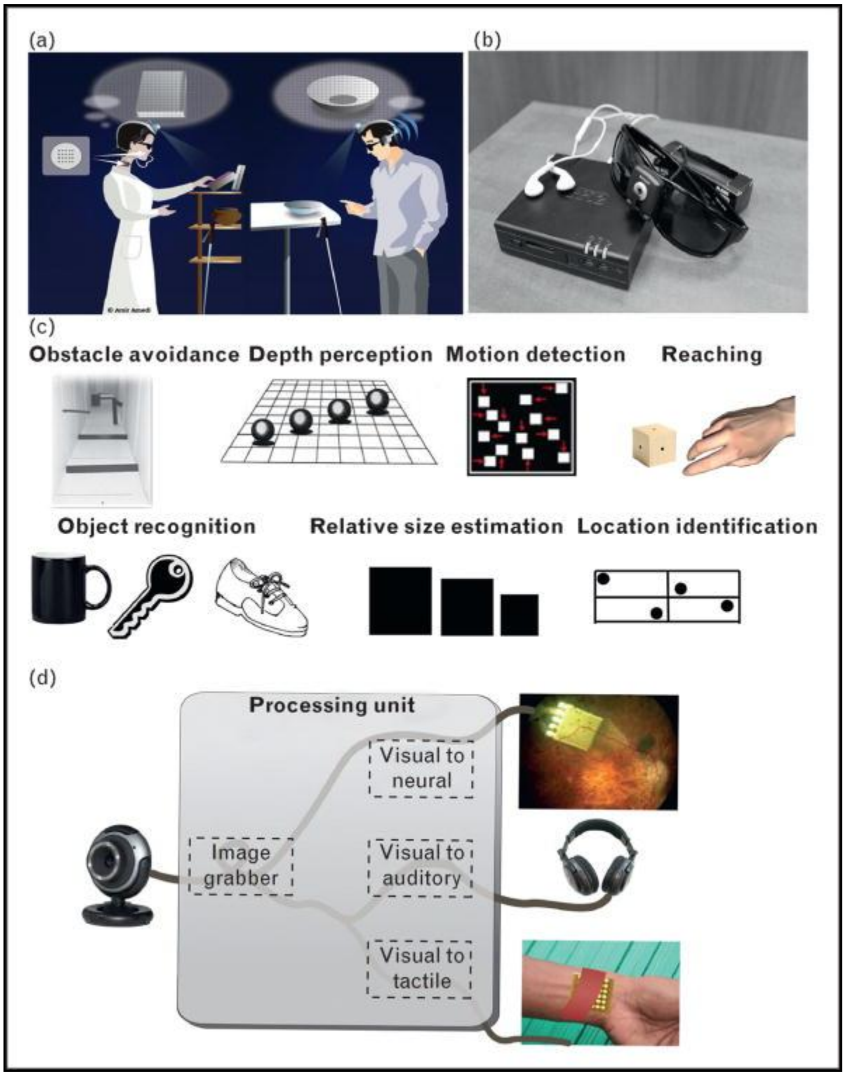


Figure 1: An overview of different SSDs for patients with visual impairment. (a) An example of visual-to-somatosensory and visual-to-auditory SSDs. (b) Equipment of a portable visual-to-auditory SSD. A webcam mounted on eyeglasses conveys visual information to a computing device and headphones. (c) Example of different tasks that were successfully performed with current SSDs. (d) Example of an invasive retinal prosthesis combined with SSDs. Visual information is captured by the webcam and conveyed to the processing unit. The information is then converted to auditory and tactile stimulations. In addition, visual information can also be conveyed to retinal prosthesis electrodes during rehabilitation of cortical function after prolonged blindness. (Figure taken from Reich, et al., (2012)).

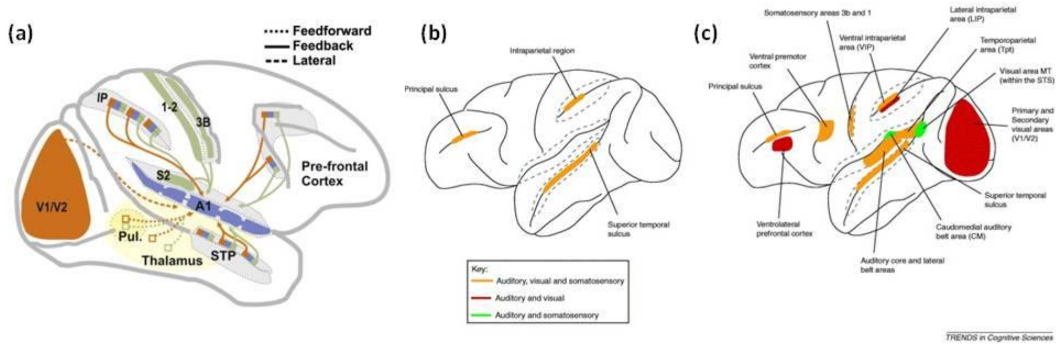
2 Anatomical substrate of multisensory integration

2.1 Integration within low-level cortices

2.1.1 Evidence from anatomical studies

Whenever you open a textbook, sensory processing is described as purely hierarchical. More precisely, sensory input is thought to be independently processed in specialized cortical areas (visual input in visual cortices, auditory input in auditory cortices, etc.), and information is combined only in higher-level parietal and frontal regions. This model has been based on the findings of Dr. Wilder Penfield (1891 - 1976) and Jones and Powell (1970). Penfield was a neurosurgeon specialized in epilepsy. During his operations he stimulated cortical areas in connection with either motor responses or reports of evoked sensory perceptions of patients. Thus, Penfield has become the most famous neurocartographer, and his cortical (motor and sensory) homunculi are still in use today. Penfield postulated that sensory cortices are unisensory, since patients never reported alternative sensory evoked perceptions when he applied cortical stimulation to a given cortical sensory area. Later, Jones and Powell (1970), further supported this model, stating that in the cerebral cortex of the monkey primary sensory cortices were not interconnected. These authors further postulated that only parts from the temporal, frontal and parietal lobes are concerned with multisensory integration (Jones & Powell, 1970).

A substantial body of work has challenged this view, demonstrating multisensory interactions to occur also (and already) at early latencies and within low-level cortices (for a review see (Ghazanfar & Schroeder, 2006; Murray, Cappe, Romei, Martuzzi, & Thut, 2012). Evidence supporting the model postulating the existence of heteromodal/multisensory connections between different cortical sensory areas has been collected and replicated repeatedly by anatomical studies in monkeys, ferrets, rats and cats (Barth, Goldberg, Brett, & Di, 1995; Bizley, Nodal, Bajo, Nelken, & King, 2007; Cappe, Rouiller, & Barone, 2009; Innocenti, Berbel, & Clarke, 1988; Miller & Vogt, 1984; Schroeder et al., 2001; Watanabe & Iwai, 1991)(see Figure 2). Additionally, recent studies have shown that multisensory interactions are also gated by thalamic interconnections, suggesting that multisensory convergence already takes place at pre-cortical areas, and that part of the information that is transmitted to the low-level sensory areas has already been merged within subcortical structures (Budinger, Laszcz, Lison, Scheich, & Ohl, 2008; Cappe, Morel, & Rouiller, 2007; Hackett et al., 2007). The thalamus is considered to be a primary candidate for such early integration sites, due to its strong ascending connections from several peripheral sensory inputs towards cortical areas, in addition to gating descending motor outputs towards the spinal cord. The thalamo-cortical connections have been seen as feedback projections in classical models, by which cortical areas control thalamic nuclei from which they receive ascending inputs. However, recent studies have shown the existence of feedforward connections originating in thalamic nuclei (Sherman, 2007). Further evidence for these feedforward projections, arises from anatomical studies labeling neurons with anterograde and retrograde tracers in monkeys (Cappe et al., 2007). In a more recent study Cappe and colleagues (2009), found overlaps between thalamo-cortical connections (input to cortices) and cortico-thalamic connections (output from cortical areas) in the thalamus (Cappe, Morel, Barone, & Rouiller, 2009). These results suggest that the thalamus, not only conveys multisensory input towards cortical areas, but also plays a crucial role in sensorimotor integration and behavioral responses. Thus, in addition to multisensory integration sites in cortical areas of the brain, possible contributions from subcortical regions must be taken into account (see Figure 3).



TRENDS in Cognitive Sciences

Figure 2: Schematic view of the macaque brain. (a) Example of multisensory projections toward unisensory cortex. Representation of non-auditory sources and their feedforward (dashed lines) and feedback from non-auditory (solid lines) projections to auditory cortex. Colors represent visual (orange), somatosensory (green) and auditory (blue) sensory representations. (Taken from Musacchia & Schroeder, 2009). (b) Schematic representations of the traditional view of multisensory convergence areas in the primate brain. (c) Modern view of cortical loci of multisensory integration. Colors represent regions were anatomical and electrophysiological data have demonstrated evidence for multisensory interactions. (Taken from Ghazanfar & Schroeder, 2006)

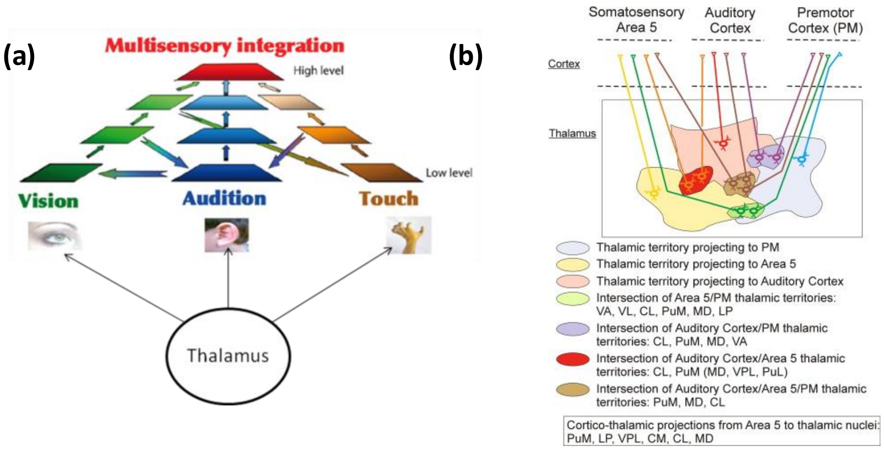


Figure 3: Multisensory convergence pathways within subcortical structures. (a) A schematic representation of thalamic influences on low-level sensory areas. Further, multisensory cortico-cortical connections between different low-level sensory areas are schematized. (b) Origin of unisensory and multisensory thalamo-cortical projections towards low-level sensory areas. (Figure taken from Cappe, Rouiller, et al., 2009; modified with permission of C. Cappe)

Whether this anatomical model can be directly transposed to humans has not yet been established. Recent neuroimaging findings (diffusion tensor imaging ; DTI) would suggest that this is the case (Beer, Plank, & Greenlee, 2011; Beer, Plank, Meyer, & Greenlee, 2013; van den Brink et al., 2013)**.** Beer et al. (2011; 2013) found fiber tracks originating in the planum temporale terminating within occipital regions (see Figure 4). On the other hand, Van den Brink et al. (2013) found that the strength of anatomical connections between sub-cortical auditory regions and primary auditory cortices was predictive of the behavioral benefit observed after multisensory events in a visual search task. Despite these recent efforts the functional properties of these connections and their behavioral relevance need to be addressed more exhaustively in humans.

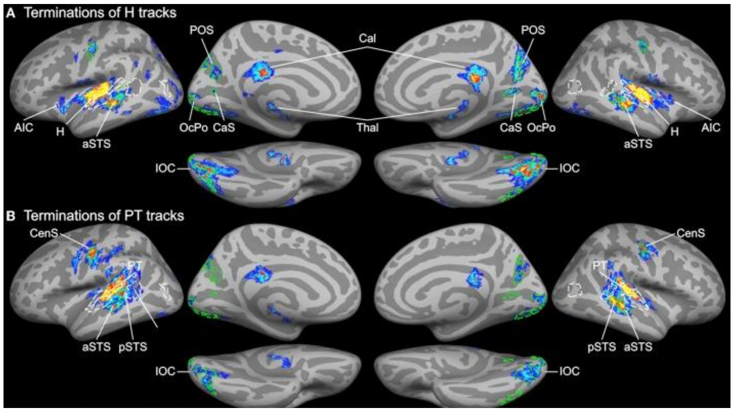


Figure 4: Diffusion tensor imaging (DTI) results. Termination maps of fiber tracks found by seding within (a) Heschls' region (H) and (b) the planum temporale (PT). Colors denote whether terminations were found in 3 (blue) or 10 (yellow) out of 10 subjects. (Taken from: Beer et al., 2013)

Generally, these questions can be addressed in behavioral paradigms in conjunction with neurophysiological imaging approaches. On the one hand, varying the stimulus features presented during the tasks can be informative of the underlying neuronal substrates which are involved. The connections described in the neuroanatomical models, go from posterior auditory regions (known to be involved in processing complex auditory stimuli as compared to the core region, involved in pitch processing of auditory cues) toward peripheral regions of the visual V2 cortices and prostriata areas and would suggest perceptual enhancement to occur for stimuli presented within the peripheral visual field (as compared to (para)foveal visual presentations) (Falchier et al., 2010; Musacchia & Schroeder, 2009). Behaviorally, this issue can be investigated by combining auditory stimuli with visual stimuli varying in spatial position, taking advantage of the retinotopic organization of the visual cortex (eccentricity manipulation). Further, in order to address the question of the origin of these connections within the auditory regions, pitch and bandwidth manipulations are informative (distinguishing between core and belt areas of primary auditory cortices) (see Spierer, Manuel, Bueti, & Murray, 2013).

In addition to varying the physical properties of the stimuli, functional connectivity between low-level auditory and visual areas can be addressed by transcranial magnetic stimulation (TMS). TMS is thought to impact ongoing neuronal activity within near-surface cortical regions situated directly underneath the stimulated scalp region. Multisensory studies have taken advantage of the excitatory TMS effect observed within primary visual areas, which induces the perception of phosphenes in subjects. A phosphene is a perceived flash of light in the absence of a physical visual stimulation after single-pulse TMS over occipital regions, and represents a measure of cortical excitability. For example, Romei et al. (2009) found that phosphene perception was enhanced when TMS-pulses where coupled with auditory stimulus presentations (Romei, Murray, Cappe, & Thut, 2009). Combining TMS and variations in the physical stimulus properties described above provides insights into the neuronal basis and functional relevance of auditory-visual integration. In a recent TMS study Spierer and colleagues (2013) have addressed this question directly (Spierer et al., 2013). TMS-pulses over occipital cortex were paired with sounds varying in center frequency (high vs. low) and bandwidth (narrow vs. broad). The results showed that phosphene perception was significantly enhanced when TMS pulses were paired with high vs. low frequency sounds, and for narrow vs. broad-band sounds. The monosynaptic projections toward low-level visual cortices in monkeys have been found to originate the caudal regions auditory cortices (Smiley & Falchier, 2009). These regions have been shown to preferentially respond to broadband stimuli compared to core regions of primary auditory cortex (Rauschecker & Tian, 2004). Strikingly, Spierer and colleagues (2013) found greater enhancement of phosphene perceptions to occur when narrow-band sounds were presented. These findings illustrate well the exiting gap between animal models and their direct application to humans.

Evidence for multisensory interactions taking place in early and low-level cortices in humans has also been found in functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) studies. The former imaging technique measures fluctuations in the concentration of oxy- vs. deoxy-hemoglobin within cortical areas. The Blood-oxygen-level-dependent measure (BOLD) can reveal cortical areas implicated in a given task. Further, fMRI has a high spatial resolution, revealing the loci of multisensory interactions. For example, Martuzzi and colleagues (2007) asked subjects to perform a simple detection task in the scanner. The functional imaging results showed visual cortex responses to auditory stimuli and auditory cortex responses to visual stimuli (Martuzzi et al., 2007). Moreover, the authors directly addressed the technical limitations inherent to BOLD imaging. The hemodynamic response is relatively slow (2 - 14 seconds) compared to processing speed of the nervous system (Kim, Richter, & Uğurbil, 1997). Martuzzi and colleagues approached this limitation by jittering the acquisition of the BOLD signal with respect to the stimulus onset upon each trial. This leads to a BOLD responses being sampled every 200ms. Consequently, the data were able to reveal response latency shifts within primary and peri-primary sensory areas under multisensory vs. unisensory conditions. The results demonstrated multisensory interactions to occur within low-level sensory areas and that these responses occur earlier in time under multisensory conditions. Several other studies have investigated multisensory interactions by functional hemodynamic neuroimaging techniques (some examples: Amedi, Jacobson, Hendler, Malach, & Zohary, 2002; Atteveldt, Formisano, Goebel, & Blomert, 2004; Bueti & Macaluso, 2010; Zangenehpour & Zatorre, 2010; for a general review see: Driver & Noesselt, 2008). These studies have revealed multisensory interactions within primary sensory cortices under a variety of tasks and stimulus presentations. Amedi and colleagues (2002) found activations within visual cortices elicited by haptic exploration of objects. Similarly, Van Atteveldt and colleagues (2004) found that activity within auditory cortices upon presentation of speech sounds was modulated by concurrently presented (written) letters. Bueti and Macaluso (2010) investigated the effect of temporal expectation upon cortical activation. Their results showed that activity within visual cortex was modulated by the anticipation of an up-coming auditory event (i.e. when viewing a visual sequence of hand-clapping). Zangenehpour and Zatorre (2010) showed that brief exposure to combined auditory-visual stimulus pairs produced visual cortex activations upon presentations of sounds ~45min and ~1day after the combined stimuli had been presented (see also: Driver & Noesselt, 2008).

2.1.2 Evidence from electrophysiological studies

Next to anatomical studies, evidence for early multisensory processing comes from electrophysiological studies in animals and humans. Electrophysiology has the advantage of high temporal resolution, and investigations in animal models can inform about the laminar dynamics of stimulus processing. Temporal and laminar profiles of sensory processing can help us to untangle feedforward from feedback interactions occurring between (relatively) distant so-called unisensory areas.

Studies in macaque monkeys have revealed multisensory interactions to occur within auditory and visual cortices (Foxe & Schroeder, 2005; Ghazanfar & Schroeder, 2006; Kayser & Logothetis, 2007; Kayser, Petkov, & Logothetis, 2009). For example, several studies have repeatedly reported low-level integrative effects to occur within regions of the superior temporal plane. Schroeder and Foxe (2002) found that sensory processing within this area was modulated by non-auditory inputs. Microelectrodes implanted within the superior temporal plane showed temporal convergence between auditory and somatosensory to occur within granular layers (layer 4 of the cortical column), suggesting feedforward input from somatosensory to auditory cortices. On the other hand, visual input modulated activity within this area above and below layer 4, suggestive of feedback modulations (Schroeder & Foxe, 2002) (see Figure 5).

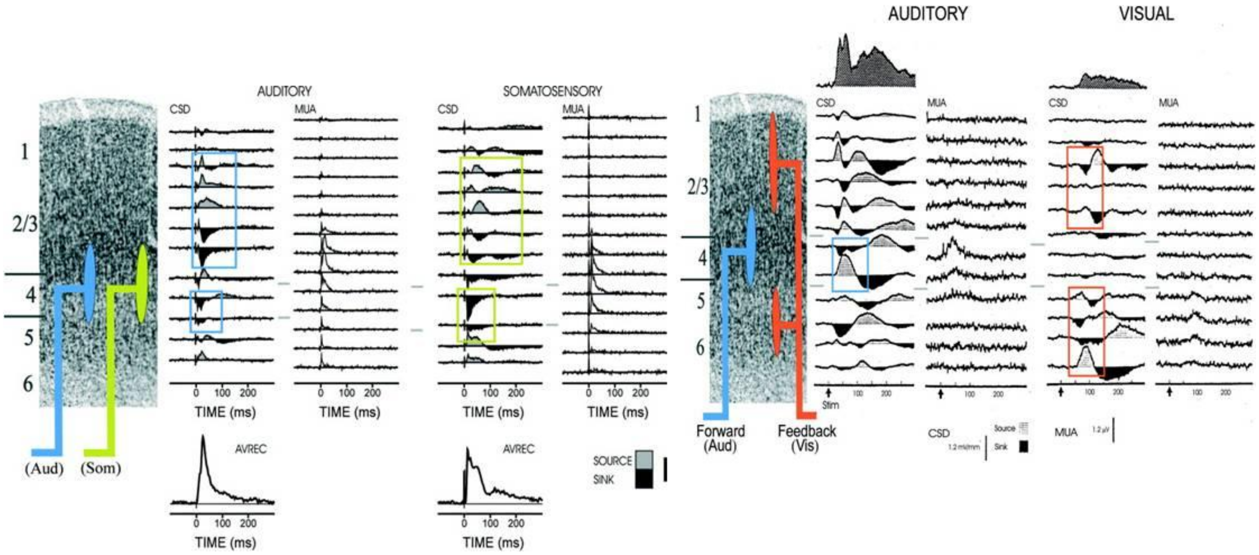


Figure 5: Laminar current source density (CSD) and multiunit activity (MUA) recorded within posterior lateral regions to primary auditory cortex. The left panel shows activity evoked by auditory (blue) and somatosensory (green) stimulations. The right panel shows activity evoked by auditory (blue) and visual (red) stimulations within the auditory cortex. Boxes highlight the initial excitatory response of the granular layer 4 and subsequent excitation of the cell ensembles within layers 2/3 and/or 5/6. The temporal and laminar profile of sensory responses reveal feedforward (auditory, somatosensory) and feedback (visual) inputs. (Taken from Schroeder & Foxe, 2002)

Importantly, some of the above-mentioned investigations have suggested that such multisensory interactions within low-level sensory cortices might be mediated by association areas, rather than being conveyed by direct connections between these regions (Kayser et al., 2009; Schroeder & Foxe, 2002). The main candidate mediating auditory-visual inputs is thought to be the superior temporal sulcus (STS). The STS has been shown to be functionally coupled to auditory cortices and has been classically seen as a multisensory convergence cortex (Kayser & Logothetis, 2009; Maier, Chandrasekaran, & Ghazanfar, 2008).

In humans, several electrophysiological studies have shown non-linear responses to multisensory stimuli to occur at early latencies of sensory processing (Cappe, Thelen, Romei, Thut, & Murray, 2012; Cappe, Thut, Romei, & Murray, 2010; Giard & Peronnet, 1999; Molholm et al., 2002; Van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011). Collectively these studies have shown auditory-visual interactions to occur within the first ~50-100ms after stimulus onset. Further, the scalp topography of the evoked responses suggested differential sensory processing under multisensory conditions which source estimations suggest occur within or near primary visual cortices.

Although non-linear responses (i.e. comparing the paired multisensory presentation to the algebraic sum of the unisensory presentations, AV ≠ A + V) has been generally accepted as evidence for multisensory integration, Teder-Sälejärvi and colleagues (2002) have criticized this approach (Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002). These authors claim that the sum of the unisensory responses is not free of "common" activity, such as anticipatory potentials and motor responses. Thus comparing the sum of such responses to a single response in the paired condition could lead to false results. A recent study by Cappe and colleagues (2010) approached the debate by investigating non-linear multisensory interactions upon stimulus presentations which did not require motor responses from the subjects. Moreover, a variable inter-stimulus-interval (ISI) ensured that anticipatory modulations of the recorded responses could be excluded (Cappe et al., 2010). The results showed that even when controlling for possible biases, non-linear responses were found at ~40-50ms post-stimulus onset within parieto-occipital regions, similar to what has been reported by van der Burg and colleagues (2011). Further, multisensory processing was shown to modulate activity within visual, auditory and posterior STS at ~60 - 95ms post-stimulus onset.

Collectively, these studies have shown that multisensory interactions occur at early stages of sensory processing and that these interactions modulate responses within primary and peri-primary sensory cortices.

2.2 Multisensory integration within higher-order cortices

In addition to evidence arguing for early interactions of sensory information, multisensory integration has been shown to take place in higher-level association cortices. Several cortical areas are considered as association sites, due to their anatomical connections with unisensory cortices (Jones & Powell, 1970).

2.2.1 The lateral occipital-temporal complex: the superior temporal sulcus (STS)

Anatomically, the superior temporal region, situated between auditory, visual and somatosensory cortices, represents an ideal candidate for multisensory integration. In fact, the STS has been repeatedly shown to respond to auditory and visual stimuli (Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005; Beauchamp, Lee, Argall, & Martin, 2004; Macaluso, George, Dolan, Spence, & Driver, 2004; van der Linden et al., 2010). More precisely, studies have shown the STS to play a crucial role in auditory-visual speech perception, in object feature binding as well as in object recognition (Atteveldt et al., 2004; Beauchamp et al., 2004; Tanabe, Honda, & Sadato, 2005; Vander Wyk et al., 2010). Further the STS seems to show a greater response to congruent meaningful stimuli, such as linguistic cues, biological movement and objects.

Calvert et al. (2000) compared cortical responses to audiovisual speech, which was presented either synchronously, asynchronously or in both unisensory conditions only (Calvert, Campbell, & Brammer, 2000). They found that when presenting multisensory synchronous stimuli the left STS showed supra-additive (AV > A + V) responses, while reduced response levels where found for asynchronous multisensory trials. When manipulating speech and language cues, Wright, et al. (2003) found that responses to visual stimuli were biggest in the posterior half of the STS, whereas the anterior part of the STS was activated only by auditory and auditory-visual stimuli (Wright, 2003).

In addition to responses to linguistic multisensory stimuli, the STS also responds to animals and man-made objects (tools)(Beauchamp et al., 2004; Beauchamp, 2005). The response enhancement observed within the STS upon multisensory vs. unisensory object presentations seems to be dependent upon the congruency between object features. In fact Dahl and colleagues (2010) reported stronger activation upon the presentation of meaningful object pairs compared to meaningless object presentations (Dahl, Logothetis, & Kayser, 2010; but see Taylor et al., 2006).

2.2.2 The Intraparietal sulcus (IP)

The intraparietal region is part of a larger network, implied in the organization of motor activity based on multisensory information. It contains spatial representations and is involved in attention and goal-directed behaviors. Neurons in the lateral IP (LIP) are known to receive multisensory inputs of eye position, in addition to inputs from visual and auditory cortices. The neuronal activity in the LIP is modulated in delayed-saccade tasks, and this independently of the type of stimulus modality (visual or auditory)(Mazzoni, Bracewell, Barash, & Andersen, 1996).

Ventral IP (VIP) neurons respond to visual, auditory, somatosensory and vestibular stimulations, where the receptive fields of different sensory modalities usually overlap in space. Most VIP cells are modulated by multisensory stimuli, but they appear to require spatial and temporal coincidence to do so, suggesting that these cells have overlapping receptive fields (RFs) (Avillac, Ben Hamed, & Duhamel, 2007). Thus, the results suggest that area VIP is involved in the representation of external space. This representation is modality-independent *per se* but the activity within VIP is modulated by multisensory congruency.

Further, the temporo-parietal junction (TPJ) has been implicated in the multisensory representation of self-body perception in humans (Blanke & Arzy, 2005; Lenggenhager, Smith, & Blanke, 2006). In fact, Ionta and colleagues (2011) found that visuo-tactile and visuo-vestibular integration within the TPJ plays a central role in self-location perception (real. vs out of body experiences) (Ionta et al., 2011).

2.2.3 Frontal and prefrontal cortices

Evidence for multisensory integration in prefrontal areas (PFC) has been collected in monkeys when presented with a delayed match-to-sample task, where an association between high or low frequency tones to two different colors had to be made (Fuster, Bodner, & Kroger, 2000). Neurons in the prefrontal cortices responded to both visual and auditory stimuli (see Figure 6). Further Romanski (2007) argued that the anterior ventrolateral prefrontal cortex (VLPFC) receives dense projections from the auditory cortex. In addition, projections from the inferotemporal sulcus carrying visual information arrive in the more posterior part of the VLPFC. This cortical region appears to play a crucial role in treating information of face and vocalization stimuli. Consequently, the connectivity of the frontal lobes makes them a likely candidate for integrating sensory signals that are related to communication (Romanski, 2007). These findings are in agreement with neuroimaging studies in humans, which show activations in the frontal lobe regions during auditory-visual integration (Calvert, Hansen, Iversen, & Brammer, 2001; Miller & D'Esposito, 2005). In addition, studies on schizophrenia patients, who are known to demonstrate deficits in communicational auditory-visual cue integration, show structural abnormalities in the frontal lobes (de Gelder et al., 2005; de Gelder, Vroomen, Annen, Masthof, & Hodiamont, 2003).

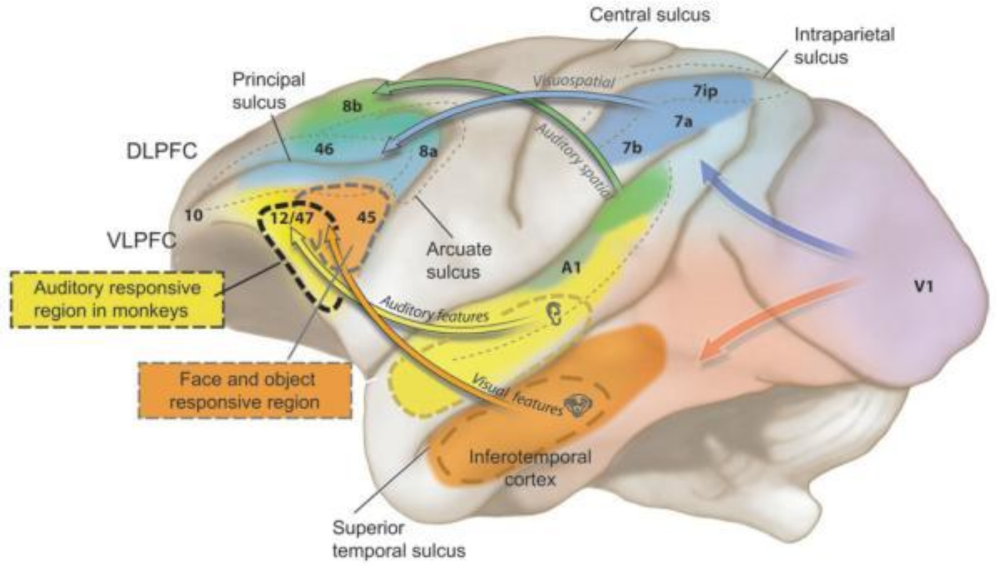


Figure 6: Schematic representation of the macaque brain, depicting the flow of auditory and visual information originating within primary sensory areas and terminating within prefrontal cortex (PFC). The colors represent visual object-related information (orange), visual spatial representations (blue); auditory object-related information (yellow), auditory spatial representation (green). Dorso-lateral PFC (DLPFC) is essential for spatial working memory. Ventro-lateral PFC (VLPFC) shows an overlap of auditory and visual object-related representations. (Taken from Romanski, 2007)

3 Mechanisms of multisensory integration

After having discussed the loci of multisensory integration (where), the question of the functional mechanisms (how?) will be discussed.

At the single-cell level, multisensory integration is defined as a statistically significant increase (response enhancement) or decrease (response depression) of the firing rate of a neuron under multisensory conditions as compared to the most efficient unisensory stimulus in isolation (Meredith & Stein, 1983; Stein & Meredith, 1993; for a review see: Stein & Stanford, 2008). In other words, the modulation of the spiking rate of a neuron is thought to either enhance or suppress the saliency of a multisensory event compared to either unisensory component. Seminal works in the superior colliculus of the cat have provided evidence for such multisensory interactions at the single cell level (Jiang, Wallace, Jiang, Vaughan, & Stein, 2001; Stein & Meredith, 1993; Wallace, Meredith, & Stein, 1998). These studies have shown differential firing rates to occur for multisensory stimuli presented within the receptive fields of neurons in this brain structure (for more details see the following section **Organizational principles of multisensory interaction**).

A way to quantify multisensory integration at the cell population level is by recording multiunit activity (MUA) and the analysis of the local field potentials (LFP), and current source densities by extension (CSD; the second spatial derivate of LFPs). MUA recordings measure the spiking rate of multiple cells, which is considered a measure of the output signal of a given neuronal population (See: Chen, Dhamala, Bollimunta, Schroeder, & Ding, 2011). LFP recordings on the other hand, reflect synchronous neuronal activity of inward and outward currents caused by action potentials at the synaptic level, and reflect activity related to information input within a tissue volume. For example, Ghazanfar and colleagues (2005) have shown that integration of voices and faces modulates LFP responses within the auditory cortex (Ghazanfar, Maier, Hoffman, & Logothetis, 2005). Integration was seen as either enhancement or depression of the peak amplitude of the LFP as compared to the LFP peak of the most effective unisensory response (multisensory enhancement index; MSI). The response enhancement vs. suppression was influenced by the relative voice presentation offset compared to the onset of the dynamic face stimulus presentation. When voices were presented within ~97ms after the onset of the dynamic visual stimulus, responses under multisensory conditions were enhanced. Contrariwise, when auditory stimulus onset was delayed to ~97-332ms after visual stimulus onset, multisensory response suppression was observed. These findings provide evidence for the impact of temporal synchrony upon multisensory response profiles (enhancement vs. depression). In another study, Lakatos and colleagues (2007) have reported evidence for somatosensory-auditory interactions within auditory cortex (Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; see also Schroeder & Foxe, 2002 for prior findings). Interestingly, these authors found that somatosensory inputs could reset ongoing oscillatory activity within the primary auditory cortex. In fact, the study set out to test the hypothesis that somatosensory inputs could enhance auditory processing by resetting the phase of ongoing activity within auditory cortex, so that incoming auditory stimuli would fall into a high-excitability phase and are amplified. This hypothesis is supported by findings of, Lakatos and colleagues (2005) who found that the oscillatory phase of the transmembrane current flow is systematically related to the spontaneous firing rate of neurons in primary auditory cortex (Lakatos et al., 2005). These results suggest that oscillatory phase describes moments of high vs. low neuronal excitability and the probability of the occurrence of action potentials. Consequently, ongoing oscillations within a neuronal ensemble have a strong impact upon the processing (amplification vs. suppression of firing rate) of incoming stimuli (reviewed in Sarko, Ghose, & Wallace, 2013; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008).

These and other studies in animals (for a review see: Ghazanfar & Schroeder, 2006) have provided evidence that interactions of neuronal oscillations across cortical areas and sensory modalities might provide an effective mechanism for multisensory integration. Collectively, these data have shown that information from other sensory areas and multisensory convergence sites can alter and shape sensory processing within low-level sensory areas. Similar to what has been observed in animals with intracranial recordings, evidence for neuronal oscillations playing a central role for multisensory integration has been found in scalp recordings (EEG/MEG) in humans. For example, Romei and colleagues (2012) showed that auditory stimuli can reset ongoing oscillatory activity with visual cortices (Romei, Gross, & Thut, 2012). In their study, they manipulated auditory-visual onset asynchrony in relation to ongoing alpha oscillations (~10Hz) within visual cortices. First, these authors showed that TMS-induced phosphene perception (perceived vs. not perceived) was affected by temporal delay between the auditory stimulus and the TMS pulse over occipital cortices. In other words, auditory stimuli presented ~100ms and ~200ms prior to the TMS-pulse delivery enhanced phosphene perception. In addition, analyses of concurrently recorded EEG data showed an enhanced phase-locking within the alpha band over auditory and parieto-occipital cortices ~50-250ms post-auditory stimulus onset.

Cortical oscillations have also been found to play a central role in auditory and auditory-visual speech perception (Giraud & Poeppel, 2012; Schroeder et al., 2008). Giraud and Poeppel have proposed that activity within the auditory cortex at rest is characterized by sustained oscillations in discrete frequency bands in the delta-theta range (1-8Hz). Speech input temporally resets oscillatory activity to specific frequencies related to phonemic and syllabic sampling within low-frequency bands (1-8Hz). Consequently, resetting within the theta frequency range is followed by a transient pause in gamma activity (25-35Hz), which thereupon is newly observed. This is followed by stronger coupling of the neuronal generators of theta and gamma activity under active speech processing. Gamma-band oscillations are thought to control the neuronal excitability involved in the propagation of speech information from primary auditory cortex to higher-order processing areas. Thus, the onset of active "tracking" of speech within the delta-theta range followed by "nested" gamma activity facilitates the propagation of speech-related information within cortical areas. Similarly, Schroeder and colleagues (2008) have proposed that the "nesting" of higher-order frequencies could optimize the processing of vocalizations in monkeys (Schroeder et al., 2008). Mouth and hand movements (visual input) related to speech utterances are observed to follow a temporal rhythm of 1-3Hz, and onset prior to generation of vocalizations (i.e. you see the mouth movement before you hear the utterance). This onset jitter between visual and auditory information enables visual inputs to modulate ongoing theta-band activity within auditory cortex, aligning incoming auditory stimuli to the ideal excitability phase of theta oscillations. Thus, visual input resets ongoing activity within auditory cortex, explaining speech perception facilitation under auditory-visual as compared to auditory context (Zion Golumbic, Cogan, Schroeder, & Poeppel, 2013) (see Figure 7).

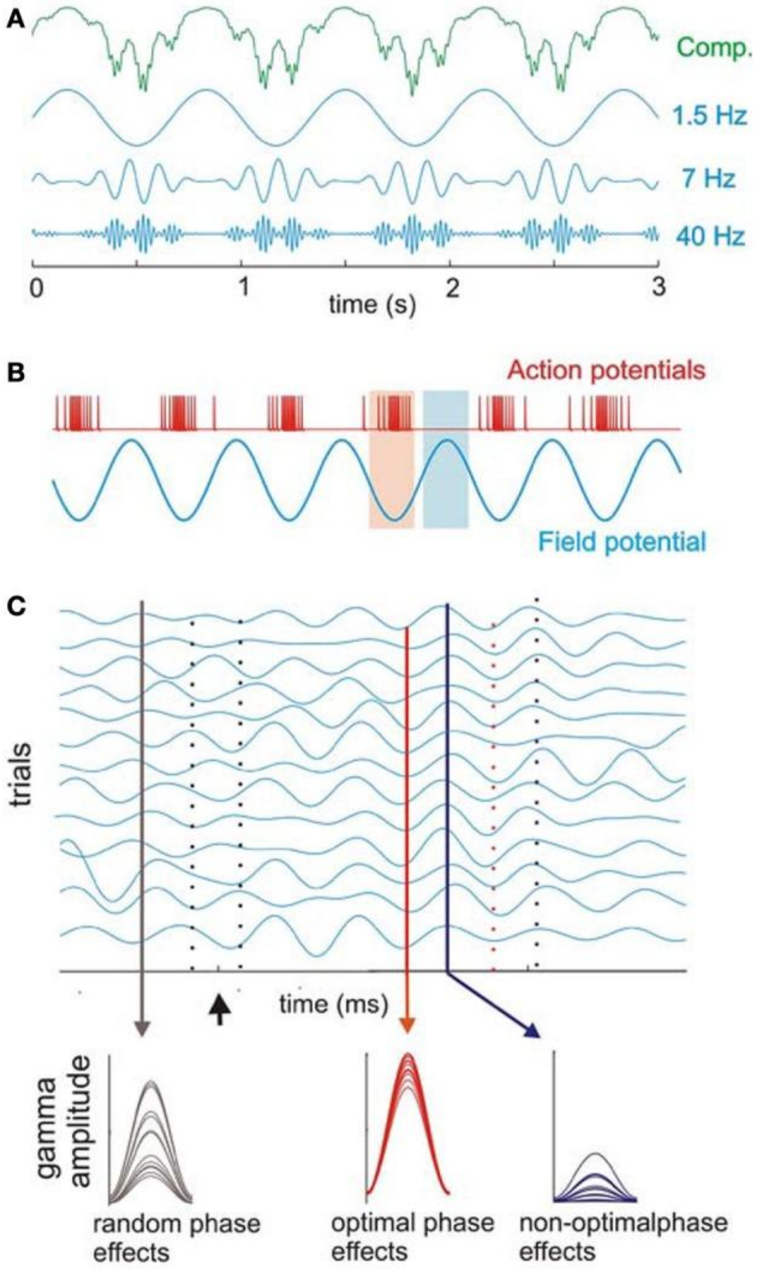


Figure 7: Schematic illustration of oscillatory phase-related modulations in neuronal excitability and cross-frequency coupling. (A) Electrophysiological signal recorded within the brain (green). Oscillatory activity can be decomposed into a mixture of components oscillating a different frequencies (blue lines). Note the strong phase-amplitude coupling between these components, i.e. the "nesting" of the oscillatory activity within higher frequencies upon the phase of slow-wave oscillations. (B) Depicts the relationship between neuronal excitability, measured in the number of spikes (red), and the phase of the ongoing oscillation in the neural ensemble (blue). The boxes indicate periods of high excitability (red) and low-excitability (blue). (C) The blue lines simulate single-trial responses within the auditory cortex. At rest, activity within the auditory cortex is characterized by high trial-by-trial phase variability (grey). The presentation of a visual stimulus (black arrow) can cause phase rest of ongoing oscillations, leading to strong phase-coherence across trials. Consequently, we observe separate alignment of optimal (red) and non-optimal (blue) periods for stimulus interactions.

In addition, to evidence of direct interactions of multisensory integration within low-level sensory areas, other studies have concentrated in how multisensory features are bound into a coherent percept. For example, Yuval-Greenberg and Deouell (2007) have emphasized the role high-frequency oscillatory activity (gamma band responses; GBR >30Hz) in object feature binding (Yuval-Greenberg & Deouell, 2007). In their study, subjects were asked to identify objects within the auditory or visual modality. Auditory-visual stimuli presented to the subjects could entail semantically congruent vs. incongruent multisensory pairs. The results showed that while there was no difference in terms to evoked GBR (eGBR) according to semantic contingencies, induced GBR (iGBR) was greater for congruent vs. incongruent trials. While the eGBR is linked to stimulus-onset, iGBR is thought to represent semantic processing, feature binding and ultimately object representation of the multisensory stimulus pair. This hypothesis is further supported by results of visual-to-auditory priming paradigm (Schneider, Debener, Oostenveld, & Engel, 2008), where GBR were enhanced upon trials of succeeding congruent vs. incongruent auditory-visual stimuli.

Generally, these data provide evidence that neuronal oscillations play a central role in the mechanisms through which multisensory information is bound at the population level and across (relatively) distant cortical areas. By providing a rhythmic substrate of activity within different frequency bands, oscillatory activity can act as a gate between sensory areas, suppressing the interaction of non-related sensory inputs by desynchronization and phase-jittering. On the other hand, neuronal oscillations can enhance perception, through synchronization, phase-resetting and neuronal coherence (for a recent review see: Sarko, Ghose, & Wallace, 2013)

4 Organizational principles of multisensory integration

Current multisensory research is based on a body of work that has been provided by Stein and colleagues (Stein & Meredith, 1993). Stein and colleagues studied (single) neurons in the superior colliculus (SC), a structure in the brainstem, which is known to be involved in visual-motor coordination. In the cat this structure supports its ability to orientate its gaze in direction of an auditory and/or somatosensory stimulus, or vice-versa. This structure receives ascending and descending information from visual, somatosensory, auditory and motor inputs, and cells in the SC appear to be multisensory. Consequently, this structure provided a model for investigating multisensory processes and its governing principles.

In order to introduce the principles put forth by Stein & Meredith (1993), we must clarify that these rules emerged from observations of changes in the firing rate of single neurons, i.e. response enhancement or depression. Further sensory neurons respond to stimuli occurring in their receptive fields (RFs). A RF is a region in space (i.e. environment or body surface) in which the presence of a stimulus will alter the firing rate of the neuron. These RFs for different sensory modalities of a multisensory neuron generally overlap, so that it responds to stimuli from the same region in space. Across many cells in the SC, their RFs are arranged to provide a functional map of the outside world (Stein & Meredith, 1993).

One of these organizational principles states that multisensory enhancement is observed for stimuli originating in close spatial proximity, whereas spatially disparate stimuli lead to response depression or no interaction (*spatial rule*). Whether this rule can be directly transposed to more complex neuronal networks or even perception and cognition is debated (for a recent review see Charles Spence, 2013). For instance, investigation of somatosensory-visual interactions have provided evidence for multisensory facilitation (speeded RTs) to occur irrespective of spatial overlap between stimuli (Sperdin, Cappe, & Murray, 2010; Tajadura-Jiménez et al., 2009; Zampini, Torresan, Spence, & Murray, 2007). Similarly, recent work has suggested that auditory enhancement observed at the behavioral level in a visual search task occurs in the absence of spatial overlap between the stimuli (van den Brink et al., 2013; Van der Burg et al., 2008). Similarly, Teder-Sälejärvi and colleagues (2007) have shown that spatial congruity between visual and auditory stimuli did not impact behavioral measures in terms of detection rate and response speed (Teder-Sälejärvi, Di Russo, McDonald, & Hillyard, 2005). Nonetheless, these authors have found that event-related potentials (ERPs) differed according to spatial congruity in terms latency and amplitude at early sensory processing stages within occipito-temporal and superior temporal regions. Thus, these studies suggest that auditory enhancement of visual perception at the behavioral level occurs in the absence of spatial congruity between the stimuli, but affects the underlying neuronal response. Contrariwise, Bolognini and colleagues (2010) have investigated the spatial constraints of auditory enhancement of visual cortex excitability, as measured by transcranial magnetic stimulation (TMS) induced phosphene perception (Bolognini, Senna, Maravita, Pascual-Leone, & Merabet, 2010). The results showed that subjects reported significantly more phosphenes, when the auditory stimulus was presented in the same spatial location as the reported phosphene within the peripheral visual field. Likewise, Leo and colleagues (2011) found visual orientation discrimination sensitivity was enhanced when looming auditory sounds were presented within the same hemifield vs. the opposite hemifield as the visual stimulus (Leo, Romei, Freeman, Ladavas, & Driver, 2011). Generally the differential findings have been explained by the task-demands at hand. While redundant signal effects (RSE) occur when subjects are performing a non-spatial task, “spatial congruity” effects are observed when subjects are asked to overtly shift their attention to a specific location in space, and when a orienting response toward the stimulus location is required (note that the SC is a brainstem structure involved in exactly those types of orienting behaviors).

Another “rule of multisensory integration” states that stimuli are integrated within certain “temporal binding window” (TBW) rather than being dependent on the exact overlap of their physical onset (*temporal rule*). The existence of the TBW is thought to deal with differences in the transmission time in the environment of sensory information (e.g. light travels faster than sound). At the cell population level this is reflected by multisensory enhancement vs. suppression observed at small vs. big stimulus onset asynchronies (SOA) between auditory and visual stimuli (see: Ghazanfar et al., 2005). Interestingly, similarly to what has been observed for the spatial rule, task-dependent as well as stimulus-dependent differences have also been found to impact the efficacy of the TBW. Most strikingly, a recent study has shown that the differences is due to whether subjects are asked to judge the common source (simultaneity judgment) of auditory-visual stimuli or rather discriminate the temporal synchrony vs. asynchrony of these stimuli (temporal order judgment) (Stevenson & Wallace, 2013).

The third principle of multisensory integration states, that response enhancement under multisensory as compared to unisensory presentations is strongest when either unisensory stimuli in isolation elicit weak responses (*principle of inverse effectiveness*; PoIE). Although this principle is commonly used to demonstrate multisensory integration at the cellular level, transposing this rule to behavioral responses has been debated (Holmes, 2007; Otto et al., 2013). Similarly, we have found selective enhancement for auditory-visual looming stimuli compared to receding stimuli (Cappe et al., 2012; Cappe, Thut, Romei, & Murray, 2009). Looming stimuli dynamically enhance their effectiveness, whereas the contrary is true for receding stimuli. Thus our results are somewhat in contradiction with the PoIE, showing greater multisensory facilitation effects to occur for behaviorally more effective stimuli.

In light of these findings, the question arises as to the cause of such variability and the plasticity of multisensory interactions. In other words, it remains unknown whether multisensory interactions are governed by strict rules (although the abovementioned literature would suggest otherwise), whether the organizational principles of these rules can be altered by experience (i.e. development and/or training protocols) and the role and causal dependency upon anatomic hard-wiring (e.g. more fibers between specific sensory representations or faster transmission between certain representations) within the brain.

4.1 Open Questions

Generally, while several studies have investigated how these fundamental principles of multisensory integration impact behavior in isolation, the interaction between principles has not been addressed directly. Here we propose the investigation of looming/receding (i.e. approaching/retracting) stimuli to directly address how the three principles interact. The change in size of an approaching/retracting object corresponds to a change in the stimulus effectiveness across time. Further, by coupling either static or receding auditory stimuli to looming visual objects, spatial congruity can be directly investigated. Similarly, one can imagine varying onset latencies between the auditory and the visual stimuli in order to directly investigate the impact of temporal synchrony.

In a previous study, Cappe and colleagues (2009) have shown that processing of looming signals can be enhanced (i.e. faster detection responses) under auditory-visual conditions (Cappe, Thut, Romei, & Murray, 2009). In this study, subjects were asked to perform a simple go/no-go motion detection task, irrespective of the direction of motion, or whether auditory-visual stimuli were congruent (same vs. opposite direction). The results showed that irrespective of direction and congruency subjects responses were significantly speeded under multisensory vs. unisensory conditions. More interestingly, the results showed selective multisensory facilitation to occur upon congruent looming trials. Nonetheless, the neuronal structures and mechanisms involved in such selective multisensory integration have not been addressed, and remain unknown.

(To learn about the underlying neuronal generators of auditory-visual looming signals see Cappe, Thelen, et al., 2012, *JNeurosci*)

5 Multisensory object perception

Objects in our environment are usually perceived through multiple sensory inputs. The question is thus, how the different and/or same sensory inputs interact and how features are bound into a unified percept. Research has provided evidence that the neuronal substrates involved in multisensory object perception are modulated by the semantic relationship between auditory-visual features and task demands (see Figure 8, taken from Doehrmann & Naumer, 2008).

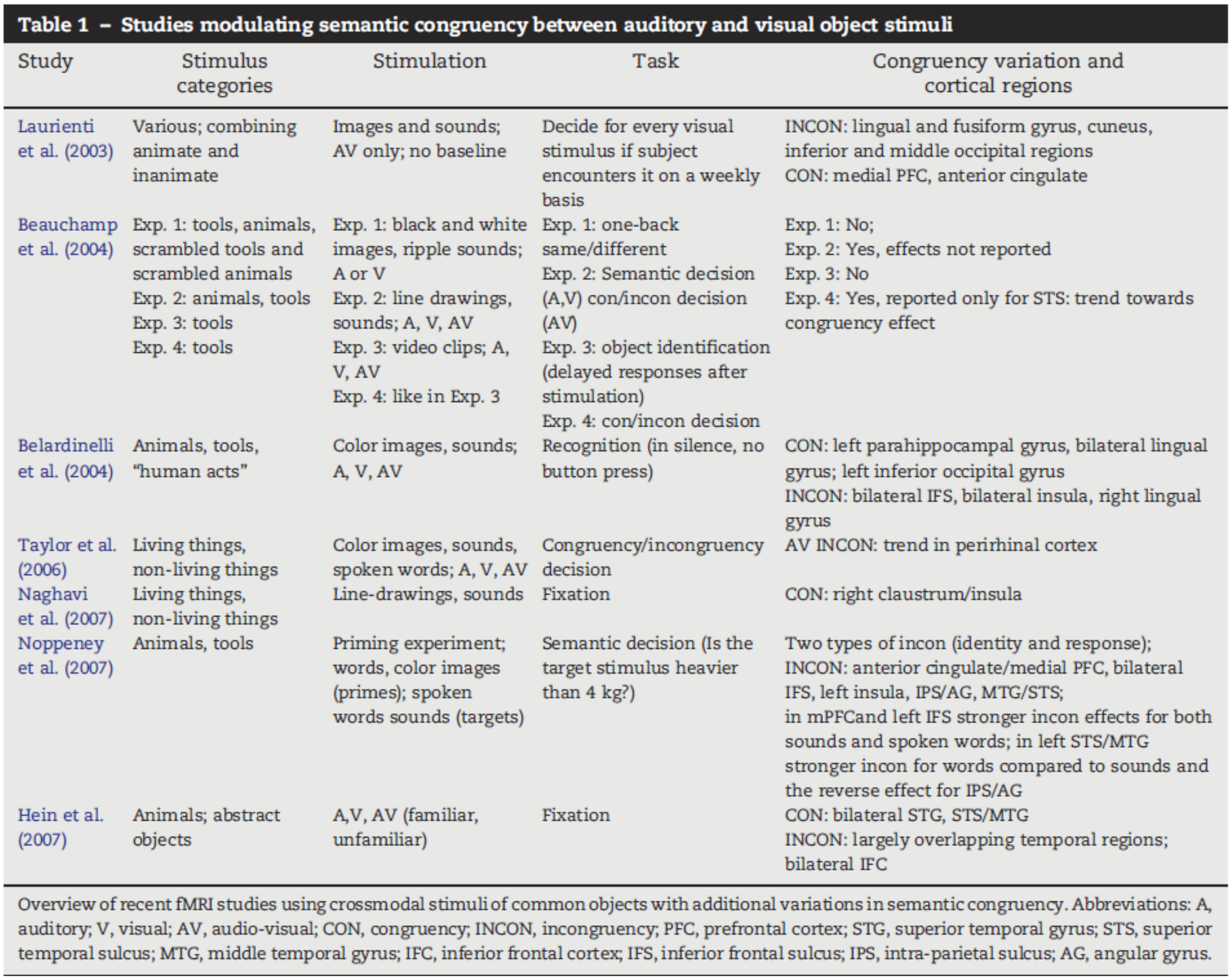


Table 1 – Studies modulating semantic congruency between auditory and visual object stimuli

Study	Stimulus categories	Stimulation	Task	Congruency variation and cortical regions
Laurienti et al. (2003)	Various; combining animate and inanimate	Images and sounds; AV only; no baseline	Decide for every visual stimulus if subject encounters it on a weekly basis	INCON: lingual and fusiform gyrus, cuneus, inferior and middle occipital regions CON: medial PFC, anterior cingulate
Beauchamp et al. (2004)	Exp. 1: tools, animals, scrambled tools and scrambled animals Exp. 2: animals, tools Exp. 3: tools Exp. 4: tools	Exp. 1: black and white images, ripple sounds; A or V Exp. 2: line drawings, sounds; A, V, AV Exp. 3: video clips; A, V, AV Exp. 4: like in Exp. 3	Exp. 1: one-back same/different Exp. 2: Semantic decision (A,V) con/incon decision (AV) Exp. 3: object identification (delayed responses after stimulation) Exp. 4: con/incon decision	Exp. 1: No; Exp. 2: Yes, effects not reported Exp. 3: No Exp. 4: Yes, reported only for STS: trend towards congruency effect
Belardinelli et al. (2004)	Animals, tools, "human acts"	Color images, sounds; A, V, AV	Recognition (in silence, no button press)	CON: left parahippocampal gyrus, bilateral lingual gyrus; left inferior occipital gyrus INCON: bilateral IFS, bilateral insula, right lingual gyrus
Taylor et al. (2006)	Living things, non-living things	Color images, sounds, spoken words; A, V, AV	Congruency/incongruency decision	AV INCON: trend in perirhinal cortex
Naghavi et al. (2007)	Living things, non-living things	Line-drawings, sounds	Fixation	CON: right claustrum/insula
Noppeney et al. (2007)	Animals, tools	Priming experiment; words, color images (primes); spoken words sounds (targets)	Semantic decision (Is the target stimulus heavier than 4 kg?)	Two types of incon (identity and response); INCON: anterior cingulate/medial PFC, bilateral IFS, left insula, IPS/AG, MTG/STS; in mPFCand left IFS stronger incon effects for both sounds and spoken words; in left STS/MTG stronger incon for words compared to sounds and the reverse effect for IPS/AG
Hein et al. (2007)	Animals; abstract objects	A,V, AV (familiar, unfamiliar)	Fixation	CON: bilateral STG, STS/MTG INCON: largely overlapping temporal regions; bilateral IFC

Overview of recent fMRI studies using crossmodal stimuli of common objects with additional variations in semantic congruency. Abbreviations: A, auditory; V, visual; AV, audio-visual; CON, congruency; INCON, incongruency; PFC, prefrontal cortex; STG, superior temporal gyrus; STS, superior temporal sulcus; MTG, middle temporal gyrus; IFC, inferior frontal cortex; IFS, inferior frontal sulcus; IPS, intra-parietal sulcus; AG, angular gyrus.

Interestingly, only few neuroimaging studies have investigated how and where multisensory information about objects is bound in the cortex (for examples see Amedi et al., 2005; Beauchamp et al., 2004; James et al., 2002; Naghavi, Eriksson, Larsson, & Nyberg, 2011). These studies have repeatedly reported the involvement of a common network implicated in multisensory object perception composed of the lateral occipital cortex (LOC), superior and middle temporal cortices as well as inferior frontal regions. Generally, differential roles in multisensory object recognition have been attributed to these areas. The LOC seems to be implicated in object recognition, while superior temporal regions have been related to object feature association. Inferior frontal cortices on the other hand seem to be involved in more cognitive processes such as object categorization and detection of semantic inconsistencies.

Beauchamp and colleagues (2004) found enhanced BOLD responses within ventral temporal regions when subjects were processing auditory-visual vs. visual-only object information. In another setup, Naghavi and colleagues (2011) found greater LOC activations to occur upon congruent auditory-visual object presentations as compared to incongruent presentations. Similar to auditory-visual interactions, Amedi and colleagues (2001) found responses to tactile object recognition to occur within the LOC during object recognition (Amedi, Malach, Hendler, Peled, & Zohary, 2001).

Beauchamp and colleagues (2004) found that the posterior superior temporal sulcus (pSTS) responds to both auditory and visual objects. Interestingly, this response was further enhanced when subjects were exposed to the combined auditory-visual percept. Naghavi and colleagues (2011) found that the middle temporal gyrus and the superior temporal gyrus (MTG/STG) showed stronger responses to incongruent auditory-visual pairs as compared to congruent presentations. In another study, Tanabe and colleagues (2005) investigated visual-tactile association learning. Their results showed a negative correlation between activity within the STS and the accuracy enhancement over several learning blocks (Tanabe et al., 2005).

In Beauchamp and colleagues' study (2004) dorso-lateral prefrontal cortices (dlPFC) also showed greater responses to multisensory vs. unisensory presentations. Moreover activation within this frontal region corresponded to the task-related responses rather than to perceptual processing of the stimuli. Naghavi and colleagues (2011) found activations with inferior frontal cortex (IFC) to occur upon presentation of congruent and incongruent auditory-visual pairs when subjects were passively exposed to the stimuli. They hypothesized that the IFC is involved in the multisensory association required for object related memories. Further, Taylor and colleagues (2006) found auditory-visual responses with medial frontal regions and the anterior cingulate cortex (Taylor, Moss, Stamatakis, & Tyler, 2006).

Recent studies have further emphasized the role of anterior temporal cortex (ATC) in auditory-visual object processing (Naci, Taylor, Cusack, & Tyler, 2012; Taylor et al., 2006; Taylor, Stamatakis, & Tyler, 2009). These studies have reported the involvement of ATC when subjects were asked to report whether auditory-visual object features where congruent vs. incongruent. Taylor and colleagues (2006) found greater activations within perirhinal cortex (PRC) upon incongruent multisensory presentations, suggesting that this area is sensitive to the semantic relationship between object features. In a subsequent study, Taylor and colleagues (2009) found that PRC but not the pSTG/MTG was not necessary for auditory-visual object integration. In their study, healthy subjects showed multisensory integration related activity within pSTG/MTG as well as within ATC (and more specifically the PRC). On the other hand, only patients with lesions within the AT, but not patients with lesions within the pSTG/MTSG, showed performance impairment, suggesting that the PRC is involved in multisensory integration of meaningful objects. Recently, Naci and colleagues (2012) have investigated auditory-visual congruency discrimination in healthy subjects. They found that activity within AT regions was modulated by the semantic relationship between object features as early as ~50-100ms post-stimulus onset. These results suggest that the AT region is implicated in familiarity-based object recognition, gating information from higher-order regions (object representations in memory) toward sensory areas.

Evidence for the involvement of the PRC in object recognition and identification has also been found in animals (Minamimoto, Saunders, & Richmond, 2010; Murray & Richmond, 2001; Richmond & Sato, 1987). Richmond and Sato (1987) have shown neurons within inferior temporal cortex (ITC) to respond when monkeys were a visual discrimination task. More recently, Murray and Richmond (2001) have related PRC activity to object recognition and identification. Further, they report greater activity within the PRC to occur upon presentation of novel vs. familiar objects, in line with the findings of Naci and colleagues (2012).

6 Multisensory Perceptual Memories

While several studies have investigated how multisensory interactions impact object perception when multiple sensory features are presented simultaneously, the following section will discuss the impact of prior multisensory encounters upon unisensory processing.

6.1 Behavioral advantages of multisensory memory traces

The interest of studying perceptual memory has risen from clinical evidence. Patients with memory impairments such as amnesia show strong impairment of episodic memory. They cannot explicitly retrieve autobiographic experiences. On the other hand perceptual memory seems to remain intact in such patients (Hamann & Squire, 1997). Perceptual memory traces are linked to (explicit) episodic memory trace retrieval, in that they provide the contextual information (where, what), but they appear to be retrieved in an implicit manner and independently of an explicit, conscious recognition. In a subsequent study Stark and Squire (2000) found evidence for the dissociation between repetition priming and recognition and recall (Stark & Squire, 2000). The patient E. P., with severe anterograde and retrograde amnesia showed impaired performances in a delayed recognition task, but performed normally on a stem completion task and perceptual identification priming, other than intact category learning. The authors argue that these findings suggest that intact memory traces formed during repetition priming are not accessible for familiarity-based recognition. These findings were also supported by studies showing that patients with focal lesions in the occipital cortices did not benefit from visual perceptual priming tasks, while they performed normally recognition memory tasks (Keane, Gabrieli, Mapstone, Johnson, & Corkin, 1995). These findings support the model of recognition memory comprising two independent processes. The first process is linked to recognition and recall of the previous experience. The second is involved in familiarity detection. While the first process depends on explicit episodic memory the second appears to be accessed implicitly.

Studies on perceptual learning have shown that the presentation of multisensory stimuli can enhance the training outcome compared to unisensory learning conditions. Van der Linden and colleagues (2010), studied the effect of multisensory category learning and its' effect on the formation of multisensory object representations. They used an auditory-visual paradigm, where subjects performed a one-back task. During the training session, this consisted in the presentation of images of birds that were coupled to auditory sounds. Subjects were presented with a series of multisensory trials, where they had to indicate whether two consecutive birds were the same type or not. During the post-training scanner session unisensory (auditory-only or visual-only) or multisensory stimuli were presented. The multisensory stimuli could be congruent or incongruent. Further, during the scanner session novel stimuli were intermixed to the previously studied bird types. The results indicated that stimuli that had been previously presented were discriminated faster and more accurately. The fMRI data showed that the superior temporal sulcus (STS) becomes involved in multisensory object representation. Interestingly this training effect did not generalize to the incongruent multisensory pairings. In fact the STS did not show differential responses to incongruent associations compared with novel stimuli (van der Linden et al., 2010).

Additionally, Shams and colleagues (2006) showed beneficial impact of auditory-visual training over unisensory perceptual learning in a motion detection task. Subjects which underwent multisensory training conditions outperformed subjects exposed to a visual-only training condition (Seitz, Kim, & Shams, 2006). Complementary results were found in a study focusing on auditory recognition (von Kriegstein & Giraud, 2006). The authors found that voice recognition improved, when subjects were presented with auditory-visual stimuli during the training session. These studies highlight the importance of auditory-visual congruency (Seitz et al., 2006; van der Linden, et al., 2010) and ethological validity (von Kriegstein & Giraud, 2006) in order to impact subsequent unisensory discrimination.

More recently, Chen and Spence (2010) investigated the existence of a common multisensory auditory-visual semantic system (Chen & Spence, 2010). The authors found that a semantically congruent sound can influence the discrimination accuracy of a masked visual image. More specifically the study found that identification performance was enhanced for those masked images that were presented simultaneously with a congruent sound, compared to those that had been presented with a burst of white noise or no sound at all. An opposing effect of discrimination impairment was found when the images were paired to an incongruent sound. In addition to the congruency effect the authors showed that the facilitations occurred when the auditory stimulus was presented either simultaneously or was delayed a little over than 300ms. When the sound was presented 533ms after the visual masked stimulus, no congruency effect was observed. These findings lead to the conclusion that the semantic systems of visual and auditory systems are not completely independent. Further, these authors refer to the theory of conceptual short-term-memory (CSTM) (Potter, 1993). According to this model the meaning of visual information is accessed rapidly (within 100ms), and is retained in CSTM for 300ms. If another redundant sensory cue is presented within this retention period, the semantic representation is consolidated and encoded into memory. This could explain the findings of Chen and Spence (2010). In fact the authors argue that the observed enhancement of object discrimination when presenting a congruent sound within a time window of 0 – 300ms after the presentation of the visual stimulus can be explained by a facilitated access to the meaning of the visual stimulus. The presentation of white noise or an incongruent auditory stimulus would interfere with the access to said meaning.

6.2 Neuronal networks implicated in multisensory memories

Very few studies have investigated the neuronal underpinnings of how multisensory encounters at one point in time impact later unisensory object processing. These studies have provided evidence that brain areas involved in the encoding of multisensory events are also involved during subsequent unisensory retrieval (James et al., 2002; Nyberg et al., 2000; von Kriegstein & Giraud, 2006; Wheeler et al., 2000).

James and colleagues (2002) asked subjects to explore meaningless clay objects either visually or haptically during the study phase. During a separate test phase within the scanner, subjects viewed images of objects that had been previously studied either visually or haptically. The results showed that images of objects that had been presented during the study phase either visually or haptically produced more activation within the middle occipital cortex (MOC) and the LOC compared to images of control objects ("new" objects that had not been encountered during the exploration phase). Further, haptic exploration of novel clay objects within the scanner produced activations within occipital cortices.

Nyberg and colleagues (2000) investigated the impact of multisensory memories upon word recognition. Subjects were asked to encode words either in a unisensory or a multisensory (visual-auditory) manner during the encoding phase. After the encoding phase, subjects immediately performed a visual recognition task. The results showed activation within auditory cortices upon unisensory retrieval. Similarly, Wheeler and colleagues (2000) asked subjects to study images and sounds over a period of 2 days. On the third day subjects performed a memory task, where upon presentation of the written label of a studied object they had to recall the previous encounter context (i.e. had it been studied as an image or as a sound). The imaging results showed activation within auditory cortices upon retrieval of objects that had been learned in an auditory context.

Von Kriegstein and Giraud (2006), on the other hand investigated whether similar effects also occurred for auditory object recognition. Subjects were asked to learn voice-face association during the study phase. Subsequently, subjects were asked to perform a speaker recognition task within the scanner. The results showed that auditory stimuli that had been paired with faces during the initial study phase produced activations within the fusiform face area (FFA).

Taken together these studies have provided evidence supporting the psychological postulate of "redintegration" (Hamilton, 1859). According to this postulate, the encounter with a single component of the original experience is sufficient to reactivate the whole experience previously encoded. In other words, the sole visual part of a previously auditory-visual experience, could elicit activity within auditory cortices during active retrieval of the original encounter context.

Similar evidence has also come from studies in monkeys (Colombo & Gross, 1994; Gibson & Maunsell, 1997; Haenny, Maunsell, & Schiller, 1988; Maunsell, Sclar, Nealey, & DePriest, 1991). Haenny and colleagues (1988) performed intracranial microelectrode recordings in the visual cortex of rhesus monkeys. The monkeys performed a delayed orientation match-to-sample task, where they had to release a lever upon trials where the sample matched the orientation of an initial tactile or visual cue stimulus. The single neuron recordings in V4 showed activation to tactile cues during the delay period similar to those observed after a visual cue had been delivered. In a subsequent study, Maunsell and colleagues (1991) further addressed the question of the specificity of the information relayed to visual cortex neurons through the tactile cue stimulus. Generally, these two studies have shown that haptic information activates neurons within visual cortex. Moreover, these authors found that haptic information activated visual neurons which coded for the same orientation as during visual cue presentation (orientation selectivity).

Colombo and Gross (1994) found similar findings within the inferior temporal regions during and auditory-visual delay match-to-sample task. Monkeys were presented with an auditory cue and had to decide whether a delayed visual sample was a match (previously learned association). Microelectrode recordings were performed within inferior temporal (IT) regions and the hippocampus. Previous studies had shown that IT neurons are involved in visual pattern information processing (Colombo & Gross, 1994; Gross, Rodman, Gochin, & Colombo, 1993). During the delay period neurons within the IT responded to visual cues in the visual-visual delayed match-to-sample condition. Interestingly, the results showed that these IT neurons also responded to auditory cues. Subsequently, Gibson and Maunsell (1997) extended these findings showing that delay-sensitive neurons with the IT responded to either a visual cue matched to an auditory sample as well as to an auditory cue matched with a visual sample. In the case of an auditory cue matched to an auditory sample no evidence of selective delay period activity was recorded, suggesting that the results were not due to general delay period selective responses.

6.3 Open Questions

In all the aforementioned studies, extensive studying of the multisensory association was performed before testing. Consequently, the paradigms reported above entailed the clear separation between encoding and retrieval related processes. Further, the results have provided evidence that well-established multisensory memories can impact subsequent unisensory processing. More specifically, these studies have implicated large-scale memory processes, where object representations are part of consolidated memories. Thus, whether multisensory memories impact subsequent unisensory processing after a single encounter with the combined percept remains unresolved.

Further, in the study of Wheeler and colleagues (2000) subjects were explicitly asked to recall the initial encounter context in which they had studied objects. This task might have led to (re)activation within auditory cortex due to mental imagery processes. Similar, in the delayed-match to sample tasks monkeys had to actively hold in memory the cue stimulus as well as the expected sample in order to perform the task. Thus, whether activity within the non-stimulated sensory area was due to the multisensory nature of the memory trace or mental representation of the cue/stimulus is unclear.

Another common point to the abovementioned studies is that subjects had to explicitly engage in processing both sensory cues. James and colleagues (2002) asked subjects to haptically explore clay objects for ~30seconds, which could have lead subjects to construct a (mental) visual representation. Similarly, in the delayed match-to-sample tasks, monkeys had to actively combine the sensory features that compose the association. Whether multisensory memory traces impact subsequent unisensory processing even when information in the additional sensory modality is task-irrelevant has not been directly addressed.

Lastly, Von Kriegstein and Giraud (2006) contrasted auditory speaker recognition with ring tone recognition. Subjects studied either voice-face or voice-name vs. ring tone-phone (image) or ring tone-brand name (word) associations. The results showed that subjects benefitted from multisensory memories only for voice recognition. Moreover this behavioral impact was not observed for voice-name associations. The authors postulated that multisensory memories only impact unisensory recognition when the association entails ethologically relevant information. Whether these findings are linked to the specific material (voices and faces) presented to subjects in this study, or can occur also for other auditory-visual combinations remains unclear.

6.4 Single-trial perceptual memories

To address these open questions we have adopted a continuous recognition task. Subjects were asked to indicate on each trial whether an object (visual or auditory) was presented for the first or the second time during the current block of trials. This task has also been used for studies on the interplay between short and long-term memory and memory encoding/retrieval processes in clinical populations (James, Morand, Barcellona-Lehmann, Michel, & Schnider, 2009; Lehmann, Morand, James, & Schnider, 2007).

6.4.1.1 Behavioral impact of single-trial multisensory memories

In previous studies Murray and colleagues coupled meaningful pictures with corresponding sounds (i.e. congruent) upon initial encounter. Subjects were asked to attend only to images and ignore the auditory stimuli. Thus, Murray and colleagues were able to address the question of whether a single encounter with a task-irrelevant multisensory coupling was sufficient to impact subsequent unisensory (visual-only) processing. Compared to images that had been presented only in a unisensory manner upon initial presentations, images previously coupled with a congruent sound were recognized significantly better upon repeated, unisensory presentation (Lehmann & Murray, 2005; Murray et al., 2004; Murray, Foxe, & Wylie, 2005). In a subsequent study, Lehmann and Murray (2004) investigated the impact of initial incongruent and meaningless auditory-visual couplings upon subsequent visual object recognition. In the case of initially incongruent couplings, visual recognition upon repeated trials was not significantly different from recognition accuracy of images encountered only visually. Interestingly, images that had been coupled to meaningless 1000Hz tone upon initial encounter were recognized significantly less accurately than images encountered only visually. These findings challenge the hypothesis of von Kriegstein and Giraud (2006), showing that a single encounter with a meaningless auditory-visual association (i.e. ethologically irrelevant) is sufficient to impact subsequent visual object recognition.

In addition to investigating auditory-visual couplings, Lehmann and Murray (2005) also investigated whether tactile-visual encounters could impact subsequent visual object recognition. The results showed that recognition accuracy between images encountered only visually and images previously paired to a haptic stimulation did not differ. Somatosensory stimuli are less informative (i.e. cat and dog both have fur) in terms of object discrimination and stimulus durations were not representative of natural haptic exploration. Further, the same 50Hz stimulation was coupled to images across trials. This suggests that subjects rapidly learned that the haptic stimulation did not convey object-related information. Whether longer and more variable haptic stimulations upon initial visual object presentations interfere with later unisensory object recognition will have to be further investigated.

Taken together, these results suggest that object based multisensory integration is sensitive to determination of object identity and the object features in other sensory modalities. Several studies have shown that vision is more reliable in object identification than audition (and somatosensation), leading to more accurate and faster object recognition (Welch & Warren, 1980; Yuval-Greenberg & Deouell, 2007, 2009). Lehmann and Murray (2005) proposed that upon presentation of a redundant auditory sound, object representations can be accessed through additional sensory channels, effectively enhancing the fidelity of the perceptual memory trace. This is not the case when the auditory stimulus does not convey redundant information (i.e. incongruent), the visual information is less ambiguous, the brain treats the auditory object independently. The authors based this hypothesis on evidence suggesting rapid parallel processing of objects (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Rousselet, Fabre-Thorpe, & Thorpe, 2002). In the case of meaningless auditory-visual encounters, the decrease in accurate object recognition could have been caused by the experimental set-up. Across experimental blocks, the same 1000Hz tone was coupled to images upon initial encounters, while unique meaningful sounds were presented in the congruent and incongruent encounter context. This continuous association/dissociation of the tone with a rapid succession of images could have led to enhance the signal-to-noise ratio (SNR) upon auditory-visual trials, interfering with memory trace formation. Thus, the decrease in recognition accuracy upon repeated unisensory trials could be explained by more ambiguous memory trace formation, rather than to the multisensory encounter context *per* se. To test this hypothesis we coupled unique sounds to visual objects upon the initial encounter (Thelen, Cappe, & Murray, 2012).

6.4.2 Neural correlates of single-trial multisensory memories

In addition to behavioral measures, Murray and colleagues (2004, 2005) also investigated the neuronal generators involved in the differential processing of visual images due to prior encounter context. In both imaging studies (electrical neuroimaging in 2004, and hemodynamic imaging in 2005) multisensory presentations entailed congruent auditory-visual couplings upon initial encounter context.

The neuroimaging results showed that upon repeated unisensory encounters, brain responses to images previously encountered in a multisensory context differed as soon as ~60-136ms post-stimulus onset. Moreover, the source of this differential processing was found within the LOC and the results of the source estimation were confirmed through hemodynamic imaging.

The timing and location of differential visual processing reported in these studies showed that multisensory memories incidentally impact subsequent visual processing at early stages and within "sensory-specific" areas. Further, the hemodynamic imaging results did not reveal multisensory-related activity within auditory areas. Thus, in contrast to what has been reported previously, that brain areas involved in encoding a multisensory event are reactivated upon subsequent unisensory retrieval.

The role of the LOC in multisensory object recognition has been reported in studied in animals (Colombo & Gross, 1994; Gibson & Maunsell, 1997; Haenny et al., 1988; Maunsell et al., 1991). As mentioned before, microelectrode recordings have shown that neuronal activity within lateral visual areas is differentially modulated in multisensory vs. unisensory delay match-to-sample tasks. Further, Doninger and colleagues (2001) found that visual perceptual learning modulated ERP responses over occipito-temporal regions during perceptual closure

(Doniger et al., 2001). Perceptual learning lead to greater ERP amplitudes, indicating an increase in activity within the underlying generators coupled to enhanced object recognition. Thus, the findings of Murray and colleagues (2004, 2005) reveal that the LOC shows greater responses for accurate vs. incorrect/unachieved visual object recognition. Enhanced visual object recognition for stimuli previously encountered in a congruent auditory-visual context, suggests that multisensory perceptual memory traces impact visual object processing within early stages and within sensory-specific cortices.

(To learn more about how meaningless auditory-visual couplings impact subsequent visual processing see: Thelen et al. 2012, *NeuroImage*; and what about auditory object processing? For the answer see: Thelen et al. *JCognition*, *in review*)

7 Electrical Neuroimaging

In order to investigate the neuronal underpinnings of multisensory integration, in this thesis, behavioral investigations were coupled with electrical neuroimaging methods. This method allows the investigation of sensory and cognitive processing in the temporal domain at the sub-second level.

The electroencephalogram (EEG) measures electrical potentials at the level of the scalp. The measure obtained at each electrode position is calculated as the difference of voltage between the electrode and a reference electrode. The value of the reference electrode is zero microvolt by definition. Therefore, changing the spatial location of the reference electrode will automatically change the voltage measured at the other electrodes. Thus, waveforms (i.e. electric potential as a function of time) are highly dependent on the placement of the reference electrode. Shifting the spatial location of the reference will impact the position and amplitude of the peak of the waveform in time. When averaging waveforms over trials (i.e. calculating event related potentials; ERPs) this reference dependency can alter the variance measured at a given electrode and consequently cause a change of the statistical results, which can lead to misinterpretation of the data. In contrast, a change in reference position will not affect the electrical field map (i.e. topography map), since a map is defined as a relative potential difference across the electrode montage. Thus, changing the reference will solely shift the map (horizontally) with respect to the zero plane, but will leave its' configuration untouched.

The ERP analysis method proposed here allows for the statistical assessment of both quantitative (in terms of response strength modulations measured by the global field power, GFP) and qualitative (in terms of electric field modulation or topography) differences of brain responses between conditions. GFP is calculated as the standard deviation of all electrodes at a given point in time (Brunet, Murray, & Michel, 2011; Murray, Brunet, & Michel, 2008; Lehmann & Skrandies, 1980). In other words, the GFP represents a measure of the strength of the electric field at the scalp. Moreover, this measure is independent of topographical modulations. On the other hand, the topographic analysis provides information about the spatial configuration of the electric field independently from its' strength. A topographic difference forcibly indicates a change the configuration of the neuronal generators involved between conditions. The inverse is not true. Statistically, different neuronal generator configurations could give rise to identical electric field configurations at the level of the scalp (Lehmann, 1987).

Due to the orthogonal nature of these two measures, a difference in GFP without a topographic difference between conditions is interpreted an increase/decrease of activity within statistically indistinguishable intracranial generators. A difference in topography in the absence of GFP modulations is interpreted as a differential neuronal generators being active between conditions. Nonetheless, a difference in topography at a given point in time can also results from a onset latency shift between conditions. Thus, when analyzing ERPs over a certain time window vs. at a single point in time, a segmentation of the topographical maps into stable functional microstates (i.e. template maps) across the time window of interest can be informative. This is achieved, through hierarchical clustering. The group-averaged ERPs for each experimental condition is submitted to the algorithm, which generates a set of stable maps which best represent the group-averaged ERPs for each experimental condition. The hierarchical clustering algorithm does not take into account onset latency shifts and period of presence over time of each map between conditions. The resulting microstate maps are then statistically tested by a "back-fitting" procedure to the original data set. This procedure calculates the spatial correlation between the template maps generated by the algorithm at the group-averaged ERP level and the single-subject ERPs (Brunet et al., 2011; Murray et al., 2008).

In a last step, source estimations can be performed. Localization algorithms allow the identification and quantification of intracranial sources (Michel et al., 2004). This is done by resolving the so-called "inverse problem", i.e. the estimation of the sources in a statistical model which takes into account bio-physical constraints of current propagation within the brain and the surrounding tissues. At the group-level statistical analyses can then be performed within the "inverse space" in order to identify a change in activity (i.e. strength) or neuronal generators implicated between conditions.

8 Results

8.1 Evidence for synergy between principles of multisensory integration

Looming signals reveal synergistic principles of multisensory integration

Céline Cappe*, Antonia Thelen*, Vincenzo Romei, Gregor Thut, and Micah M. Murray

Journal of Neuroscience. 2012 Jan;32(4):1171-82

*equal contributions

CONTRIBUTION: analyzed and interpreted the data; wrote the manuscript

Abstract

Multisensory interactions are a fundamental feature of brain organization. Principles governing multisensory processing have been established by varying stimulus location, timing and efficacy independently. Determining whether and how such principles operate when stimuli vary dynamically in their perceived distance (as when looming/receding) provides an assay for synergy among the above principles and also means for linking multisensory interactions between rudimentary stimuli with higher-order signals used for communication and motor planning. Human participants indicated movement of looming or receding versus static stimuli that were visual, auditory, or multisensory combinations while 160-channel EEG was recorded. Multivariate EEG analyses and distributed source estimations were performed. Nonlinear interactions between looming signals were observed at early post-stimulus latencies (~75 ms) in analyses of voltage waveforms, global field power, and source estimations. These looming-specific interactions positively correlated with reaction time facilitation, providing direct links between neural and performance metrics of multisensory integration. Statistical analyses of source estimations identified looming-specific interactions within the right claustrum/insula extending inferiorly into the amygdala and also within the bilateral cuneus extending into the inferior and lateral occipital cortices. Multisensory effects common to all conditions, regardless of perceived distance and congruity, followed (~115 ms) and manifested as faster transition between temporally stable brain networks (vs summed responses to unisensory conditions). We demonstrate the early-latency, synergistic interplay between existing principles of multisensory interactions. Such findings change the manner in which to model multisensory interactions at neural and behavioral/perceptual levels. We also provide neurophysiologic backing for the notion that looming signals receive preferential treatment during perception.

8.2 Impact of single-trial multisensory memories

Electrical neuroimaging of memory discrimination based on single-trial multisensory learning

Antonia Thelen, Céline Cappe, and Micah M. Murray

NeuroImage. 2012 Sep;62(3):1478-88. Epub 2012 May 18.

CONTRIBUTION: performed research; analyzed data; wrote the manuscript

Abstract

Multisensory experiences influence subsequent memory performance and brain responses. Studies have thus far concentrated on semantically congruent pairings, leaving unresolved the influence of stimulus pairing and memory sub-types. Here, we paired images with unique, meaningless sounds during a continuous recognition task to determine if purely episodic, single-trial multisensory experiences can incidentally impact subsequent visual object discrimination. Psychophysics and electrical neuroimaging analyses of visual evoked potentials (VEPs) compared responses to repeated images either paired or not with a meaningless sound during initial encounters. Recognition accuracy was significantly impaired for images initially presented as multisensory pairs and could not be explained in terms of differential attention or transfer of effects from encoding to retrieval. VEP modulations occurred at 100-130ms and 270-310ms and stemmed from topographic differences indicative of network configuration changes within the brain. Distributed source estimations localized the earlier effect to regions of the right posterior temporal gyrus (STG) and the later effect to regions of the middle temporal gyrus (MTG). Responses in these regions were stronger for images previously encountered as multisensory pairs. Only the later effect correlated with performance such that greater MTG activity in response to repeated visual stimuli was linked with greater performance decrements. The present findings suggest that brain networks involved in this discrimination may critically depend on whether multisensory events facilitate or impair later visual memory performance. More generally, the data support models whereby effects of multisensory interactions persist to incidentally affect subsequent behavior as well as visual processing during its initial stages.

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

Antonia Thelen, Durk Talsma, and Micah M. Murray

Journal Cognition, in review

CONTRIBUTION: designed research; performed research; analyzed data; wrote the manuscript.

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. The implication is that multisensory exposures establish distinct object representations that are accessible during later unisensory processing. The generalization of such effects to auditory object recognition has yet to be established and was the focus of the present study. First, we demonstrate that visual object recognition 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 recognition. Auditory object recognition 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 the unisensory object recognition 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 independent object processing mechanisms between modalities. We discuss these findings in terms of the conceptual short term memory and the modality appropriateness models. Our results suggest differential recruitment and modulation of conceptual memory networks according to the sensory task at hand.

9 Discussion

In this chapter I will give a brief summary of our relevant results and discuss the findings obtained in our three studies (Cappe et al., 2012; Thelen, Cappe, & Murray, 2012; Thelen, Talsma, & Murray, *in review*).

9.1 Selective perceptual enhancement for multisensory looming stimuli

9.1.1 Summary of Results

The behavioral data collected concurrently to the EEG acquisition have been published separately (Cappe et al., 2009). Generally, subjects' response speed showed evidence for redundant signal effects (RSE). In other words, subjects' reaction times (RTs) in the motion detection task were faster under multisensory vs. unisensory conditions. This response facilitation was observed irrespective of the motion direction of stimuli (looming/receding) and congruity of the stimulus couplings (congruent/incongruent). Further, when testing whether probability summation of the unisensory components could explain the observed RSE on RTs, the results showed that for all multisensory presentations (looming/receding, congruent/incongruent) RT distributions violated Millers' race model (Miller, 1982). This violation is interpreted as evidence for multisensory integration. In contrast, when the race model is satisfied, this is interpreted as two independent sensory processes "racing/competing" to elicit a motor response. In addition, Cappe and colleagues (2009) further investigated whether subjects showed preferential integration for looming vs. receding stimuli. The results showed that under auditory-visual congruent looming conditions RTs were significantly faster as compared to any other multisensory stimulus coupling.

Interestingly, the selective facilitation at the RT level was reflected in a separate task where subjects were asked to rate the perceived motion strength. Results showed that motion perception was generally stronger under multisensory vs. unisensory conditions. Further, motion strength perception of congruent auditory-visual looming stimuli was significantly higher compared to any other multisensory stimulus combination. Taken together, these results provided evidence for selective integration of multisensory looming stimuli at the behavioral level.

We then proceeded to investigate the neuronal processes involved in the differential perceptual processing of auditory-visual looming stimuli. Similar to analyses at the behavioral level, brain responses to combined auditory-visual presentations (paired) were compared to the algebraic sum of the unisensory components (sum). We found non-linear (pair ≠ sum) brain responses in terms of ERP waveforms upon congruent auditory-visual looming stimuli as soon as ~68ms post-stimulus onset. This was not the case for any other multisensory condition, which resulted in non-linear effects at later latencies (for example at ~119ms post-stimulus onset for the congruent multisensory receding condition).

As mentioned in the methods section, waveform analyses are biased by the choice of the recording reference (or in our case of the average reference), and thus global electric field analyses were conducted. The results from hierarchical topographic cluster analysis revealed a template map which had different onset latencies between the pair and the sum conditions in a time window from 73-145ms post-stimulus onset. This results suggests a latency shift of the onset of activity within underlying intracranial sources. In other words, upon paired multisensory stimulus presentations, the onset of the template map was observed significantly earlier (at ~113ms post-stimulus onset) as compared to the sum of the unisensory conditions. The temporal onset shift was used to further explore whether differences in electric field configurations were complemented by differences in global response strength differences between the pair and sum of multisensory conditions. The results showed that within the first 73-113ms post-stimulus onset responses to congruent paired multisensory looming presentations showed significant non-linear response strength enhancement. Again, this early difference was specific for paired looming stimuli as compared to any other multisensory condition. Moreover this selective GFP enhancement under multisensory looming conditions was highly correlated with the response enhancement observed at the behavioral level (faster RTs). This finding suggests that early and low-level multisensory interactions are relevant for behavior.

Source estimations of activity recorded during this time window revealed differential processing upon congruent paired looming presentations to occur with the right claustrum/insula, the right anterior inferior temporal lobe, right lingual gyrus and the right middle occipital gyrus, as well as bilaterally within the cuneus. Moreover, activity within these regions was subadditive (AV < A+V). Single-cell recordings have shown that multisensory interactions lead to super, respectively subadditive effects depending on whether unisensory stimuli elicit weak or strong responses. Our behavioral data suggest that looming cues are perceived as salient signals, and thus elicit strong responses under unisensory conditions. Consequently, the subadditive effects reported here are consistent with prior literature (Carriere, Royal, Wallace, & Carolina, 2008; Krueger, Royal, Fister, & Wallace, 2009).

9.1.2 Behavioral relevance of motion processing

Previous work in animals has shown preferential responses for looming stimuli at the behavioral and the neuronal level (de Vries & Clandinin, 2012; Maier et al., 2008; Maier, Neuhoff, Logothetis, & Ghazanfar, 2004; Maier & Ghazanfar, 2007; Sun & Frost, 1998). Generally, these studies have suggested that this preferential integration can be explained by the ethological relevance of these signals. Response facilitation to approaching stimuli is of crucial importance for behavior. Throughout evolution, this has lead to the emergence of specialized neuronal networks involved in the detection and the enhancement of the saliency through bottom-up processes of such cues, which in turn can explain the greater attentional and emotional capture observed in humans.

Selective processing of looming cues at the behavioral has been consistently reported in humans (Bach, Neuhoff, Perrig, & Seifritz, 2009; Cappe et al., 2009; Harrison, 2012). Harrison (2012) asked subjects to discriminate the direction of auditory motion stimuli (looming/receding) and ignore simultaneously presented visual stimuli. The results showed a general decrease in accurate motion discrimination upon incongruent multisensory trials (ALVR; ARVL, for acronyms see Cappe et al. 2009). When investigating dynamic capture effects (either AL-ALVR; AL-ALVL; or AR-ARVR; AR-ARVL), the results showed significantly greater effects for ARVL vs. ALVR trials. In other words, upon all incongruent multisensory trials auditory motion discrimination was "captured" by the simultaneously presented and task-irrelevant visual stimulus. Furthermore, this dynamic capture effect was significantly greater when auditory receding stimuli were presented with a visual looming stimulus.

In all the abovementioned studies, the preferential processing of looming stimuli, is linked to the ethological relevance of the stimulus. In fact, it is rather advantageous to pay more attention to an approaching object than to a retracting object. This is of crucial importance for survival, since an approaching threat has to be quickly identified, in order to elicit fight/flight behavior (or prepare for communication). On the other hand, a retracting object is of less crucial importance for survival (unless you are a hungry animal watching your prey escape). Bach and colleagues (2009) directly addressed this question. In their study subjects' perceived loudness ratings of approaching vs. retracting stimuli were significantly greater. Further, these authors measured skin conductance responses (SCR) upon trials and found an increase in the SCR amplitude upon auditory looming trials, suggesting an increase in alertness. In a second experiment, subjects were asked to give explicit emotional ratings of the motion cues. The results showed that subjects rated approaching sounds as more unpleasant, more arousing, more potent, more intense, more salient and more threatening that receding sounds. These results support the assumption that looming stimuli are generally perceived as warning signals.

9.1.3 Neuronal basis of perceptual enhancement of motion perception

Up until now, only few studies have investigated the neural structures and mechanisms involved in looming/receding motion processing in non-human primates and humans (Bach et al., 2008; Cappe et al., 2012; Maier et al., 2008; Maier & Ghazanfar, 2007; Seifritz et al., 2002; Tyll et al., 2013). Although the studies in pigeons and drosophilae should be mentioned (de Vries & Clandinin, 2012; Sun & Frost, 1998). Sun and Frost (1998) investigated neuronal responses to visual looming signals in pigeons. They found that presentation of approaching objects ("soccer ball"-like visual stimulus) elicited responses in neurons within the nucleus rotundus (homologue to the mammalian thalamic pulvinar nucleus). Moreover, these responses were coupled to increased electromyogram (EMG) responses recorded in the pigeons' major flight muscle as well as increased heart rate, associated with natural defensive responses (avoidance/flight). No such response was observed when self approaching motion was mimicked (the pigeon flying towards the soccer ball). Similarly, de Vries and Clandinin (2012) found a specific neuronal population within the optic lobe of drosophilae which responded to looming signals. Interestingly, after genetic manipulation of these neurons, escape behavior was extinguished. These findings suggest the existence of specialized neuronal pools and/or networks dealing with processing of looming cues in order to facilitate avoidance/flight behavior in these animals.

Evidence for selective processing of looming/approaching signals has also come from studies in monkeys. Maier and colleagues (2004) found "preferential looking" towards auditory-visual looming signals as compared to receding stimuli in rhesus monkeys. Maier and Ghazanfar (2007) investigated whether the preferential processing of looming sounds observed at the behavioral level was reflected in selective responses within the auditory cortex of the rhesus monkey. The results showed the existence of a bias for looming vs. receding sounds within the lateral belt of auditory cortex. This response enhancement was characterized by sustained activity (in terms of local field potentials; LFP) within the gamma-band (45-90Hz) upon presentation of looming stimuli that was not seen for receding stimuli. These findings were supported by multiunit activity (MUA) analyses, which revealed higher firing rates upon looming vs. receding stimulus presentations. In a subsequent study Maier and colleagues (2008) investigated the impact of auditory-visual looming/receding stimuli upon neuronal activity within the lateral belt of auditory cortex and the upper bank of the STS. A sustained increase in oscillatory activity within both areas in the gamma-band range (45-90Hz) was observed upon presentation of looming stimuli. Moreover, the authors observed an increase in gamma-band coherence and phase synchrony between these areas upon presentation of multisensory looming trials. This increase in oscillatory coherence and phase synchrony suggest selective integration of multisensory looming vs. receding cues.

Similar to what has been observed in monkeys, human functional imaging studies have shown greater responses to auditory looming stimuli within auditory cortical areas (Bach et al., 2008) and within auditory motion processing areas (Seifritz et al., 2002). In addition to activity enhancements within the intraparietal sulcus and the temporal plane, Bach and colleagues (2008) also found greater BOLD responses within the amygdala. The enhanced responses in the amygdala were linked to an increase in salience of the auditory looming stimuli indicating the presence of a behaviorally relevant stimulus. This is in line with their behavioral findings, suggesting that auditory looming stimuli act as warning cues (Bach et al., 2009).

Recently, Tyll and colleagues (2013) performed a hemodynamic imaging study on the neuronal basis of auditory-visual looming perception. The authors found enhanced BOLD responses to multisensory vs. unisensory looming stimuli within a network including low-level visual and auditory areas, in addition to activations within the STS. Further, the authors investigated functional connectivity between these regions. They found enhanced functional connectivity between the STS and low-level visual areas under multisensory conditions.

Enhanced activity within low-level visual areas upon presentation of auditory looming stimuli has also been reported by Romei and colleagues (2009). These authors applied single-pulse TMS over occipital areas while presenting auditory looming and receding sounds. Results showed that phosphene induction was significantly enhanced when upon trials where auditory looming vs. receding stimuli were presented (Romei et al., 2009).

The collective findings here suggest that the perceptual enhancement observed for auditory-visual looming cues is caused by greater activity and connectivity between low-level auditory and visual areas with the STS. Whether the enhanced connectivity in humans reported by Tyll and colleagues (2013) is mediated by oscillatory mechanisms such as enhanced coherence and phase synchrony, similar to what has been reported by Maier and colleagues (2008) in monkeys, will be the focus of future investigations. In addition to modulating activity within low-level cortices, the behavioral and neuroimaging studies in humans suggest that the looming/approaching signals elicit alerting responses (Bach et al., 2008, 2009; Cappe et al., 2012). These results seems particularly relevant in the light of recent findings showing enhanced motion processing within autistic subjects, which also show strong deficits in social interactions (Foss-Feig, Tadin, Schauder, & Cascio, 2013; Woynaroski et al., 2013).

The importance of the investigation of motion perception is also highlighted by recent studies in children/young adults and elderly populations (Poulter & Wann, 2013; Purcell, Wann, Wilmut, & Poulter, 2012; Wann, Poulter, & Purcell, 2011). These studies have found significant impairment of motion processing within these populations, suggesting that the neuronal mechanisms involved in processing looming stimuli mature relatively late during development. Thus, the investigation of looming motion can shed a light upon the development and maturation as well as vulnerability of multisensory processes across the lifespan. This could be a good model to investigate how principles of multisensory integration develop and their susceptibility to training-induced changes.

9.2 Efficacy of multisensory memories

9.2.1 Summary of results

Our collective results have repeatedly shown that a single encounter with a multisensory event is sufficient to impact subsequent unisensory (visual and auditory) processing (for a review see Thelen & Murray, 2013).

For visual object recognition, we found that recognition accuracy was enhanced for images that had been presented in a congruent multisensory context upon initial encounter, while initial meaningless associations lead to a decrease in recognition accuracy (Thelen et al., 2012; Thelen et al., *in review*). This was always compared to images that had been encountered in a unisensory context upon initial encounter. These results replicated and extended previous findings (Lehmann & Murray, 2005; Murray et al., 2004, 2005; for a review see Murray & Sperdin, 2010), emphasizing the robustness of these effects. Interestingly, while Lehmann and Murray (2005) failed to show the impact of initial incongruent encounter context upon subsequent visual object recognition, we found a decrease in recognition accuracy for images that had been paired with an incongruent sound upon initial encounter.

Further, we are the first to report that auditory object recognition is affected by single-trial multisensory memories (Thelen et al. *in review*). Similar to what we have observed for visual object recognition, auditory objects that had been paired with congruent images upon initial encounter are recognized significantly better than auditory objects encountered in a unisensory context. Further, initial incongruent encounter context lead to significant recognition impairment upon repeated presentations of auditory objects. In contrast to what has been observed for the visual modality, initial meaningless associations did not impact subsequent auditory object recognition.

The impact of single-trial multisensory memories upon object recognition accuracy was observed in the absence of modulation on RTs upon repeated presentations. Nonetheless, we observed significant response slowing upon initial multisensory vs. unisensory encounters. This could be explained by a distraction mechanism. Subjects are explicitly told to attend to a single sensory modality. Consequently, the presentation of an unexpected event within another sensory modality lead to a momentary distraction (Spence, Nicholls, & Driver, 2001). On the other hand, an alternative hypothesis would suggest that the RT slowing is linked to memory trace formation of the multisensory association.

In addition to behavioral data, we also recorded EEG data while subjects were performing a visual object recognition task (Thelen et al., 2012). Subjects were asked to discriminate between initial and repeated presentations of visual objects. Upon half of the initial encounters, images were paired with unique meaningless sounds. We compared brain responses upon unisensory repeated presentations, in order to investigate the neuronal substrates involved the differential object recognition accuracy found at the behavioral level.

The results showed that upon repeated presentations brain responses differed as soon as ~100ms post-stimulus onset according to initial encounter context. The differential brain responses were reflected in a change in the electric field configuration (i.e. different topographies). Source localization showed that upon presentation of images that had been paired with a meaningless sound upon initial encounter, a cluster within the right posterior insular cortex extending to the STG was active. Alternatively, a smaller cluster within the LOC showed stronger responses upon presentation of images that had been encountered in a unisensory context.

We also found differential brain responses during a later time window starting at ~270ms post-stimulus onset. Again, this differential response was characterized by topographical changes, suggesting the differential involvement of neuronal sources. Source estimation of differential activations revealed a cluster within the right STG extending anteriorly to the middle temporal gyrus (MTG). Most interestingly, the activity within this cluster was positively correlated with the recognition accuracy impairment observed at the behavioral level. No other correlations between brain activity and behavior were found.

9.2.2 Impact of single-trial memories upon unisensory object recognition

Previous studies, investigating the impact of multisensory memories upon subsequent unisensory processing asked subjects to study multisensory associations over extensive periods where stimuli were presented multiple times (see Kim, Seitz, & Shams, 2008; van der Linden et al., 2011; Wheeler et al., 2000 for examples). Kim and colleagues (2008) showed benefits of multisensory training in a visual motion direction discrimination task. Subjects were trained over several days and were trained on 6000 trials (5 days, 1200 trials each day). Similarly, van der Linden and colleagues (2011) trained subjects to associate bird types with the respective bird sounds over a three day period (1 $^{1}/_{2}$ hours per day). Wheeler and colleagues (2000) exposed subjects to picture-sound pairs over a period of 2 days (10 repetitions per stimulus, per day). Here we show that even single encounters with a multisensory pair are sufficient to impact subsequent unisensory processing. These findings have strong clinical implications for memory rehabilitation procedures. In fact, during rehabilitation, patients are subjected to large batteries of tests (due to high comorbidity of deficits), which is taxing and exhausting. The present findings suggest that adopting multisensory single-trial learning procedures could be beneficial. Lehmann and colleagues (2007) investigated memory encoding and retrieval within a patient with post-anoxic amnesia. In their study the same continuous recognition task was used, although all stimuli were presented only visually (Lehmann et al., 2007). Compared to the control group, the patient performed at similar levels upon initial trails (indicating that an image was "new"). Contrariwise, recognition accuracy of the patient was significantly impaired (47% vs. 94.2%) upon repeated presentations. Future investigations are necessary to test whether initial congruent multisensory presentations can help to improve such deficient performance within patients.

Further, some groups have suggested that multisensory memories only impact subsequent unisensory processing when the multisensory pair is composed by congruent and/or ethologically relevant information (Kim et al., 2008; von Kriegstein & Giraud, 2006). Kim and colleagues (2008) showed that incongruent multisensory training did not impact unisensory motion discrimination. Subjects who had been exposed to training sessions entailing auditory-visual pairs moving in opposite directions performed at similar levels than subjects trained only visually. In another study, von Kriegstein and Giraud (2006) found that recognition accuracy was significantly affected only for voice presentations that had been paired with a face upon initial presentations. No impact upon subsequent recognition accuracy was found for ring tones. Similarly, van der Linden and colleagues (2011) suggested that multisensory memories impact recognition accuracy when objects are familiar. Neuroimaging results showed that the STS only responded upon presentation of previously learned congruent auditory-visual presentations as compared to incongruent presentations. Our findings have challenged these hypotheses, by showing that single-trial incongruent and meaningless associations can impact (albeit, impair) subsequent unisensory processing. Similarly, Gottfried and colleagues (2004) found that meaningless olfactory-visual associations could impact subsequent visual object recognition (Gottfried et al., 2004).

Moreover, we have provided evidence that multisensory learning affects auditory object recognition. Up until now, only two studies have investigated similar effects within the auditory modality (Cohen, Horowitz, & Wolfe, 2009; von Kriegstein & Giraud, 2006). Cohen and colleagues (2009) found that recognition accuracy for sounds was significantly impaired compared to visual object recognition. In a subsequent experiment sounds were paired with either corresponding images or written labels. Recognition accuracy for sounds that had been paired with congruent pictures upon initial encounter did not differ from sounds only encountered in a unisensory context. Interestingly, this was not the case in our present findings. We show recognition accuracy enhancement for sounds that had been paired with congruent images upon initial encounter compared to sounds presented in a unisensory context. These discordant findings could be explained by the differential experimental setups. First, Cohen and colleagues (2009) asked their participants explicitly to commit the presented sounds into memory. On the other hand, in our study participants were asked to perform a simple old/new recognition task. Second, the experiment of Cohen and colleagues (2009) was divided into two separate sessions (study phase and test phase), while we presented stimuli in a continuous and interleaved manner (old/new trials within a same presentation block). Thus, different types of memory were engaged. While our experimental setup addressed short-term perceptual memory trace retrieval, long-term memory trace retrieval was engaged in Cohen and colleagues' task.

Our findings within the auditory modality also challenge the findings of von Kriegstein and Giraud (2006). In their study, participants only benefitted from multisensory associations upon unisensory retrieval in the case of initial voice-face pairings. Thus, the authors postulated that auditory object recognition could benefit from multisensory memories only when the auditory-visual pairing entailed ethologically valid information. We found that auditory object recognition benefitted from prior multisensory exposure whenever the coupling was congruent and independently of object category tested (animals, tools, cars, etc.). As discussed in Thelen and colleagues (*in review*) we hypothesize that these discordant results may be caused by the type of material presented to subjects (i.e. voices-faces vs. schematized objects).

9.2.3 Impact on brain responses of single-trial multisensory memories

Our collective work has shown that single-trial multisensory memories impact subsequent visual object recognition (Murray et al., 2004, 2005; Thelen et al., 2012).

Differential brain responses to images according to past encounter context were observed within early sensory processing stages (~100ms post-stimulus onset). These findings are concordant with previous work showing that multisensory memories impact unisensory object recognition (Giard & Peronnet, 1999).

Interestingly, source localization of these differential early effects differed according to the initial encounter context. When past encounters entailed congruent auditory-visual presentations, visual recognition elicited stronger responses within low-level visual areas (LOC). On the other hand, initial meaningless auditory-visual associations lead to stronger activity with the pSTG upon repeated visual trials. We have proposed that these differential results are linked to behavioral responses rather than to the semantic contingencies *per se*. In fact, we found stronger activity within the LOC for images that had been presented in a unisensory context compared to images initially encountered in a meaningless auditory-visual context. In other words, the common findings between our past and recent results suggests that LOC activity enhancement was observed upon presentation of those stimuli that are recognized more accurately within the experimental setup (i.e. upon V+c trials in Murray et al., 2004, 2005; and upon V- trials in Thelen et al., 2012, for acronyms please see Thelen & Murray, 2013). This hypothesis is further supported by the findings of Nyberg and colleagues (2000). In fact, the behavioral results showed that subjects word recognition was impaired when they had been studied in an auditory-visual context (76% vs. 84% although no statistical analysis was performed on the behavioral data). Thus, although Nyberg and colleagues (2000) have interpreted activity within auditory cortices upon visual retrieval to be evidence for "redintegration" processes, we propose that these activations are related to the recognition impairment observed at the behavioral level.

Moreover, in our recent EEG study, we found that presentations of images that had been paired with a meaningless sound upon initial encounter lead to stronger activation with the STG/MTG at ~270ms post-stimulus onset. Importantly, activity within this cluster was correlated with the decrease in accurate recognition, suggesting the involvement of these areas in more complex retrieval processes. This interpretation is supported by a hemodynamic study implicating the STS in the establishment of multisensory associations (Tanabe et al., 2005). In fact, these authors have shown that activity within the superior temporal region decreases the better an arbitrary auditory-visual association is established. In other words, Tanabe and colleagues (2005) asked subjects to learn the association between meaningless auditory and visual features. Across experimental blocks, subjects performance (yes/no, this pair of stimuli belongs together) increased, suggesting that subjects successfully learned to associate the object features. The fMRI results showed that while recognition accuracy increased over trials, activity within the STS decreased.

Overall, these findings suggest that brain activity upon unisensory repetitions of previously multisensory stimulus presentations is affected by the relative efficacy of the memory traces, rather than by the semantic context they were established in.

9.2.4 Impact of single-trial multisensory memories upon auditory recognition

In a current Master Project (Amrein, Thelen, Murray, *in preparation*), we addressed the question, of the underlying generators involved in differential sound recognition according to initial encounter context observed in our behavioral study (Thelen et al., *in review*).

The experiment entailed unisensory, congruent and meaningless auditory-visual initial pairings, and we investigated subsequent sound recognition. The paradigm was similar to what we have reported previously during the imaging studies in the visual modality. The behavioral results replicated and extended our previous findings. In addition, to the statistical analysis of the accuracy, we calculated and performed statistical analyses on the response sensitivity (*d'*) and the response criterion (*c*). These measures are derived by calculating the *Hit Rate* (correct responses), the *Misses* (trials without responses), and *False Alarms* (incorrect discrimination, i.e. responding that a stimulus is "new" instead of "old" and vice-versa). This technique is used to differentiate between perceptual (in terms of shifts in sensitivity, d') and decisional (in terms of response criterion, c) factors contributing to detection performance (Macmillan, 1985; Verde & MacMillan, 2006). We found enhanced sensitivity for sounds that had been paired with a congruent image or were coupled with a meaningless image upon initial encounter compared to sounds previously encountered in a unisensory context (see Figure 9).

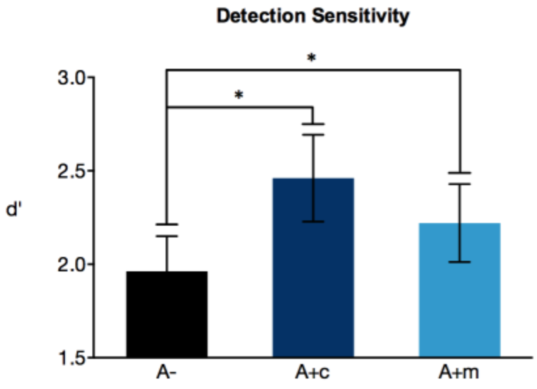
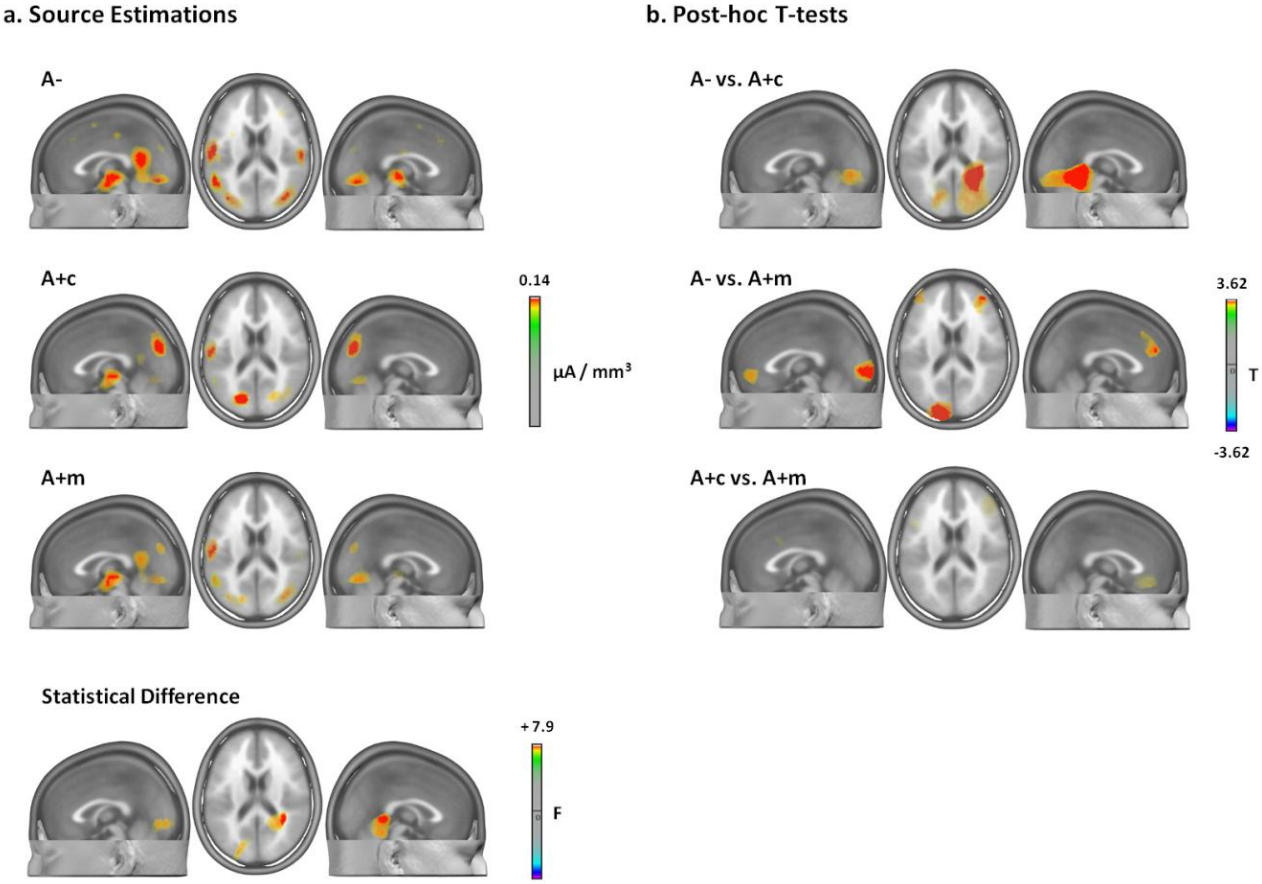


Figure 9: Sensitivity (*d'*) upon repeated presentations of sounds. Sensitivity was significantly enhanced for sounds that had been encountered in a congruent multisensory context upon initial presentations (A+c) compared to sounds encountered in a unisensory manner (A-) and sounds encountered in a meaningless multisensory context (A+m). (Figure, Courtesy of S. Amrein)

In addition to changes upon subsequent recognition accuracy, we also found differential brain responses to sounds according to single-trial multisensory encounter context. Topographical analyses of the ERPs revealed a significant main effect of encounter context at ~240-262ms post-stimulus onset. Source estimations revealed activations within clusters in the occipital and the temporo-parietal lobes to be activated upon repeated presentations of sounds. Statistical analyses revealed activity within low-level visual areas (lingual gyrus, middle occipital gyrus and the cuneus) and within more anterior-ventral- temporal areas (fusiform gyrus and the parahippocampal gyrus) to show differential activity according to the initial encounter context of sounds (see Figure 10).



a. Source Estimations

b. Post-hoc T-tests

Figure 10: (a) Source estimation of peak activity for each sound category (first three rows) over ~240-262pms post-stimulus onset. Statistically significant results of the one-way ANOVA are shown in the lowest row. (b) Post-hoc t-tests within, revealing differential activity within this network between conditions. Colors indicate *t*-values (red, stronger activity upon condition 1) (Courtesy of S. Amrein)

Generally, these results reveal that upon repeated presentation of sounds, visual cortices are recruited, independently of the initial unisensory vs. multisensory context. These results are in contrast with our prior findings, showing that under auditory conditions, visual cortices are recruited in order to perform the discrimination task, suggesting the involvement of mental imagery processes. We hypothesize that the increased task difficulty under auditory vs. visual conditions can explain this pattern. In fact, as mentioned before, we propose that sensory cortices are recruited during active retrieval according to task difficulty and performance outcome, rather than purely being linked to initial encounter context (see also next section).

This hypothesis is further supported by the results of the post-hoc t-tests performed between conditions. Recognition of sounds that had been encountered in a unisensory context upon initial presentations (A-) led to greater activations within visual areas. Recognition sensitivity under these conditions was significantly impaired compared to sounds that had been presented in a multisensory context (A+c and A+m). Interestingly, we did not find any significant differences in terms of neuronal generators when comparing sounds that had been presented either in a congruent or a meaningless multisensory context. This is not surprising, since the behavioral results showed that recognition sensitivity for sounds that had been paired with a meaningless image upon initial encounters also enhanced compared to sounds encountered in a unisensory context.

These are preliminary findings and we are currently finalizing the analyses.

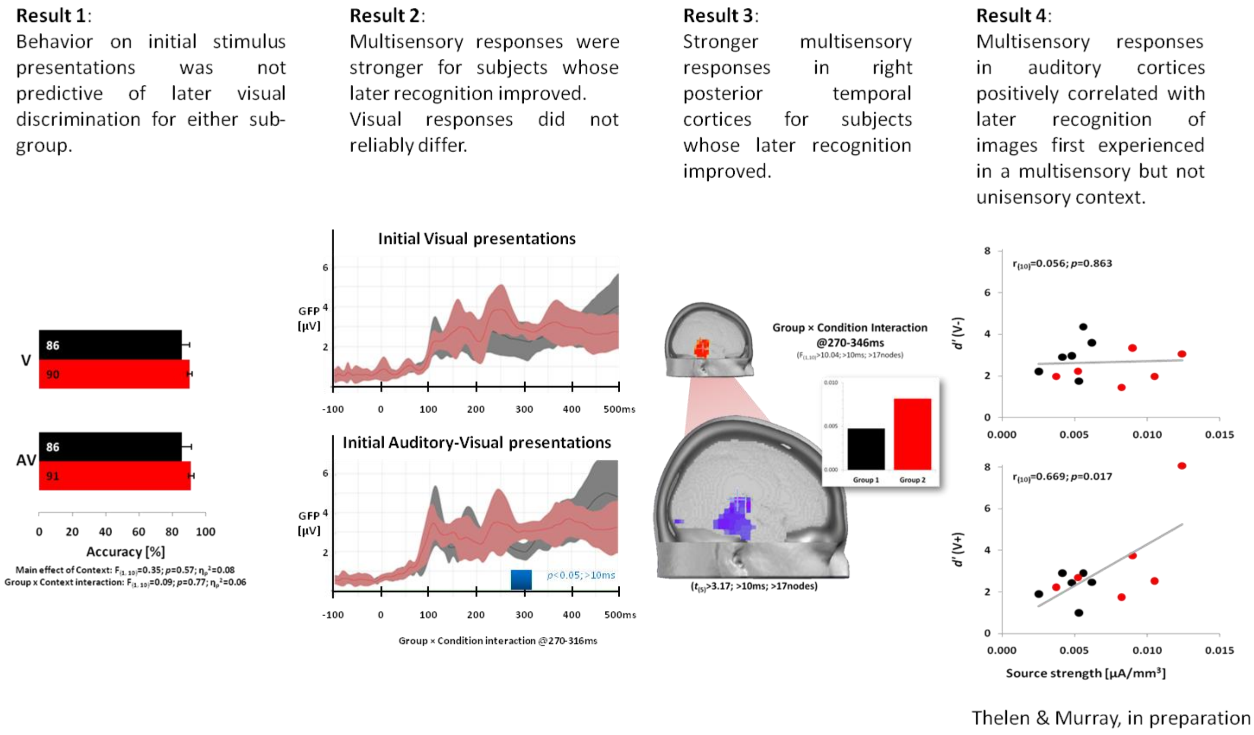
9.2.5 Predicting benefits of multisensory memories

In addition, we have addressed the question whether we could predict the efficacy of single-trial multisensory memories (Thelen & Murray, *in preparation*).

Recent findings from Naghavi and colleagues (2011) suggest that brain activity upon initial encounters is linked to the behavioral outcome (Naghavi et al., 2011). Subjects were presented with auditory-visual pairs within the scanner. Multisensory pairs were composed of congruent and incongruent auditory-visual object pairs. After the scanning sessions, subjects were asked to perform a surprise associative recognition test where they had to indicate whether a given auditory-visual pair (independently if congruent or incongruent) had been presented during the scanning session. Two clusters within the occipito-temporal and inferior frontal gyrus were found when comparing brain activation between correctly remembered and forgotten auditory-visual pairings. Further, remembered congruent (RC) pairings were associated with greater activity with the LOC, the fusiform gyrus and the inferior frontal gyrus than compared to remembered incongruent (RI) trials. Contrariwise, RI trials were associated with greater activations within the posterior and anterior STG/MTG as well as the inferior frontal gyrus.

These findings suggest the existence of links between brain activity during encoding with subsequent retrieval performance. Nonetheless, whether brain activity is predictive of behavioral outcome was not addressed. To assess the predictive value of multisensory memories, we *a posteriori* grouped subjects according to whether or not they benefitted from single-trial multisensory memories upon subsequent visual object recognition in the data set previously described in Thelen and colleagues (2012). In other words, we divided subjects into two groups according to whether recognition accuracy showed V+<V- (impairment, group 1) or V+>V- (benefit, group 2) (V+: visual repetitions of previously multisensory presentations; V- visual repetitions of previously unisensory presentations). We then analyzed the behavioral and EEG data of initial trials in order to investigate the existence of evidence of differential stimulus processing.

We did not find any indication for differential object processing in either encounter context condition (visual, V; auditory-visual AV) between groups at the behavioral level (Group x Encounter context interaction: $F_{(1,\ 10)}=0.09$; $p=0.77$; $\eta_p^2=0.06$). On the other hand, we found group differences to occur during processing of AV stimuli in terms of electric field strength (GFP) at ~270-316ms post-stimulus onset. In fact GFP was stronger within this time window for subjects in group 2 as compared to group 1. We then performed a Group by Encounter context repeated measures ANOVA within the inverse space, in order to test whether differences observed at the level of the scalp where linked to differential activation strength of the underlying neuronal generators. We found that during the same time window (~270-346ms post-stimulus onset) as the GFP difference, activity within a cluster in the STG/MTG was significantly stronger for subjects in group 2 than in group 1. Moreover, activity within this cluster upon initial multisensory encounters (AV) was correlated ($r_{(10)}=0.669$; $p=0.017$) with sensitivity (*d'*) to these images upon repeated unisensory presentations (V+) (see Figure 11).



Result 1: Behavior on initial stimulus presentations was not predictive of later visual discrimination for either sub-group.

Result 2: Multisensory responses were stronger for subjects whose later recognition improved. Visual responses did not reliably differ.

Result 3: Stronger multisensory responses in right posterior temporal cortices for subjects whose later recognition improved.

Result 4: Multisensory responses in auditory cortices positively correlated with later recognition of images first experienced in a multisensory but not unisensory context.

Thelen & Murray, in preparation

Figure 11: Summary results. There was no evidence of differential processing of initial stimulus presentation between groups at the behavioral level. ERP analyses revealed a GFP enhancement at ~270-316ms post-stimulus onset upon initial multisensory presentations between groups, in the absence of similar differences upon presentation of unisensory stimuli. Source estimations revealed enhanced activity within a cluster in the superior temporal region upon auditory-visual presentations in group 2 at ~270-346ms post-stimulus onset. Moreover, activity within this cluster was predictive of recognition sensitivity to the same stimuli upon unisensory recognition.

In line with the previously discussed findings from Tanabe and colleagues (2005), we propose that the STG/MTG activations are linked to multisensory object encoding. In fact the STS has been shown to play a key role in object feature association (Beauchamp et al., 2004). Beauchamp and colleagues (2004) have shown that the pSTS plays a central role in the formation of object feature association within and between sensory modalities. We propose that activity levels within pSTG/MTG found here are related to the encoding or retrieval of meaningless auditory-visual presentations.

More precisely, in our prior study (Thelen et al. 2012) we have reported activity within a spatially and temporally (~270-310ms post-stimulus onset) corresponding cluster to be correlated with the behavioral decrement observed upon repeated presentations images that had been previously paired with a meaningless sound (V+). The higher the activity was within this cluster upon retrieval, the greater the recognition impairment observed at the behavioral level. Contrariwise, the activity within this cluster upon initial encounters (AV) is correlated with a response sensitivity enhancement upon subsequent trails (V+). Thus, the activity within the pSTG/MTG differs according to the efficacy of multisensory memories to improve or impair unisensory recognition. As mentioned above, recognition sensitivity (*d'*) is a measure derived from signal detection theory. We propose that the differential recruitment of pSTG/MTG is related to inter-subject variability in the choice of strategy adopted to perform the task. In fact, although the explicit task-instructions are to ignore the task-irrelevant sounds and to attend solely to stimuli in the visual stream, subjects might have engaged differently in the active suppression of the auditory stimuli. Subjects in group 1, attended exclusively to the visual stream, (most probably) actively suppressing the integration of the auditory stimuli. This might have lead to a cost in terms of the deployment of attentional resources and consequently the establishment of unambiguous perceptual memory traces. On the other hand, subjects in group 2 processed the task-irrelevant auditory stimuli. This is supported by the neuroimaging results, showing greater activity for group 2 upon auditory-visual trials providing evidence for multisensory object association processes. Consequently, whether subjects' recognition accuracy for images that had been paired with a meaningless sound upon initial encounter is enhanced vs. impaired compared to images presented alone, depends on the strategy adopted to perform the task. This hypothesis will have to be addressed by manipulating the task-instructions in the future.

10 Conclusion and Outlook

During my thesis I have investigated the neuronal mechanisms underlying multisensory perceptual enhancement and the efficacy of multisensory memories.

In both projects we have found evidence that multisensory interactions impact early stages of sensory processing within low-level sensory areas. Moreover, the results presented here have shown that these early auditory-visual integrations effects are relevant for behavior. In the first project concurrent auditory-visual presentations lead to speeded responses in a motion detection task. On the other hand, we have shown that multisensory memory traces impact subsequent unisensory processing, and that activity upon multisensory encoding are predictive of unisensory retrieval performance.

Nonetheless, future investigations will have to address several open questions. For example, in the motion detection task, we have investigated the interaction between the spatial principle of multisensory integration and the PoIE. In a next step, the temporal rule should be integrated to the experimental plan. Further, recent studies have suggested that such low-level auditory-visual interactions are mediated by direct monosynaptic connections between primary and peri-primary sensory areas in monkeys (Falchier et al., 2010; Musacchia & Schroeder, 2009). Whether such findings extend to human perception are still unknown, although evidence has been reported suggesting that this is the case (Beer et al., 2011, 2013; Romei et al., 2012, 2009).

Clarifying whether low-level neuronal connections are susceptible to implicit training is of crucial importance for clinical research. Recent work has shown that the emergence and development of multisensory processing continues into late childhood and adolescence (Fredembach, de Boisferon, & Gentaz, 2009; Gori, Sandini, & Burr, 2012; Wallace, Carriere, Perrault, Vaughan, & Stein, 2006) as well as disrupted multisensory processing within clinical population with autism spectrum disorder (Wallace, 2009; Woynaroski et al., 2013). How they develop, and whether deficiencies observed in clinical populations can be addressed with simple training protocols has been insufficiently addressed until now. These findings also suggest that the low-level cortical interactions investigated during my thesis are susceptible to play a central role in the development of normal and pathological multisensory integration mechanisms.

Nonetheless, the exact mechanisms underpinning normal and pathological emergence of multisensory processing are unknown.

Similarly, the application of multisensory protocols in aging to improve life quality and or in rehabilitation procedures needs further investigation. For example, whether amnesic patients can benefit from single-trial multisensory perceptual memories has to be investigated. Also, whether the efficacy of these memories can be further enhanced when using more realistic materials (i.e. photographs instead of schematized line drawings) is of relevance for memory rehabilitation protocols. Further, the specificity of the memory trace has not been addressed. More precisely, whether the memory traces involved in our study are purely perceptual (stimulus-specific priming) and how much long-term memory representations play a role in this recognition task (semantic-priming) has to be addressed in the future. Due to the great variability in specific deficits reported in case studies, a greater effort towards single-patient adapted rehabilitation procedures has to be made. Together, these studies will reveal whether the continuous recognition task presented in our research can be adapted to the individual demands of the patient (specific material, rehabilitation of semantic vs. perceptual deficits...etc.).

11 References

Amedi, a, Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature neuroscience*, *4*(3), 324–30. doi:10.1038/85201

Amedi, a, von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *166*(3-4), 559–71. doi:10.1007/s00221-005-2396-5

Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral cortex (New York, N.Y.: 1991)*, *12*(11), 1202–12. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12379608

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., … Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature neuroscience*, *10*(6), 687–9. doi:10.1038/nn1912

Atteveldt, N. Van, Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of Letters and Speech Sounds in the Human Brain. *Neuron*, *43*(July), 271–282.

Auvray, M., Hanneton, S., Lenay, C., & O'Regan, K. (2005). There is something out there: distal attribution in sensory substitution, twenty years later. *Journal of integrative neuroscience*, *4*(4), 505–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16385644

Avillac, M., Ben Hamed, S., & Duhamel, J.-R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *27*(8), 1922–32. doi:10.1523/JNEUROSCI.2646-06.2007

Bach, D. R., Neuhoff, J. G., Perrig, W., & Seifritz, E. (2009). Looming sounds as warning signals: the function of motion cues. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology*, *74*(1), 28–33. doi:10.1016/j.ijpsycho.2009.06.004

Bach, D. R., Schächinger, H., Neuhoff, J. G., Esposito, F., Di Salle, F., Lehmann, C., … Seifritz, E. (2008). Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cerebral cortex (New York, N.Y.: 1991)*, *18*(1), 145–50. doi:10.1093/cercor/bhm040

Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, *221*(5184), 963–4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5818337

Bach-y-rita, P., & Kercel, S. W. (2003). Sensory substitution and the human – machine interface, *7*(12), 541–546. doi:10.1016/j.tics.2003.10.013

Barth, D. S., Goldberg, N., Brett, B., & Di, S. (1995). The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. *Brain Research*, *678*(1), 177–190. Retrieved from http://www.sciencedirect.com/science/article/pii/000689939500182P

Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Current opinion in neurobiology*, *15*(2), 145–53. doi:10.1016/j.conb.2005.03.011

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15813999

Beer, A. L., Plank, T., & Greenlee, M. W. (2011). Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *213*(2-3), 299–308. doi:10.1007/s00221-011-2715-y

Beer, A. L., Plank, T., Meyer, G., & Greenlee, M. W. (2013). Combined diffusion-weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing. *Frontiers in integrative neuroscience*, *7*(February), 5. doi:10.3389/fnint.2013.00005

Bernstein, L. E., Auer, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, *44*(1-4), 5–18. doi:10.1016/j.specom.2004.10.011

Bertelson, P. (1999). Ventriloquism: A case of cross-modal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 347–362). Amsterdam: Elsevier.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & psychophysics*, *29*(6), 578–84. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7279586

Bizley, J. K., Nodal, F. R., Bajo, V. M., Nelken, I., & King, A. J. (2007). Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cerebral cortex (New York, N.Y.: 1991)*, *17*(9), 2172–89. doi:10.1093/cercor/bhl128

Blanke, O., & Arzy, S. (2005). The out-of-body experience: disturbed self-processing at the temporo-parietal junction. *The Neuroscientist: a review journal bringing neurobiology, neurology and psychiatry*, *11*(1), 16–24. doi:10.1177/1073858404270885

Bolognini, N., Senna, I., Maravita, A., Pascual-Leone, A., & Merabet, L. B. (2010). Auditory enhancement of visual phosphene perception: the effect of temporal and spatial factors and of stimulus intensity. *Neuroscience letters*, *477*(3), 109–14. doi:10.1016/j.neulet.2010.04.044

Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. *Computational intelligence and neuroscience*, *2011*, 813870. doi:10.1155/2011/813870

Budinger, E., Laszcz, A., Lison, H., Scheich, H., & Ohl, F. W. (2008). Non-sensory cortical and subcortical connections of the primary auditory cortex in Mongolian gerbils: bottom-up and top-down processing of neuronal information via field AI. *Brain research*, *1220*, 2–32. doi:10.1016/j.brainres.2007.07.084

Bueti, D., & Macaluso, E. (2010). Auditory temporal expectations modulate activity in visual cortex. *NeuroImage*, *51*(3), 1168–83. doi:10.1016/j.neuroimage.2010.03.023

Calvert, G. a, Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current biology: CB*, *10*(11), 649–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10837246

Calvert, G. a, Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*, *14*(2), 427–38. doi:10.1006/nimg.2001.0812

Capelle, C., Trullemans, C., Arno, P., & Veraart, C. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE transactions on bio-medical engineering*, *45*(10), 1279–93. doi:10.1109/10.720206

Cappe, C, Morel, a, & Rouiller, E. M. (2007). Thalamocortical and the dual pattern of corticothalamic projections of the posterior parietal cortex in macaque monkeys. *Neuroscience*, *146*(3), 1371–87. doi:10.1016/j.neuroscience.2007.02.033

Cappe, C, Morel, A., Barone, P., & Rouiller, E. M. (2009). The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. *Cerebral cortex (New York, N.Y.: 1991)*, *19*(9), 2025–37. doi:10.1093/cercor/bhn228

Cappe, C, Rouiller, E. M., & Barone, P. (2009). Multisensory anatomical pathways. *Hearing research*, *258*(1-2), 28–36. doi:10.1016/j.heares.2009.04.017

Cappe, Céline, Thelen, A., Romei, V., Thut, G., & Murray, M. M. (2012). Looming signals reveal synergistic principles of multisensory integration. *Journal of Neuroscience*, *32*(4), 1171–1182. doi:10.1523/JNEUROSCI.5517-11.2012

Cappe, Céline, Thut, G., Romei, V., & Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia*, *47*(4), 1045–52. doi:10.1016/j.neuropsychologia.2008.11.003

Cappe, Céline, Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *30*(38), 12572–80. doi:10.1523/JNEUROSCI.1099-10.2010

Carriere, B. N., Royal, D. W., Wallace, M. T., & Carolina, N. (2008). Spatial heterogeneity of cortical receptive fields and its impact on multisensory interactions. *Journal of neurophysiology*, *99*(5), 2357–68. doi:10.1152/jn.01386.2007

Chen, Y., Dhamala, M., Bollimunta, A., Schroeder, C. E., & Ding, M. (2011). Current Source Density Analysis of Ongoing Neural Activity: Theory and Application. In R. P. Vertes & R. W. J. Stackman (Eds.), *Electrophysiological Recording Techniques* (Vol. 54, pp. 27–41). Springer Science+Business Media. doi:10.1007/978-1-60327-202-5

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404. doi:10.1016/j.cognition.2009.10.012

Cohen, M. a, Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 6008–6010. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2667065&tool=pmcentrez&rendertype=abstract

Colombo, M., & Gross, C. G. (1994). Responses of inferior temporal cortex and hippocampal neurons during delayed matching to sample in monkeys (Macaca fascicularis). *Behavioral neuroscience*, *108*(3), 443–55. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7917038

Dahl, C. D., Logothetis, N. K., & Kayser, C. (2010). Modulation of visual responses in the superior temporal sulcus by audio-visual congruency. *Frontiers in integrative neuroscience*, *4*(April), 10. doi:10.3389/fnint.2010.00010

De Gelder, B., Vroomen, J., Annen, L., Masthof, E., & Hodiamont, P. (2003). Audio-visual integration in schizophrenia. *Schizophrenia research*, *59*(2-3), 211–8. Retrieved from http://www.schres-journal.com/article/S0920-9964(01)00344-9/abstract

De Gelder, B., Vroomen, J., de Jong, S. J., Masthoff, E. D., Trompenaars, F. J., & Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics. *Schizophrenia research*, *72*(2-3), 195–203. doi:10.1016/j.schres.2004.02.013

De Vries, S. E. J., & Clandinin, T. R. (2012). Loom-sensitive neurons link computation to action in the Drosophila visual system. *Current biology: CB*, *22*(5), 353–62. doi:10.1016/j.cub.2012.01.007

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain research*, *1242*, 136–50. doi:10.1016/j.brainres.2008.03.071

Doniger, G. M., Foxe, J. J., Schroeder, C. E., Murray, M. M., Higgins, B. a, & Javitt, D. C. (2001). Visual perceptual learning in human object recognition areas: a repetition priming study using high-density electrical mapping. *NeuroImage*, *13*(2), 305–13. doi:10.1006/nimg.2000.0684

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. *Neuron*, *57*(1), 11–23. doi:10.1016/j.neuron.2007.12.013

Falchier, A., Schroeder, C. E., Hackett, T. a, Lakatos, P., Nascimento-Silva, S., Ulbert, I., … Smiley, J. F. (2010). Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cerebral cortex (New York, N.Y.: 1991)*, *20*(7), 1529–38. doi:10.1093/cercor/bhp213

Foss-Feig, J. H., Tadin, D., Schauder, K. B., & Cascio, C. J. (2013). A substantial and unexpected enhancement of motion perception in autism. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *33*(19), 8243–9. doi:10.1523/JNEUROSCI.1608-12.2013

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport*, *16*(5), 419–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15770144

Fredembach, B., de Boisferon, A. H., & Gentaz, E. (2009). Learning of arbitrary association between visual and auditory novel stimuli in adults: the "bond effect" of haptic exploration. *PloS one*, *4*(3), e4844. doi:10.1371/journal.pone.0004844

Fuster, Â. M., Bodner, M., & Kroger, J. K. (2000). association in neurons of frontal cortex, *405*(May), 347–351.

Ghazanfar, A. a, Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *25*(20), 5004–12. doi:10.1523/JNEUROSCI.0799-05.2005

Ghazanfar, A. a, & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in cognitive sciences*, *10*(6), 278–85. doi:10.1016/j.tics.2006.04.008

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, *11*(5), 473–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10511637

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. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9310418

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, *15*(4), 511–7. doi:10.1038/nn.3063

Gori, M., Sandini, G., & Burr, D. (2012). Development of visuo-auditory integration in space and time. *Frontiers in integrative neuroscience*, *6*(September), 77. doi:10.3389/fnint.2012.00077

Gottfried, J. a, Smith, A. P. R., Rugg, M. D., & Dolan, R. J. (2004). Remembrance of odors past: human olfactory cortex in cross-modal recognition memory. *Neuron*, *42*(4), 687–95. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15157428

Gross, C. G., Rodman, H. R., Gochin, P. M., & Colombo, M. W. (1993). Inferior temporal cortex as a pattern recognition device. In E. Baum (Ed.), *Computational learning and recognition: Proceedings of the 3rd NEC research symposium* (pp. 44–73). Philadelphia: PA: Society for Industrial and Applied mathematics.

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *The Journal of comparative neurology*, *502*(6), 924–52. doi:10.1002/cne.21326

Haenny, P. E., Maunsell, J. H., & Schiller, P. H. (1988). State dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *69*(2), 245–59. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3345806

Haigh, A., Brown, D. J., Meijer, P., & Proulx, M. J. (2013). How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Frontiers in psychology*, *4*(June), 330. doi:10.3389/fpsyg.2013.00330

Hamann, S. B., & Squire, L. R. (1997). Intact perceptual memory in the absence of conscious memory.

Harrison, N. (2012). Auditory motion in depth is preferentially “captured” by visual looming signals. *Seeing and perceiving*, *25*(1), 71–85. doi:10.1163/187847611X620928

Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: multisensory integration versus normal variability. *Neuropsychologia*, *45*(14), 3340–5. doi:10.1016/j.neuropsychologia.2007.05.025

Innocenti, G. M., Berbel, P., & Clarke, S. (1988). Development of projections from auditory to visual areas in the cat. *The Journal of comparative neurology*, *272*(2), 242–59. doi:10.1002/cne.902720207

Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., Chapuis, D., … Blanke, O. (2011). Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. *Neuron*, *70*(2), 363–74. doi:10.1016/j.neuron.2011.03.009

James, C., Morand, S., Barcellona-Lehmann, S., Michel, C. M., & Schnider, A. (2009). Neural transition from short- to long-term memory and the medial temporal lobe: a human evoked-potential study. *Hippocampus*, *19*(4), 371–8. doi:10.1002/hipo.20526

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S., & Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, *40*(10), 1706–1714. Retrieved from http://www.sciencedirect.com/science/article/pii/S0028393202000179

Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W., & Stein, B. E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *Journal of neurophysiology*, *85*(2), 506–22. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11160489

JONES, E. G., & POWELL, T. P. S. (1970). AN ANATOMICAL STUDY OF CONVERGING SENSORY PATHWAYS WITHIN THE CEREBRAL CORTEX OF THE MONKEY. *Brain*, *93*(4), 793–820. doi:10.1093/brain/93.4.793

Joubert, O. R., Rousselet, G. a, Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: fast categorization and object interference. *Vision research*, *47*(26), 3286–97. doi:10.1016/j.visres.2007.09.013

Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain structure & function*, *212*(2), 121–32. doi:10.1007/s00429-007-0154-0

Kayser, C., & Logothetis, N. K. (2009). Directed Interactions Between Auditory and Superior Temporal Cortices and their Role in Sensory Integration. *Frontiers in integrative neuroscience*, *3*, 7. doi:10.3389/neuro.07.007.2009

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2009). Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. *Hearing research*, *258*(1-2), 80–8. doi:10.1016/j.heares.2009.02.011

Keane, M. M., Gabrieli, J. D., Mapstone, H. C., Johnson, K. A., & Corkin, S. (1995). Double dissociation of memory capacities after bilateral occipital-lobe or medial temporal-lobe lesions. *Brain: a journal of neurology*, *118 (Pt 5*, 1129–48.

Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PloS one*, *3*(1), e1532. doi:10.1371/journal.pone.0001532

Kim, S. G., Richter, W., & Uğurbil, K. (1997). Limitations of temporal resolution in functional MRI. *Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *37*(4), 631–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9094089

Krueger, J., Royal, D. W., Fister, M. C., & Wallace, M. T. (2009). Spatial receptive field organization of multisensory neurons and its impact on multisensory interactions. *Hearing research*, *258*(1-2), 47–54. doi:10.1016/j.heares.2009.08.003

Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*(2), 279–92. doi:10.1016/j.neuron.2006.12.011

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of neurophysiology*, *94*(3), 1904–11. doi:10.1152/jn.00263.2005

Lehmann, S., Morand, S., James, C., & Schnider, A. (2007). Electrophysiological correlates of deficient encoding in a case of post-anoxic amnesia. *Neuropsychologia*, *45*(8), 1757–66. doi:10.1016/j.neuropsychologia.2006.12.018

Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Brain research. Cognitive brain research*, *24*(2), 326–34. doi:10.1016/j.cogbrainres.2005.02.005

Lenggenhager, B., Smith, S. T., & Blanke, O. (2006). Functional and neural mechanisms of embodiment: importance of the vestibular system and the temporal parietal junction. *Reviews in the neurosciences*, *17*(6), 643–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17283609

Leo, F., Romei, V., Freeman, E., Ladavas, E., & Driver, J. (2011). Looming sounds enhance orientation sensitivity for visual stimuli on the same side as such sounds. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *213*(2-3), 193–201. doi:10.1007/s00221-011-2742-8

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. *NeuroImage*, *21*(2), 725–32. doi:10.1016/j.neuroimage.2003.09.049

Macmillan NA, K. H. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological bulletin*, *98*(1), 185–199.

Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. a. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current biology: CB*, *18*(13), 963–8. doi:10.1016/j.cub.2008.05.043

Maier, J. X., & Ghazanfar, A. a. (2007). Looming biases in monkey auditory cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *27*(15), 4093–100. doi:10.1523/JNEUROSCI.0330-07.2007

Maier, J. X., Neuhoff, J. G., Logothetis, N. K., & Ghazanfar, A. A. (2004). Multisensory Integration of Looming Signals by Rhesus Monkeys Max Planck Institute for Biological Cybernetics, *43*, 177–181.

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J.-P., Maeder, P. P., Clarke, S., & Meuli, R. A. (2007). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cerebral cortex (New York, N.Y.: 1991)*, *17*(7), 1672–9. doi:10.1093/cercor/bhl077

Maunsell, J. H., Sclar, G., Nealey, T. A., & DePriest, D. D. (1991). Extraretinal representations in area V4 in the macaque monkey. *Visual neuroscience*, *7*(6), 561–73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1772806

Mazzoni, P., Bracewell, R. M., Barash, S., & Andersen, R. a. (1996). Spatially tuned auditory responses in area LIP of macaques performing delayed memory saccades to acoustic targets. *Journal of neurophysiology*, *75*(3), 1233–41. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8867131

MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. doi:10.1038/264746a0

Merabet, L. B., Battelli, L., Obretenova, S., Maguire, S., Meijer, P., & Pascual-Leone, A. (2009). Functional recruitment of visual cortex for sound encoded object identification in the blind. *Neuroreport*, *20*(2), 132–8. doi:10.1097/WNR.0b013e32832104dc

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science (New York, N.Y.)*, *221*(4608), 389–91.

Meylan, R. V, & Murray, M. M. (2007). Auditory-visual multisensory interactions attenuate subsequent visual responses in humans. *NeuroImage*, *35*(1), 244–54. doi:10.1016/j.neuroimage.2006.11.033

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical neurophysiology⍰: official journal of the International Federation of Clinical Neurophysiology*, *115*(10), 2195–222. doi:10.1016/j.clinph.2004.06.001

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & psychophysics*, *40*(5), 331–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3786102

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of experimental psychology. Human perception and performance*, *17*(1), 160–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1826309

Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *The Journal of neuroscience⍰: the official journal of the Society for Neuroscience*, *25*(25), 5884–93. doi:10.1523/JNEUROSCI.0896-05.2005

Miller, M. W., & Vogt, B. A. (1984). Direct connections of rat visual cortex with sensory, motor, and association cortices. *The Journal of comparative neurology*, *226*(2), 184–202. doi:10.1002/cne.902260204

Minamimoto, T., Saunders, R. C., & Richmond, B. J. (2010). Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex. *Neuron*, *66*(4), 501–7. doi:10.1016/j.neuron.2010.04.010

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain research. Cognitive brain research*, *14*(1), 115–28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12063135

Murray, E. A., & Richmond, B. J. (2001). Role of perirhinal cortex in object perception, memory, and associations. *Current opinion in neurobiology*, *11*(2), 188–93.

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

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain topography*, *20*(4), 249–64. doi:10.1007/s10548-008-0054-5

Murray, M. M., Foxe, J. J., & Wylie, G. R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. *NeuroImage*, *27*(2), 473–8. doi:10.1016/j.neuroimage.2005.04.016

Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *NeuroImage*, *21*(1), 125–135. doi:10.1016/j.neuroimage.2003.09.035

Murray, M. M., & Sperdin, H. F. (2010). Single-trial multisensory learning and memory retrieval. In J. Kaiser & M. J. Naumer (Eds.), *Multisensory Object Perception in the Primate Brain* (pp. 191–208). Heidelberg, Germany: Springer.

Musacchia, G., & Schroeder, C. E. (2009). Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hearing research*, *258*(1-2), 72–9. doi:10.1016/j.heares.2009.06.018

Naci, L., Taylor, K. I., Cusack, R., & Tyler, L. K. (2012). Are the senses enough for sense? Early high-level feedback shapes our comprehension of multisensory objects. *Frontiers in integrative neuroscience*, *6*(September), 82. doi:10.3389/fnint.2012.00082

Naghavi, H. R., Eriksson, J., Larsson, A., & Nyberg, L. (2011). Cortical regions underlying successful encoding of semantically congruent and incongruent associations between common auditory and visual objects. *Neuroscience letters*, *505*(2), 191–5. doi:10.1016/j.neulet.2011.10.022

Nyberg, L., Habib, R., McIntosh, a R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11120–4. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=27158&tool=pmcentrez&rendertype=abstract

Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *The Journal of neuroscience□: the official journal of the Society for Neuroscience*, *33*(17), 7463–74. doi:10.1523/JNEUROSCI.4678-12.2013

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & cognition*, *21*(2), 156–61. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3338118&tool=pmcentrez&rendertype=abstract

Poulter, D. R., & Wann, J. P. (2013). Errors in motion processing amongst older drivers may increase accident risk. *Accident; analysis and prevention*, *57*, 150–6. doi:10.1016/j.aap.2013.03.031

Powers, A. R., Hevey, M. a, & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *32*(18), 6263–74. doi:10.1523/JNEUROSCI.6138-11.2012

Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *29*(39), 12265–74. doi:10.1523/JNEUROSCI.3501-09.2009

Proulx, M. J., Brown, D. J., Pasqualotto, A., & Meijer, P. (2012). Multisensory perceptual learning and sensory substitution. *Neuroscience and biobehavioral reviews*, 1–10. doi:10.1016/j.neubiorev.2012.11.017

Purcell, C., Wann, J. P., Wilmut, K., & Poulter, D. (2012). Reduced looming sensitivity in primary school children with Developmental Co-ordination Disorder. *Developmental science*, *15*(3), 299–306. doi:10.1111/j.1467-7687.2011.01123.x

RAAB, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–90.

Rauschecker, J. P., & Tian, B. (2004). Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *Journal of neurophysiology*, *91*(6), 2578–89. doi:10.1152/jn.00834.2003

Reich, L., Maidenbaum, S., & Amedi, A. (2012). The brain as a flexible task machine: implications for visual rehabilitation using noninvasive vs . invasive approaches, *25*(1), 86–95. doi:10.1097/WCO.0b013e32834ed723

Richmond, B. J., & Sato, T. (1987). Enhancement of inferior temporal neurons during visual discrimination. *Journal of neurophysiology*, *58*(6), 1292–306.

Romanski, L. M. (2007). Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cerebral cortex (New York, N.Y.: 1991)*, *17 Suppl 1*, i61–9. doi:10.1093/cercor/bhm099

Romei, V., Gross, J., & Thut, G. (2012). Sounds reset rhythms of visual cortex and corresponding human visual perception. *Current biology: CB*, *22*(9), 807–13. doi:10.1016/j.cub.2012.03.025

Romei, V., Murray, M. M., Cappe, C., & Thut, G. (2009). Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Current biology: CB*, *19*(21), 1799–805. doi:10.1016/j.cub.2009.09.027

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature neuroscience*, *5*(7), 629–30. doi:10.1038/nn866

Sarko, D. K., Ghose, D., & Wallace, M. T. (2013). Convergent approaches toward the study of multisensory perception. *Frontiers in Systems Neuroscience*, *7*, 81. doi:10.3389/fnsys.2013.00081

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. doi:10.1016/j.neuroimage.2008.05.033

Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain research. Cognitive brain research*, *14*(1), 187–98. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12063142

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in cognitive sciences*, *12*(3), 106–13. doi:10.1016/j.tics.2008.01.002

Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory Input to Auditory Association Cortex in the Macaque Monkey. *J Neurophysiol*, *85*(3), 1322–1327. Retrieved from http://jn.physiology.org/content/85/3/1322.long

Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, K., Mustovic, H., Schächinger, H., … Di Salle, F. (2002). Neural processing of auditory looming in the human brain. *Current biology: CB*, *12*(24), 2147–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12498691

Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current biology: CB*, *16*(14), 1422–7. doi:10.1016/j.cub.2006.05.048

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature Communications*, *408*(December), 2000.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain research. Cognitive brain research*, *14*(1), 147–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12063138

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences*, *12*(11), 411–7. doi:10.1016/j.tics.2008.07.006

Shams, L., Wozny, D. R., Kim, R., & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in psychology*, *2*(October), 264. doi:10.3389/fpsyg.2011.00264

Sherman, S. M. (2007). The thalamus is more than just a relay. *Current opinion in neurobiology*, *17*(4), 417–22. doi:10.1016/j.conb.2007.07.003

Smiley, J. F., & Falchier, A. (2009). Multisensory connections of monkey auditory cerebral cortex. *Hearing research*, *258*(1-2), 37–46. doi:10.1016/j.heares.2009.06.019

Spence, C, Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & psychophysics*, *63*(2), 330–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11281107

Spence, Charles. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*. doi:10.1111/nyas.12121

Sperdin, H. F., Cappe, C., & Murray, M. M. (2010). Auditory-somatosensory multisensory interactions in humans: dissociating detection and spatial discrimination. *Neuropsychologia*, *48*(13), 3696–705. doi:10.1016/j.neuropsychologia.2010.09.001

Spierer, L., Manuel, A. L., Bueti, D., & Murray, M. M. (2013). ScienceDirect Contributions of pitch and bandwidth to sound-induced enhancement of visual cortex excitability in humans. *CORTEX*, *49*(10), 2728–2734. doi:10.1016/j.cortex.2013.01.001

Stark, C. E. L., & Squire, L. R. (2000). Recognition memory and familiarity judgments in severe amnesia: No evidence for a contribution of repetition priming. *Behavioral Neuroscience*, *114*(3), 459–467. doi:10.1037//0735-7044.114.3.459

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

Stein, Barry E, & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature reviews. Neuroscience*, *9*(4), 255–66. doi:10.1038/nrn2331

Stevenson, R. a, & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *227*(2), 249–61. doi:10.1007/s00221-013-3507-3

Striem-Amit, E., Cohen, L., Dehaene, S., & Amedi, A. (2012). Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron*, *76*(3), 640–52. doi:10.1016/j.neuron.2012.08.026

Sun, H., & Frost, B. J. (1998). Computation of different optical variables of looming objects in pigeon nucleus rotundus neurons. *Nature neuroscience*, *1*(4), 296–303. doi:10.1038/1110

Tajadura-Jiménez, A., Kitagawa, N., Väljamäe, A., Zampini, M., Murray, M. M., & Spence, C. (2009). Auditory-somatosensory multisensory interactions are spatially modulated by stimulated body surface and acoustic spectra. *Neuropsychologia*, *47*(1), 195–203. doi:10.1016/j.neuropsychologia.2008.07.025

Tanabe, H. C., Honda, M., & Sadato, N. (2005). Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. *The Journal of neuroscience□: the official journal of the Society for Neuroscience*, *25*(27), 6409–18. doi:10.1523/JNEUROSCI.0636-05.2005

Taylor, K. I., Moss, H. E., Stamatakis, E. a, & Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(21), 8239–44. doi:10.1073/pnas.0509704103

Taylor, K. I., Stamatakis, E. a, & Tyler, L. K. (2009). Crossmodal integration of object features: voxel-based correlations in brain-damaged patients. *Brain*, *132*, 671–83. doi:10.1093/brain/awn361

Teder-Sälejärvi, W. . A., McDonald, J. . J., Di Russo, F., & Hillyard, S. . A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research*, *14*(1), 106–114. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12063134

Teder-Sälejärvi, W. a, Di Russo, F., McDonald, J. J., & Hillyard, S. a. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of cognitive neuroscience*, *17*(9), 1396–409. doi:10.1162/0898929054985383

Temple, E., Deutsch, G. K., Poldrack, R. A., Miller, S. L., Tallal, P., Merzenich, M. M., & Gabrieli, J. D. E. (2003). Neural deficits in children with dyslexia ameliorated by behavioral remediation: evidence from functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(5), 2860–5. doi:10.1073/pnas.0030098100

Thelen, A., Cappe, C., & Murray, M. M. (2012). Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *NeuroImage*, *62*(3), 1478–1488. doi:10.1016/j.neuroimage.2012.05.027

Thelen, A., & Murray, M. M. (2013). The Efficacy of Single-Trial Multisensory Memories. *Multisensory Research*, *0*(5), 1–21. doi:10.1163/22134808-00002426

Tyll, S., Bonath, B., Schoenfeld, M. A., Heinze, H.-J., Ohl, F. W., & Noesselt, T. (2013). Neural basis of multisensory looming signals. *NeuroImage*, *65*, 13–22. doi:10.1016/j.neuroimage.2012.09.056

Van den Brink, R. L., Cohen, M. X., van der Burg, E., Talsma, D., Vissers, M. E., & Slagter, H. a. (2013). Subcortical, Modality-Specific Pathways Contribute to Multisensory Processing in Humans. *Cerebral cortex (New York, N.Y.: 1991)*, 1–9. doi:10.1093/cercor/bht069

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & 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. doi:10.1037/0096-1523.34.5.1053

Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *NeuroImage*, *55*(3), 1208–18. doi:10.1016/j.neuroimage.2010.12.068

Van der Linden, M., van Turennout, M., & Fernández, G. (2011). Category training induces cross-modal object representations in the adult human brain. *Journal of cognitive neuroscience*, *23*(6), 1315–31. doi:10.1162/jocn.2010.21522

Van der Linden, M., van Turennout, M., & Indefrey, P. (2010). Formation of category representations in superior temporal sulcus. *Journal of cognitive neuroscience*, *22*(6), 1270–82. doi:10.1162/jocn.2009.21270

Vander Wyk, B. C., Ramsay, G. J., Hudac, C. M., Jones, W., Lin, D., Klin, A., … Pelphrey, K. a. (2010). Cortical integration of audio-visual speech and non-speech stimuli. *Brain and cognition*, *74*(2), 97–106. doi:10.1016/j.bandc.2010.07.002

Verde ME, MacMillan NA, R. C. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: the accuracy, precision, and robustness of d’, Az, and A'. *Perception & Psychophysics*, *68*(4), 643–654.

Von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS biology*, *4*(10), e326. doi:10.1371/journal.pbio.0040326

Wallace, M T, Meredith, M. a, & Stein, B. E. (1998). Multisensory integration in the superior colliculus of the alert cat. *Journal of neurophysiology*, *80*(2), 1006–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9705489

Wallace, Mark T. (2009). Dyslexia: bridging the gap between hearing and reading. *Current biology: CB*, *19*(6), R260–2. doi:10.1016/j.cub.2009.01.025

Wallace, Mark T, Carriere, B. N., Perrault, T. J., Vaughan, J. W., & Stein, B. E. (2006). The development of cortical multisensory integration. *The Journal of neuroscience□: the official journal of the Society for Neuroscience*, *26*(46), 11844–9. doi:10.1523/JNEUROSCI.3295-06.2006

Wann, J. P., Poulter, D. R., & Purcell, C. (2011). Reduced sensitivity to visual looming inflates the risk posed by speeding vehicles when children try to cross the road. *Psychological science*, *22*(4), 429–34. doi:10.1177/0956797611400917

Watanabe, J., & Iwai, E. (1991). Neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during visual fixation task. *Brain Research Bulletin*, *26*(4), 583–592. Retrieved from http://www.sciencedirect.com/science/article/pii/0361923091900996

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin*, *88*(3), 638–67. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7003641

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, *97*(20), 11125–11129. doi:10.1073/pnas.97.20.11125

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. a, Stone, W. L., & Wallace, M. T. (2013). Multisensory Speech Perception in Children with Autism Spectrum Disorders. *Journal of autism and developmental disorders*. doi:10.1007/s10803-013-1836-5

Wright, T. M. (2003). Polysensory Interactions along Lateral Temporal Regions Evoked by Audiovisual Speech. *Cerebral Cortex*, *13*(10), 1034–1043. doi:10.1093/cercor/13.10.1034

Yuval-Greenberg, S., & 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□: the official journal of the Society for Neuroscience*, *27*(5), 1090–6. doi:10.1523/JNEUROSCI.4828-06.2007

Yuval-Greenberg, S., & Deouell, L. Y. (2009). The dog's meow: asymmetrical interaction in cross-modal object recognition. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *193*(4), 603–14. doi:10.1007/s00221-008-1664-6

Zampini, M., Torresan, D., Spence, C., & Murray, M. M. (2007). Auditory-somatosensory multisensory interactions in front and rear space. *Neuropsychologia*, *45*(8), 1869–77. doi:10.1016/j.neuropsychologia.2006.12.004

Zangenehpour, S., & Zatorre, R. J. (2010). Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia*, *48*, 591–600. doi:10.1016/j.neuropsychologia.2009.10.022

Zion Golumbic, E., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a “cocktail party”. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *33*(4), 1417–26. doi:10.1523/JNEUROSCI.3675-12.2013

12 Curriculum Vitae

Name: Antonia Thelen

Contact: Faculty of Biology and Medicine - UNIL
Department of Clinical Neuroscience - CHUV
Laboratory for Investigative Neurophysiology - LINE
Rue du Bugnon 46
1011 Lausanne, Switzerland
Ph.: +41'21'314'74'76

Email:
Antonia.Thelen@chuv.ch
Thelen.Antonia@gmail.com

Primary home address:
Carrale Bacilieri 16
6500 Bellinzona
+41 (0) 91 835 44 87

Secondary home address:
Av. D'Echallens 78
1004 Lausanne
+41 (0) 78 737 12 92

Date of Birth: February 07th 1984, Bern

Nationality: German
Languages: German (mother tongue)
English
Italian
French

Education:

03.2011 – 02.2014	University of Lausanne PhD Thesis in Neuroscience Lemanic Neuroscience Doctoral School PhD thesis: *Multisensory Learning and Perceptual Enhancement*
03.2009 – 03.2011	University of Geneva Interdisciplinary Master in Neuroscience Master Thesis: *The role of single-trial, episodic multisensory learning in unisensory object discrimination*
10.2005 – 06.2009	University of Lausanne Bachelor in Psychology and Sport Sciences Personal Bachelor work in Psychophysiology:

	Motor learning and the role of motivation (*L'Apprentissage moteur et l'implication de la Motivation)*
09.2000 – 06.2005	Scuola Media Superiore, Liceo Cantonale Bellinzona Main branch: Biology and Chemistry Maturity Thesis in Latin and Greek: The origins of theatre (*Le origini del teatro)*
09. 1999 – 02.2000	Mathematisch-Naturwissenschaftliches Gymnasium Neufeld, Bern Main branch: Biology and Chemistry
08.1996 – 06.1999	Secondary School Hochfeld, Bern
08.1992 – 06.1996	Primary School, Grosses Länggass-Schulhaus
08.1992 – 06.1994	Primary School, Länggasse - Muessmatt, Bern
05.1991 – 06.1992	Primary School, Spitalacker - Viktoriaschulhaus, Bern
09.1990 – 04.1991	First Grade, New York City

Publications:

Research Articles:

Cappe C*, **Thelen A*,** Romei V, Thut G, Murray MM.(2012). Looming signals reveal synergistic principles of multisensory integration. *J Neurosci. 25;32(4)*,1171-82.

Thelen A*, Cappe C, Murray MM. (2012). Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *Neuroimage. 62(3)*,1478-88.

Thelen A*, Talsma D, Murray MM. (2013). Single-trial multisensory memories affect later visual and auditory object recognition. *Journal Cognition*, *in review*.

Thelen A*, Murray MM. (2013). Predicting benefits of multisensory memories. *In preparation.*

Reviews:

Thelen A*, Murray MM**.** (2013). The efficacy if single-trial multisensory memories. *Multisensory Research*, *26(5)* 438-502.

Book Chapters:

Draganski B, **Thelen A**. (2012). Ontogenese und Plastizität des Gehirns. In *Entwicklungspsychologie*, Schneider/Lindenberger(Hrsg.), Julius Beltz GmbH Co. KG, Weinheim

Published Abstracts:

Thelen A*, Murray MM. (2012). Determinants of the efficacy of single-trial multisensory learning. *Seeing and Perceiving. 25(Supplement 1)*, 39-39

Thelen A*, Murray MM. (2012). Heterogeneous auditory-visual integration: Effects of pitch, band-width and visual eccentricity. *Seeing and Perceiving. 25(Supplement 1)*, 89-89

Thelen A*, Murray MM. (2012). Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *Seeing and Perceiving. 25(Supplement 1)*, 180-180

Thelen A*, Murray MM. (2013). Predicting individual differences in the impact of multisensory, single-trial exposure upon subsequent object recognition. *Neuroscience 2013* Abstracts. Program No. 765.15483. Society for Neuroscience, San Diego, California, USA.

Conference attendance and presentations:

2013

Society for Neuroscience Annual Meeting, San Diego, USA. November 9th – 13th. *Predicting Benefits of Multisensory Memories*.

Tucker-Davis Symposium on Advances and Perspectives in Auditory Neurophysiology, San Diego, USA. November 8th. *Predicting Benefits of Multisensory Memories*.

Lemanic Neuroscience Annual Meeting, Les DIablerets, Switzerland. September 20th – 21st. *Predicting Benefits of Multisensory Memories*.

International Conference on Basic and Clinical multimodal imaging, Geneva, Switzerland. September 5th – 8th. *Predicting Benefits of Multisensory Memories*.

International Multisensory Research Forum, Jerusalem, Israel. June 3rd – 6th. *Predicting Benefits of Multisensory Memories*.

Swiss Society for Neuroscience Annual Meeting, Geneva, Switzerland. February 2nd. *Determinants of the efficacy of single-trial multisensory learning*.

2012

Lemanic Neuroscience Annual Meeting, Les Diablerets, Switzerland. September 7th – 8th. *Determinants of the efficacy of single-trial multisensory learning*.

Neurobiology of Cognition, Gordon Research Conference, Barga (Lucca), Italy. July 8th – 13th. *Heterogeneous auditory-visual integration: effects of pitch, band-width and visual eccentricity.*

International Multisensory Research Forum, Oxford, United Kingdom. June 19th – 22nd. 1) *Electrical neuroimaging of memory discrimination based on single-trial multisensory learning* 2) *Determinants of the efficacy of single-trial multisensory learning* 3) *Heterogeneous auditory-visual integration: effects of pitch, band-width and visual eccentricity.*

Swiss Society for Neuroscience Annual Meeting, Zürich, Switzerland. February 3rd. *Electrical neuroimaging reveals the role of single-trial, episodic multisensory learning in unisensory object discrimination*.

Alpine Brain Imaging Meeting, Champéry, Switzerland. January 8th – 13th. *Electrical neuroimaging reveals the role of single-trial, episodic multisensory learning in unisensory object discrimination.*

2011 Lemanic Neuroscience Annual Meeting, Les Diablerets, Switzerland. September 23rd – 24th. *The impact of single-trial, audio-visual learning on unisensory object discrimination*.

Munich Multisensory Perception Symposium, Holzhausen am Ammersee, Germany, June 24th – 26th. *The impact of single-trial, audio-visual learning on unisensory object discrimination*.

D-Day UNIL, Lausanne, Switzerland. April 28th. *The role of single-trial, episodic multisensory learning in unisensory object discrimination*.

Les Doctoriales de l'UNIL, Lausanne, Switzerland. March 31st. *The role of single-trial, episodic multisensory learning in unisensory object discrimination*.

Swiss Society for Neuroscience Annual Meeting, Basel, Switzerland. March 26th. *The role of single-trial, episodic multisensory learning in unisensory object discrimination*.

Science Communication:

2012 *Les doigts dans le cerveau"* at "La permance du Cerveau", Lausanne, Switzerland. Febrauary - April
Exhibit organized by L'Espace des Inventions and the Pôle de Recherche National – SYNAPSY

13 Articles

The Journal of Neuroscience, January 25, 2012 • 32(4):1171–1182 • **1171**

Behavioral/Systems/Cognitive

Looming Signals Reveal Synergistic Principles of Multisensory Integration

Céline Cappe,[1,2]* **Antonia Thelen,**[1]* **Vincenzo Romei,**[3,4] **Gregor Thut,**[5] **and Micah M. Murray**[1,6,7]

[1]The Functional Electrical Neuroimaging Laboratory, Department of Clinical Neurosciences and Department of Radiology, Centre Hospitalier Universitaire Vaudois and University of Lausanne, 1011 Lausanne, Switzerland, [2]Laboratory of Psychophysics, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland, [3]Wellcome Trust Centre for Neuroimaging at UCL, Institute of Neurology, University College London, London WC1N 3BG, United Kingdom, [4]UCL Institute of Cognitive Neuroscience, University College London, London WC1 3AR, United Kingdom, [5]Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, United Kingdom, [6]Electroencephalography Brain Mapping Core, Center for Biomedical Imaging of Lausanne and Geneva, 1011 Lausanne, Switzerland, and [7]Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, Tennessee 37232

Multisensory interactions are a fundamental feature of brain organization. Principles governing multisensory processing have been established by varying stimulus location, timing and efficacy independently. Determining whether and how such principles operate when stimuli vary dynamically in their perceived distance (as when looming/receding) provides an assay for synergy among the above principles and also means for linking multisensory interactions between rudimentary stimuli with higher-order signals used for communication and motor planning. Human participants indicated movement of looming or receding versus static stimuli that were visual, auditory, or multisensory combinations while 160-channel EEG was recorded. Multivariate EEG analyses and distributed source estimations were performed. Nonlinear interactions between looming signals were observed at early poststimulus latencies (∼75 ms) in analyses of voltage waveforms, global field power, and source estimations. These looming-specific interactions positively correlated with reaction time facilitation, providing direct links between neural and performance metrics of multisensory integration. Statistical analyses of source estimations identified looming-specific interactions within the right claustrum/insula extending inferiorly into the amygdala and also within the bilateral cuneus extending into the inferior and lateral occipital cortices. Multisensory effects common to all conditions, regardless of perceived distance and congruity, followed (∼115 ms) and manifested as faster transition between temporally stable brain networks (vs summed responses to unisensory conditions). We demonstrate the early-latency, synergistic interplay between existing principles of multisensory interactions. Such findings change the manner in which to model multisensory interactions at neural and behavioral/perceptual levels. We also provide neurophysiologic backing for the notion that looming signals receive preferential treatment during perception.

Introduction

Understanding how the brain generates accurate representations of the world requires characterizing the organizing principles governing and neural substrates contributing to multisensory interactions (Calvert et al., 2004; Wallace et al., 2004; Ghazanfar and Schroeder, 2006; Stein and Stanford, 2008). Structurally, monosynaptic projections identified between unisensory (including primary) cortices raise the possibility of interactions during early stimulus processing stages (Falchier et al., 2002, 2010; Rockland and Ojima, 2003; Cappe and Barone, 2005; Cappe et al., 2009a; see also Beer et al., 2011). In agreement, functional data support the occurrence of multisensory interactions within 100 ms poststimulus onset and within low-level cortical areas (Giard and Peronnet, 1999; Molholm et al., 2002; Martuzzi et al., 2007; Romei et al., 2007, 2009; Cappe et al., 2010; Raij et al., 2010; Van der Burg et al., 2011). Nevertheless, the organizing principles governing such multisensory interactions in human cortex and their links to behavior/perception remain largely unresolved.

Based on single-neuron recordings, Stein and Meredith (1993) formulated several "rules" governing multisensory interactions. The principle of inverse effectiveness states that facilitative multisensory interactions are inversely proportional to the effectiveness of the best unisensory response. The temporal rule stipulates that multisensory interactions are dependent on the approximate superposition of neural responses to the constituent unisensory stimuli. The "spatial rule" states that multisensory

Received Nov. 1, 2011; accepted Dec. 1, 2011.

Author contributions: C.C., A.T., V.R., G.T., and M.M.M. designed research; C.C., A.T., and M.M.M. performed research; C.C., A.T., and M.M.M. analyzed data; C.C., A.T., V.R., G.T., and M.M.M. wrote the paper.

This work has been supported by the Swiss National Science Foundation (Grant 3100AO-118419 and 310030B-133136 to M.M.M.) and the Leenaards Foundation (2005 Prize for the Promotion of Scientific Research to M.M.M. and G.T.). The Cartool software (http://sites.google.com/site/fbmlab/cartool) has been programmed by Denis Brunet, from the Functional Brain Mapping Laboratory, Geneva, Switzerland, and is supported by the EEG Brain Mapping Core of the Center for Biomedical Imaging (www.cibm.ch) of Geneva and Lausanne. Prof. Lee Miller provided helpful comments on an earlier version of this manuscript.

*C.C. and A.T. contributed equally to this work.

Correspondence should be addressed to Micah M. Murray, Neuropsychology and Neurorehabilitation Service, Department of Clinical Neurosciences, Centre Hospitalier Universitaire Vaudois and University of Lausanne, rue du Bugnon 46, 1011 Lausanne, Switzerland. E-mail: micah.murray@chuv.ch.

DOI:10.1523/JNEUROSCI.5517-11.2012

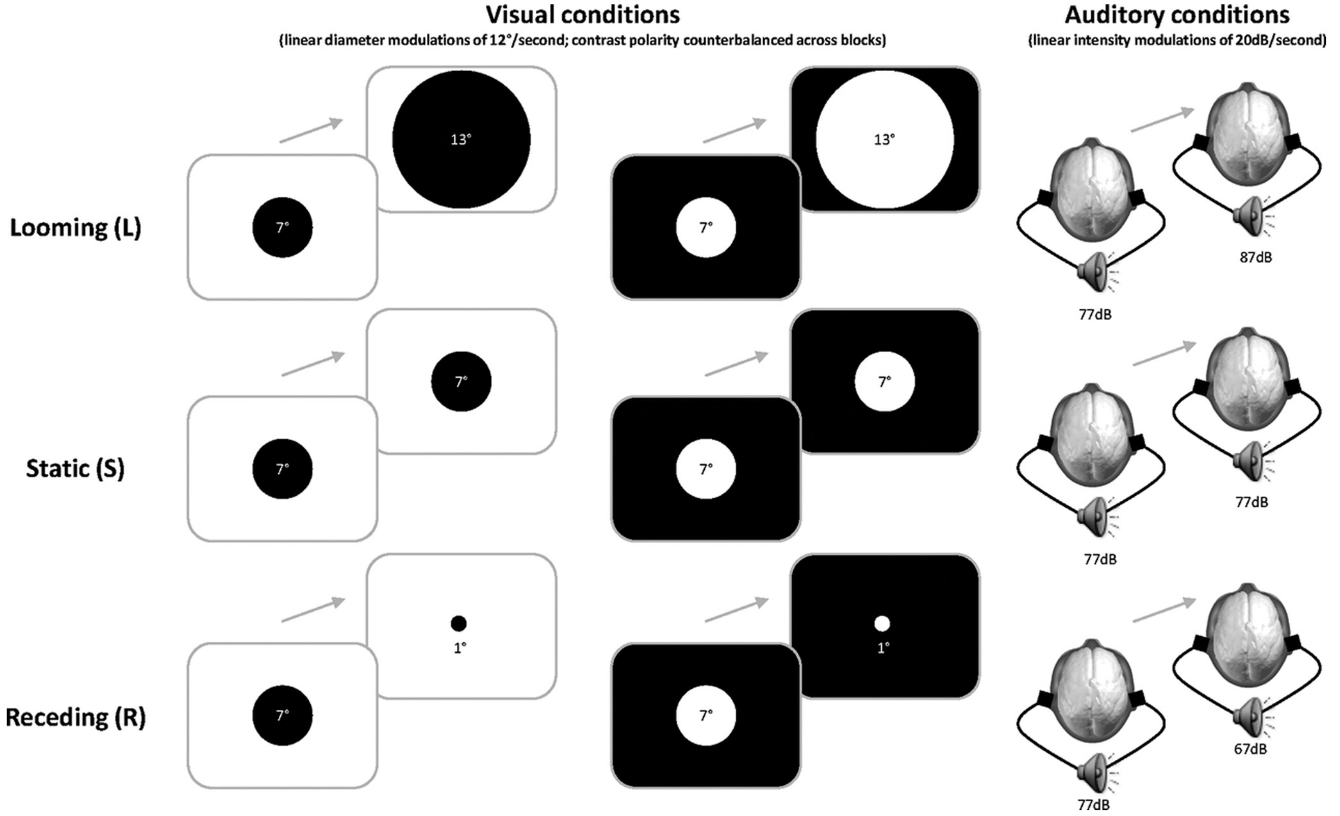


Figure 1. Stimuli and paradigm. Participants performed a go/no-go detection of moving (looming, receding) versus static stimuli that could be auditory, visual, or multisensory auditory-visual. All the stimuli were initially of the same size/intensity to ensure that subjects used dynamic information to perform the task. The perception of movement was induced by linearly changing the size of the centrally displayed disk for the visual condition and by changing the intensity of the complex tone for the auditory condition. To control for differences in stimulus energy in the visual modality, opposite contrast polarities were used across blocks of trials.

interactions are contingent on stimuli being presented to overlapping excitatory zones of the neuron's receptive field. Until now, the spatial rule has faithfully accounted for spatial modulation in azimuth and elevation. But, how spatial information in depth as well as the covariance of information in space, time and effectiveness (i.e., "interactions" between the abovementioned principles) is integrated remains unresolved and was the focus here.

The investigation of looming (approaching) signals is a particularly promising avenue to address synergy between principles of multisensory interactions. Looming signals dynamically increase in their effectiveness and spatial coverage relative to receding stimuli that diminish their effectiveness and spatial coverage. It is also noteworthy that looming cues can indicate both potential threats/collisions and success in acquiring sought-after objects/goals (Schiff et al., 1962; Schiff, 1965; Neuhoff, 1998, 2001; Ghazanfar et al., 2002; Seifritz et al., 2002; Graziano and Cooke, 2006). Recent evidence in non-human primates further suggests that processing of looming signals may benefit from multisensory conditions (Maier et al., 2004, 2008); a suggestion recently confirmed in human performance and consistent with there being synergistic interplay between principles of multisensory interactions (Cappe et al., 2009b). Parallel evidence at the single-neuron level similarly nuances how principles of multisensory interactions cooperate. Responses expressing multisensory interactions within subregions of a neuron's receptive field are heterogeneous and give rise to integrative "hot spots" (Carriere et al., 2008).

In this framework, the present study sought to demonstrate such synergy by identifying the underlying neural mechanisms of multisensory integration for depth cues in humans. We used a multivariate signal analysis approach for EEG termed "electrical neuroimaging" that differentiates modulations in response strength, topography, and latency, as well as localizes effects using a distributed source model (Michel et al., 2004; Murray et al., 2008).

Materials and Methods

Subjects

Fourteen healthy individuals (aged 18–32 years: mean = 25 years; 7 women and 7 men; 13 right-handed) with normal hearing and normal or corrected-to-normal vision participated. Handedness was assessed with the Edinburgh questionnaire (Oldfield, 1971). No subject had a history of neurological or psychiatric illness. All participants provided written informed consent to the procedures that were approved by the Ethics Committee of the Faculty of Biology and Medicine of the University Hospital and University of Lausanne.

Stimuli and procedure

The main experiment involved the go/no-go detection of moving versus static stimuli that could be auditory, visual, or multisensory auditory-visual (A, V, and AV, respectively), as described in our recent paper describing the behavioral part of this study (Cappe et al., 2009b). To induce the perception of movement, visual stimuli changed in size and auditory stimuli changed in volume so as to give the impression of either looming or receding (denoted by L and R, respectively). Static stimuli were of constant size/volume. The stimulus conditions are schematized in Figure 1. Specific multisensory conditions were generated using the full range of combinations of movement type (L, R, and S) and congruence between the senses. For convenience we use shorthand to describe experimental conditions such that, for example, ALVL refers to the multisensory combination of auditory looming and visual looming and ARVL refers to the multisensory combination of auditory receding and visual looming. There were 15 configurations of stimuli in total (6 unisensory and 9 multisensory). Go trials (i.e., those on which either or both sensory modalities contained moving stimuli) occurred on 80% of the trials. Each of the 15 conditions was repeated 252 times across 18 blocks of randomly intermixed trials. Additional details appear in Cappe et al., 2009b.

Auditory stimuli, 10 dB rising-intensity (looming signal) and falling-intensity (receding signal) 1000 Hz complex tones composed of a triangular waveform, were generated with Adobe Audition software (Adobe Systems Inc.). Prior research has shown that tonal stimuli produce more reliable perceptions of looming and receding (Neuhoff, 1998) and may also be preferentially involved in multisensory integration (Maier et al., 2004; Romei et al., 2009). Auditory stimuli were presented over insert earphones (Etymotic model ER4S). They were sampled at 44.1 kHz, had 10 ms onset and offset ramps (to avoid clicks). The visual stimulus consisted of a centrally presented disc (either black on a white background or white on a black background, counterbalanced across blocks of trials to avoid difference of contrast and size between these dynamic stimuli) that symmetrically expanded (from 7° to 13° diameter with the radius increasing linearly at a constant rate) in the case of looming or contracted (from 7° to 1° diameter) in the case of receding. Additionally, the stimuli were 500 ms in duration and the interstimulus interval varied from 800 to 1400 ms such that participants could not anticipate the timing of stimulus presentation. Stimulus delivery and response recording were controlled by E-Prime (Psychology Software Tools; www.pstnet.com).

EEG acquisition and analyses

Continuous EEG was acquired at 1024 Hz through a 160-channel Biosemi ActiveTwo AD-box (www.biosemi.com) referenced to the common mode sense (CMS; active electrode) and grounded to the driven right leg (DRL; passive electrode), which functions as a feedback loop driving the average potential across the electrode montage to the amplifier zero (full details, including a diagram of this circuitry, can be found at http://www.biosemi.com/faq/cms&drl.htm). Epochs of EEG from 100 ms prestimulus to 500 ms poststimulus onset were averaged for each of the four stimulus conditions and from each subject to calculate the event-related potential (ERP). Only trials leading to correct responses were included. In addition to the application of an automated artifact criterion of ±80 μV, the data were visually inspected to reject epochs with blinks, eye movements, or other sources of transient noise. Baseline was defined as the 100 ms prestimulus period. For each subject's ERPs, data at artifact electrodes were interpolated (Perrin et al., 1987). Data were baseline corrected using the prestimulus period, bandpass filtered (0.18–60.0 Hz), and recalculated against the average reference.

General analysis strategy. Multisensory effects and effects of spatial congruence were identified with a multistep analysis procedure, which we refer to as electrical neuroimaging and which is implemented in the freeware Cartool (Brunet et al., 2011; http://sites.google.com/site/fbmlab/cartool). Analyses were applied that use both local and global measures of the electric field at the scalp. These so-called electrical neuroimaging analyses allowed us to differentiate effects following from modulations in the strength of responses of statistically indistinguishable brain generators from alterations in the configuration of these generators (viz. the topography of the electric field at the scalp), as well as latency shifts in brain processes across experimental conditions (Michel et al., 2004, 2009; Murray et al., 2005, 2008). In addition, we used the local autoregressive average distributed linear inverse solution (LAURA; Grave de Peralta Menendez et al., 2001, 2004) to visualize and statistically contrast the likely underlying sources of effects identified in the preceding analysis steps.

ERP waveform modulations. As a first level of analysis, we analyzed waveform data from all electrodes as a function of time poststimulus onset in a series of pairwise comparisons (*t* tests) between responses to the multisensory pair and summed constituent unisensory responses. Temporal auto-correlation at individual electrodes was corrected through the application of an 11 contiguous data-point temporal criterion (~11 ms) for the persistence of differential effects (Guthrie and Buchwald, 1991). Similarly, spatial correlation was addressed by considering as reliable only those effects that entailed at least 11 electrodes from the 160-channel montage. Nonetheless, we would emphasize that the number of electrodes exhibiting an effect at a given latency will depend on the reference, and this number is not constant across choices of reference because significant effects are not simply redistributed across the montage (discussed in Tzovara et al., in press). Likewise, the use of an average reference receives support from biophysical laws as well as the implicit recentering of ERP data to such when performing source estimations (discussed by Brunet et al., 2011). Analyses of ERP voltage waveform data (vs the average reference) are presented here to provide a clearer link between canonical ERP analysis approaches and electrical neuroimaging. The results of this ERP waveform analysis are presented as an area plot representing the number of electrodes exhibiting a significant effect as a function of time (poststimulus onset). This type of display was chosen to provide a sense of the dynamics of a statistical effect between conditions as well as the relative timing of effects across contrasts. We emphasize that while these analyses give a visual impression of specific effects within the dataset, our conclusions are principally based on reference-independent global measures of the electric field at the scalp that are described below.

Global electric field analyses. The collective poststimulus group-average ERPs were subjected to a topographic cluster analysis based on a hierarchical clustering algorithm (Murray et al., 2008). This clustering identifies stable electric field topographies (hereafter template maps). The ERP topography is independent of the reference, and modulations in topography forcibly reflect modulations in the configuration of underlying generators (Lehmann, 1987). Additionally, the clustering is exclusively sensitive to topographic modulations, because the data are first normalized by their instantaneous Global Field Power (GFP). The optimal number of temporally stable ERP clusters (i.e., the minimal number of maps that accounts for the greatest variance of the dataset) was determined using a modified Krzanowski-Lai criterion (Murray et al., 2008). The clustering makes no assumption on the orthogonality of the derived template maps (Pourtois et al., 2008; De Lucia et al., 2010). Template maps identified in the group-average ERP were then submitted to a fitting procedure wherein each time point of each single-subject ERP is labeled according to the template map with which it best correlated spatially (Murray et al., 2008) so as to statistically test the presence of each map in the moment-by-moment scalp topography of the ERP and the differences in such across conditions. Additionally, temporal information about the presence of a given template map was derived, quantifying (among other things) when a given template map was last labeled in the single-subject ERPs. These values were submitted to repeated-measures ANOVA. In addition to testing for modulations in the electric field topography across conditions, this analysis also provides a more objective means of defining ERP components. That is, we here defined an ERP component as a time period of stable electric field topography.

Modulations in the strength of the electric field at the scalp were assessed using GFP (Lehmann and Skrandies, 1980; Koenig and Melie-García, 2010) for each subject and stimulus condition. GFP is calculated as the square root of the mean of the squared value recorded at each electrode (vs the average reference) and represents the spatial SD of the electric field at the scalp. It yields larger values for stronger electric fields. Because GFP is calculated across the entire electrode montage, comparisons across conditions will be identical, regardless of the reference used (though we would note that the above formula uses an average reference). In this way, GFP constitutes a reference-independent measure. GFP modulations were analyzed via ANOVAs over the periods of interest defined by the above topographic cluster analysis (i.e., 73–113 ms and 114–145 ms).

Source estimations. We estimated the localization of the electrical activity in the brain using a distributed linear inverse solution applying the LAURA regularization approach comprising biophysical laws as constraints (Grave de Peralta Menendez et al., 2001, 2004; for review, see also Michel et al., 2004; Murray et al., 2008). LAURA selects the source configuration that better mimics the biophysical behavior of electric vector fields (i.e., activity at one point depends on the activity at neighboring points according to electromagnetic laws). In our study, homogenous

Table 1. Psychophysics results

Condition	Mean reaction times (ms) ± SEM	Mean movement rating (1–5 scale) ± SEM
ALVL	439 ± 19	4.20 ± 0.15
ARVR	457 ± 20	3.46 ± 0.17
ALVR	447 ± 20	4.08 ± 0.15
ARVL	456 ± 21	3.28 ± 0.17

regression coefficients in all directions and within the whole solution space were used. LAURA uses a realistic head model, and the solution space included 4024 nodes, selected from a 6 × 6 × 6 mm grid equally distributed within the gray matter of the Montreal Neurological Institute's average brain (courtesy of R. Grave de Peralta Menendez and S. Gonzalez Andino; http://www.electrical-neuroimaging.ch/). Prior basic and clinical research from members of our group and others has documented and discussed in detail the spatial accuracy of the inverse solution model used here (Grave de Peralta Menendez et al., 2004; Michel et al., 2004; Gonzalez Andino et al., 2005; Martuzzi et al., 2009). In general, the localization accuracy is considered to be along the lines of the matrix grid size (here 6 mm). The results of the above topographic pattern analysis defined time periods for which intracranial sources were estimated and statistically compared between conditions (here 73–113 ms poststimulus). Before calculation of the inverse solution, the ERP data were downsampled and affine-transformed to a common 111-channel montage. Statistical analyses of source estimations were performed by first averaging the ERP data across time to generate a single data point for each participant and condition. This procedure increases the signal-to-noise ratio of the data from each participant. The inverse solution was then estimated for each of the 4024 nodes. These data were then submitted to a three-way ANOVA using within-subject factors of pair/sum condition, stimulus congruence/incongruence, and visual looming/receding. A spatial extent criterion of at least 17 contiguous significant nodes was likewise applied (see also Toepel et al., 2009; Cappe et al., 2010; De Lucia et al., 2010; Knebel et al., 2011; Knebel and Murray, 2012 for a similar spatial criterion). This spatial criterion was determined using the AlphaSim program (available at http://afni.nimh.nih.gov) and assuming a spatial smoothing of 6 mm FWHM. This criterion indicates that there is a 3.54% probability of a cluster of at least 17 contiguous nodes, which gives an equivalent node-level *p*-value of $p \leq 0.0002$. The results of the source estimations were rendered on the Montreal Neurologic Institute's average brain with the Talairach and Tournoux (1988) coordinates of the largest statistical differences within a cluster indicated.

Results

The behavioral results with this paradigm (Fig. 1) have been published separately (Cappe et al., 2009b). Our main findings were a selective facilitation for multisensory looming stimuli (auditory-visual looming denoted ALVL). When asking participants to detect stimulus movement, facilitation of behavior was seen for all multisensory conditions compared with unisensory conditions. Interestingly, human subjects were faster to detect movement of multisensory looming stimuli versus receding (auditory-visual receding denoted ARVR) or incongruent stimuli (auditory looming with visual receding and auditory receding with visual looming denoted ALVR and ARVL, respectively). For a movement rating task with the same stimuli, this selective facilitation for looming stimuli was shown again in higher movement ratings for looming stimuli than for receding stimuli, and even more in multisensory conditions (Cappe et al., 2009b). Only multisensory looming stimuli resulted in enhancement beyond that induced by the sheer presence of auditory-visual stimuli, as revealed by contrasts with multisensory conditions where one sensory modality consisted of static (i.e., constant size/volume) information (cf. Cappe et al., 2009b, their Fig. 5). These behavioral results are recapitulated here in Table 1. During the detection task, we recorded ERPs for each subject and we analyzed these data as described below and in the Materials and Methods.

ERP waveform analyses

Our analyses here are based on the application of an additive model to detect nonlinear neural responses interactions, wherein the ERP in response to the multisensory condition is contrasted with the summed ERPs in response to the constituent auditory and visual conditions (hereafter referred to as "pair" and "sum" ERPs, respectively). The first level of analysis focused on determining the timing differences between the multisensory pair and the sum of unisensory ERPs. Visual inspection of an exemplar occipital electrode suggests there to be nonlinear interactions beginning earlier for looming conditions (ALVL) than for receding (ARVR) or incongruent conditions (ALVR and ARVL) (Fig. 2*a*). The group-averaged ERPs from the pair and sum responses were compared statistically by paired *t* tests. These analyses were applied for each condition (ALVL, ARVR, ALVR, ARVL; for the analyses of static conditions, see Cappe et al., 2010). Statistical analyses of the pair versus sum ERP waveforms as a function of time are displayed in Figure 2*b* and show significant and temporally sustained nonlinear neural response interactions for each condition, but with different latencies (statistical criteria are defined as $p < 0.05$ for a minimum of 11 ms duration at a given electrode and a spatial criterion of at least 11 electrodes). Using these criteria, the earliest nonlinear response interactions began at 68 ms poststimulus onset for the multisensory looming condition (similar results were also found for static conditions, see Cappe et al., 2010), whereas such effects were delayed until 119 ms for the multisensory receding condition. For incongruent conditions, these differences were observed at 95 ms poststimulus for ALVR and at 140 ms poststimulus for ARVL.

Global electric field analyses

A hierarchical topographic cluster analysis was performed on the group-average ERPs concatenated across the 8 experimental conditions (pair/sum × ALVL, ARVR, ALVR, ARVL) to identify periods of stable electric field topography both within and between experimental conditions. For this concatenated dataset, 9 template maps were identified with a global explained variance of 95.1%. Two different maps were identified in the group-average data over the 73–145 ms poststimulus period that appeared to differently account for pair and sum conditions (Fig. 3*a*). The first map appeared earlier for multisensory pair than for summed unisensory responses for all conditions (i.e., all combinations of looming and receding auditory and visual stimuli). The amount of time each template map yielded a higher spatial correlation with the single-subject data from each condition was quantified over the 73–145 ms poststimulus period as "the frequency of map presence" and submitted to a repeated-measures ANOVA using within-subject factors of pair/sum condition, stimulus congruence/incongruence, visual looming/receding, and map (Fig. 3*b*). In accordance with a faster transition from one map to another under multisensory conditions, there was a significant interaction between pair/sum condition and template map ($F_{(1,13)} = 11.957$; $p = 0.004$; $\eta_p^2 = 0.479$). By extension, such topographic differences argue for a latency shift in the configuration of the underlying intracranial sources. This latency shift was further supported by an analysis of the timing at which the first of the two template maps was last observed (i.e., yielded a higher spatial correlation than the other template map) in the single-subject data (Fig. 3*c*). The same factors as above were used, save for that of template map. Consistent with the above, there was a signifi-

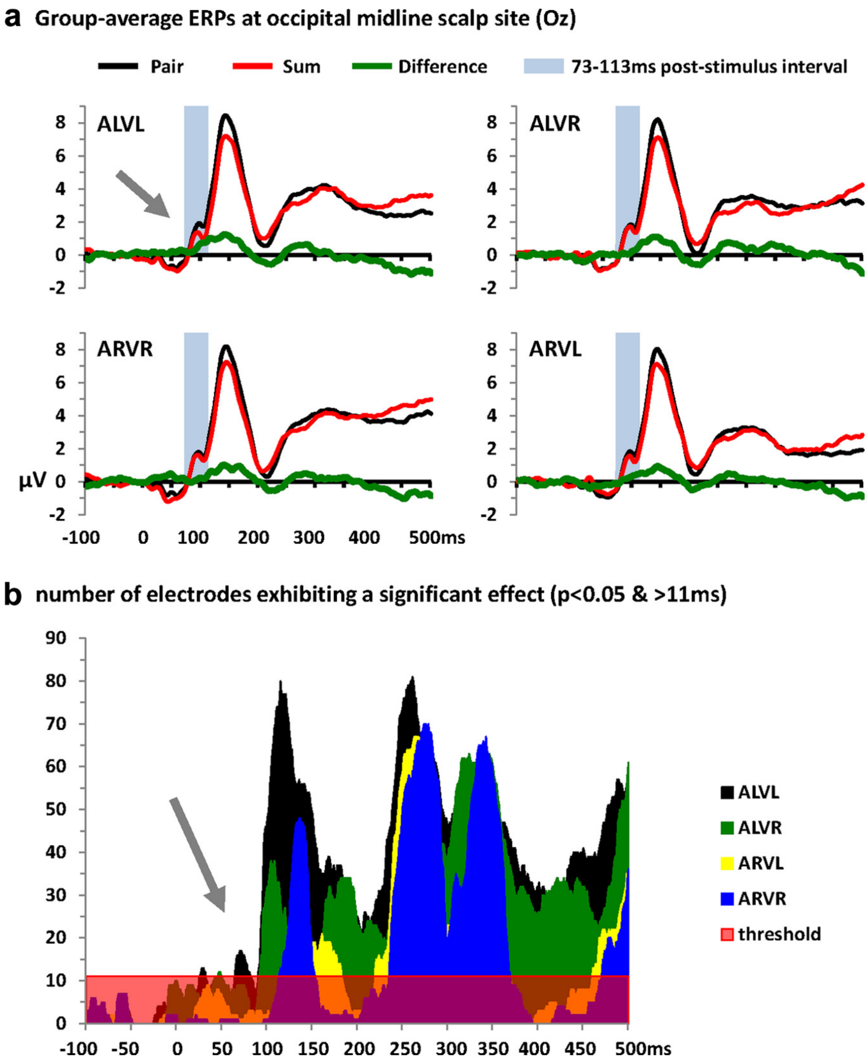


Figure 2. Group-averaged ($N = 14$) voltage waveforms and ERP voltage waveform analyses. ***a***, Data are displayed at a midline occipital electrode site (Oz) from the response to the multisensory pair (black traces), summed unisensory responses (red traces), and their difference (green traces). The arrow indicates modulations evident for multisensory looming conditions that were not apparent for any other multisensory combination over the ~70–115 ms poststimulus interval. ***b***, The area plots show results of applying millisecond-by-millisecond paired contrasts (*t* tests) across the 160 scalp electrodes comparing multisensory and the sum of unisensory stimuli. The number of electrodes showing a significant difference are plotted as a function of time (statistical criteria: $p < 0.05$ for a minimum of 11 consecutive milliseconds and 11 scalp sites). Nonlinear neural response interactions started at 68 ms poststimulus onset for multisensory looming stimuli (ALVL), at 119 ms for the multisensory receding (ARVR) condition, and at 95 and 140 ms for incongruent multisensory conditions ALVR and ARVL, respectively.

cant main effect of pair/sum condition ($F_{(1,9)} = 5.154$; $p = 0.049$; $\eta_p^2 = 0.364$; note the lower degrees of freedom in this specific analysis because not all maps were observed in all subjects, leading to missing values rather than entries of 0). Specifically, in the case of responses to multisensory stimuli, the transition occurred at 113 ms poststimulus on average; a latency that is used below to define time windows of interest for analyses of Global Field Power and source estimations.

Distinct topographies were also identified at the group-average level across conditions (pair versus sum) over the 250–400 ms poststimulus period (Fig. 3*a*). The topographic analysis indicated that predominated maps differed between the multisensory pair (one map) and unisensory sum conditions (two maps) over this period. However, the differences after 250 ms could also be due to the use of the additive model to determine nonlinear interactions (summation of motor activity; discussed by Cappe et al., 2010). We therefore will not focus on this observation (see also Besle et al., 2004; Murray et al., 2005).

For each pair and sum condition as well as each subject, the mean Global Field Power was calculated over the 73–113 ms and 114–145 ms periods and in turn submitted to a three-way ANOVA, using within-subjects factors of pair/sum condition, stimulus congruence/incongruence, and visual looming/receding. Over the 73–113 ms period (i.e., during the first period of stable topography), there were superadditive interactions for the multisensory looming condition that were not observed for other conditions (Fig. 4*a*; three-way interaction $F_{(1,13)} = 4.862$; $p = 0.046$; $\eta_p^2 = 0.272$ and *post hoc t* test for the ALVL condition $p = 0.02$). Over the 114–145 ms period (i.e., during the second period of stable topography), there were superadditive interactions for all multisensory conditions (Fig. 4*b*; main effect of pair vs sum $F_{(1,13)} = 4.913$; $p = 0.045$; $\eta_p^2 = 0.274$). The early preferential nonlinear interactions observed for multisensory looming conditions here is consistent with observations based on voltage ERP waveforms (Fig. 2).

Excluding accounts based on differences in stimulus energy

We deliberately used an experimental design wherein all stimulus conditions had the same initial volume and/or size (Fig. 1) so that subjects could not perform the task based on initial differences across conditions but instead needed to evaluate the stimuli dynamically. However, a reasonable criticism is that the conditions consequently differ in their total stimulus energy. With regard to the visual modality, we controlled for such differences by counterbalancing the contrast polarity across blocks of trials such that the total number of black and white pixels was equivalent across conditions. With regard to the auditory modality, no such control was implemented. Thus, there is a potential confound between perceived direction and stimulus intensity.

However, it is important to recall that all of the pair versus sum comparisons are fully equated in terms of stimulus energy. Likewise, a posteriori our results provide one level of argumentation against this possibility. Neither the main effect of pair/sum condition (or its interaction with map) in the topographic cluster analysis nor the three-way interaction observed in the global field power analysis can be explained by simple differences in acoustic intensity. Moreover, strict application of the principle of inverse effectiveness would predict that receding stimuli would yield greater interactions than looming stimuli. Yet, there was no evidence of such in our analyses. Rather, only the multisensory looming condition resulted in early-stage global field power (and voltage waveform) modulations, and subsequent effects were

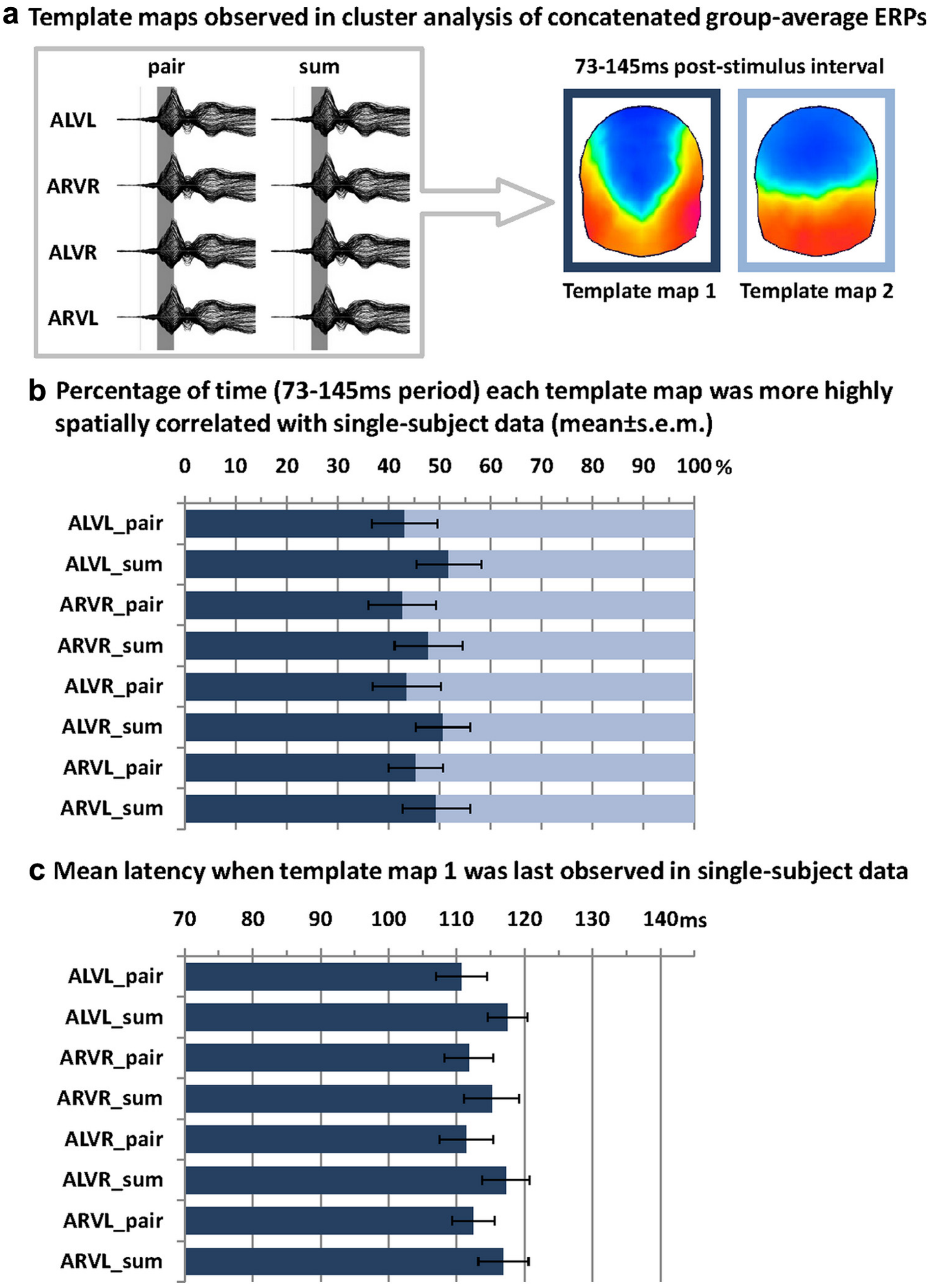


Figure 3. Topographic cluster analyses and single-subject fitting based on spatial correlation. ***a***, The hierarchical clustering analysis was applied to the concatenated group-averaged ERPs from all pair and sum conditions (schematized by the gray box) and identified two template maps accounting for responses over the 73–145 ms poststimulus period that are shown on the right of this panel. ***b***, The spatial correlation between each template map (Template maps 1 and 2) was calculated with the single-subject data from each condition, and the percentage of time a given template map yielded a higher spatial correlation was quantified (mean ± SEM shown) and submitted to ANOVA that revealed a significant interaction between pair versus sum conditions and template map. ***c***, The latency when template map 1 was last observed (measured via spatial correlation) in the single-subject data from each condition was quantified and submitted to ANOVA. There was an earlier transition from template map 1 to template map 2 under multisensory versus summed unisensory conditions.

common to all multisensory conditions regardless of looming/receding (and therefore stimulus intensity confounds).

To more directly address this potential confound, we contrasted ERPs in response to unisensory conditions, using the same analysis methods as described above for examining multisensory interactions. With regard to responses to looming and receding sounds, voltage waveform analyses revealed effects beginning at 220 ms poststimulus onset. A millisecond-by-millisecond analysis of the GFP waveforms revealed effects beginning at 234 ms poststimulus onset. Finally, a millisecond-by-millisecond analysis of the ERP topography (normalized by its instantaneous GFP) revealed effects beginning at 264 ms poststimulus onset. These analyses across local and global measures of the electric field all indicate that response differences between unisensory looming and receding stimuli are substantially delayed relative to the latency of the earliest nonlinear neural response interactions observed for all multisensory conditions as well as the preferential interactions between multisensory looming stimuli. This finding provides additional support to the proposition that discrimination/differentiation of motion signals is facilitated by multisensory interactions and extends this no-

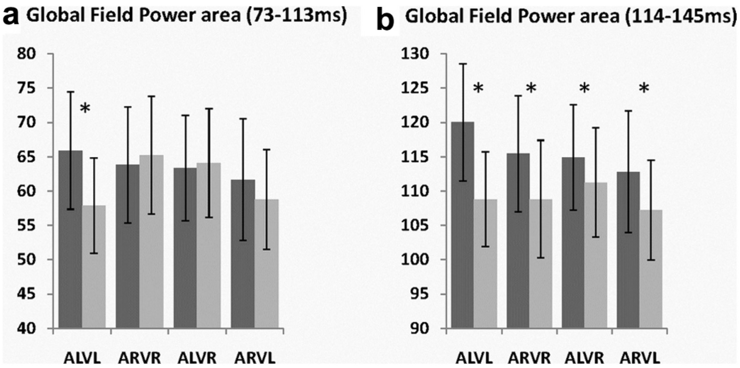


Figure 4. Global field power analyses. ***a***, ***b***, Modulations in response strength were identified using global field power (GFP), which was quantified over the 73–113 ms poststimulus period (***a***) and 114–145 ms poststimulus period (***b***) for each multisensory condition and the sum of unisensory conditions (dark and light gray bars, respectively). Mean ± SEM values are displayed, and asterisks indicate significant effects between specific pair and sum conditions. There was a significant three-way interaction over the 73–113 ms period, with evidence of selective nonlinear modulations for multisensory looming conditions. There was a significant main effect of pair versus sum conditions over the 114–145 ms period, indicative of generally stronger responses to multisensory versus summed unisensory conditions.

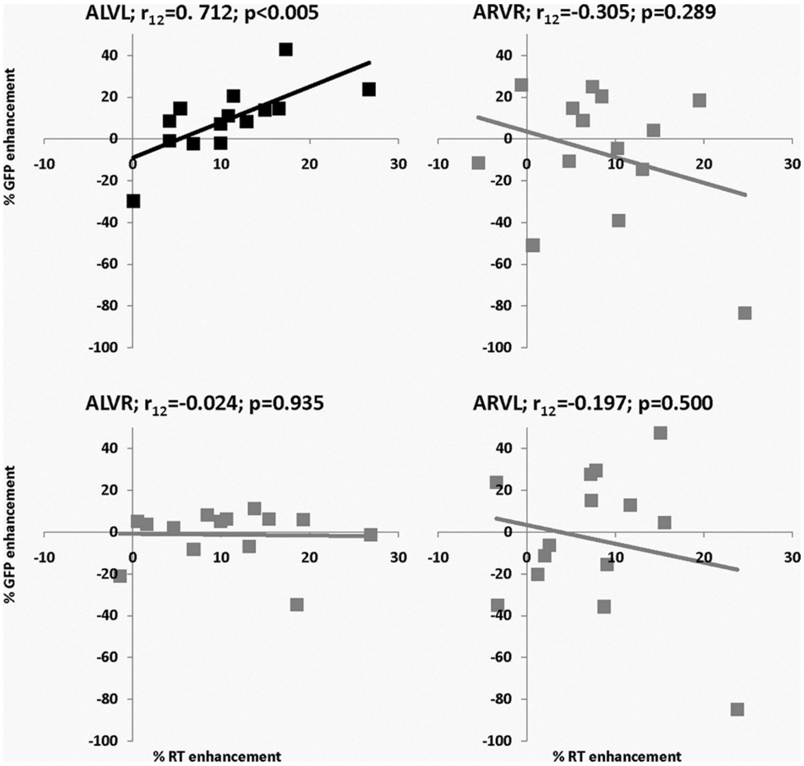


Figure 5. Relationship between RT and GFP multisensory enhancements. These scatter plots relate the percentage of RT enhancement to the percentage of GFP enhancement over the 73–113 ms period (*x*-axis and *y*-axis, respectively) for each of the multisensory conditions. The multisensory enhancement index is defined as the difference between the multisensory condition and the best unisensory condition divided by the best unisensory condition for each participant. A significant, positive, and linear correlation was exhibited only for the multisensory looming condition (ALVL).

tion to motion across perceived distances. Future work varying the acoustic structure of the stimuli will be able to capitalize on evidence that the impression of looming is limited to harmonic or tonal sounds (Neuhoff, 1998) as are multisensory effects involving looming sounds (Maier et al., 2004; Romei et al., 2009; Leo et al., 2011).

Relation between behavioral facilitation and GFP enhancement

In a further analysis, we determined an index of multisensory enhancements for reaction times (RTs) and GFP area over the 73–113 ms period. The percentage of multisensory RT enhancement was calculated as the ratio of the difference between the multisensory condition and the best constituent unisensory condition relative to the best unisensory condition for each participant (see also Stein and Meredith, 1993; Cappe et al., 2009b). The percentage of multisensory GFP enhancement was calculated as the ratio of the difference between the GFP to the multisensory pair and summed unisensory conditions relative to the multisensory pair. Interestingly, a positive correlation was exhibited between RTs and GFP multisensory enhancements for looming conditions over the 73–113 ms period ($r_{(12)} = 0.712$; $p < 0.005$; Fig. 5). No other condition showed a reliable correlation (all p-values >0.05; Fig. 5). These results suggest that early integrative effects are behaviorally relevant (particularly in the case of looming signals that convey strong ethological significance) and that greater integrative effects result in greater behavioral facilitation. Similar correlations have recently been reported by Van der Burg et al. (2011), where participants with greater early-latency interactions showed bigger benefits of task-irrelevant sounds in the context of a visual feature detection task. Such findings thus add to a growing literature demonstrating the direct behavioral relevance of early-latency and low-level multisensory interactions (Sperdin et al., 2010).

Source estimations

Given the results of the above voltage waveform and GFP analyses, we estimated sources over the 73–113 ms poststimulus period. Scalar values from the source estimations throughout the entire brain volume from each participant and condition were submitted to a three-way ANOVA (spatial criterion described in Materials and Methods). There was evidence for a three-way interaction between pair/sum condition, stimulus congruence/incongruence, and visual looming/receding within the right claustrum/insula extending into the anterior inferior temporal lobe and amygdala as well as within the bilateral cuneus extending within the right hemisphere inferiorly into the lingual gyrus and posteriorly into the lateral middle occipital gyrus (Fig. 6*a*; Table 2). To ascertain the basis for this interaction, group-average scalar values at the node exhibiting maximal *F*-values within each of these three clusters are shown as bar graphs in Figure 6*b*. In all three clusters subadditive effects were seen for multisensory looming conditions. The other conditions failed to exhibit significant nonlinear effects, with the exception of the ALVR condition (pair vs sum contrast) that exhibited significant subadditive effects in the cu-

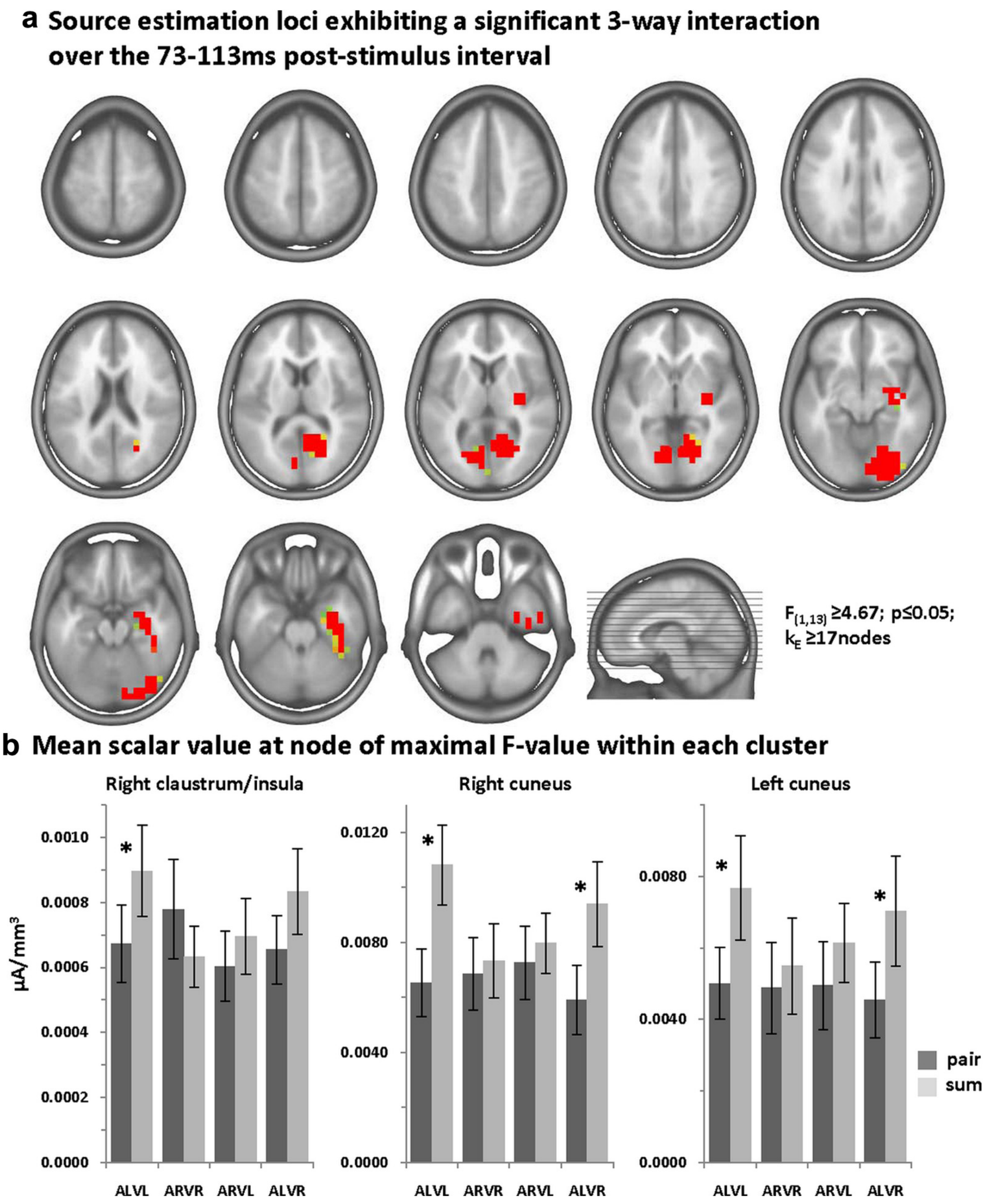


Figure 6. Statistical analyses of source estimations: three-way interaction. Group-averaged source estimations were calculated over the 73–113 ms poststimulus period for each experimental condition and submitted to a three-way ANOVA. Regions exhibiting significant interactions between pair/sum conditions, congruent/incongruent multisensory pairs, and visual looming versus receding stimuli are shown in ***a*** on axial slices of the MNI template brain. Only nodes meeting the $p \leq 0.05$ α criterion as well as a spatial extent criterion of at least 17 contiguous nodes were considered reliable (see Materials and Methods for details). Three clusters exhibited an interaction, and the mean scalar values (SEM indicated) from the node exhibiting the maximal *F*-value in each cluster are shown in ***b***. Asterisks indicate significant differences between pair and sum conditions.

Table 2. Source estimation clusters exhibiting a three-way interaction over the 73–113 ms interval

Brain area	Talairach and Tournoux (1988) coordinates of maximal *F*-value	Maximal *F*-value	Cluster size (number of nodes)
Right claustrum/insula extending inferiorly into the anterior temporal cortex and amygdala	35, −17, 2 mm	12.071	78
Right cuneus extending inferiorly into the lingual gyrus and posteriorly to lateral occipital cortex	23, −70, −4 mm	15.484	104
Left cuneus	−17, −75, 11 mm	9.157	23

neus bilaterally. Of note, however, is that nonlinear effects within the claustrum/insula were limited to multisensory looming conditions, suggesting these regions are particularly sensitive to and/or themselves integrating information regarding perceived motion direction and congruence across modalities (see also Bushara et al., 2001; Calvert et al., 2001; Naghavi et al., 2007; but see Remedios et al., 2010). While determining the precise manner of relating the directionality of changes in GFP to the directionality of effects observed within source estimation nodes awaits further investigation, it is important to note that both levels of analysis indicate there to be effects specific to multisensory looming conditions.

Aside from this three-way interaction, several other main effects and interactions were observed (Fig. 7). There was evidence

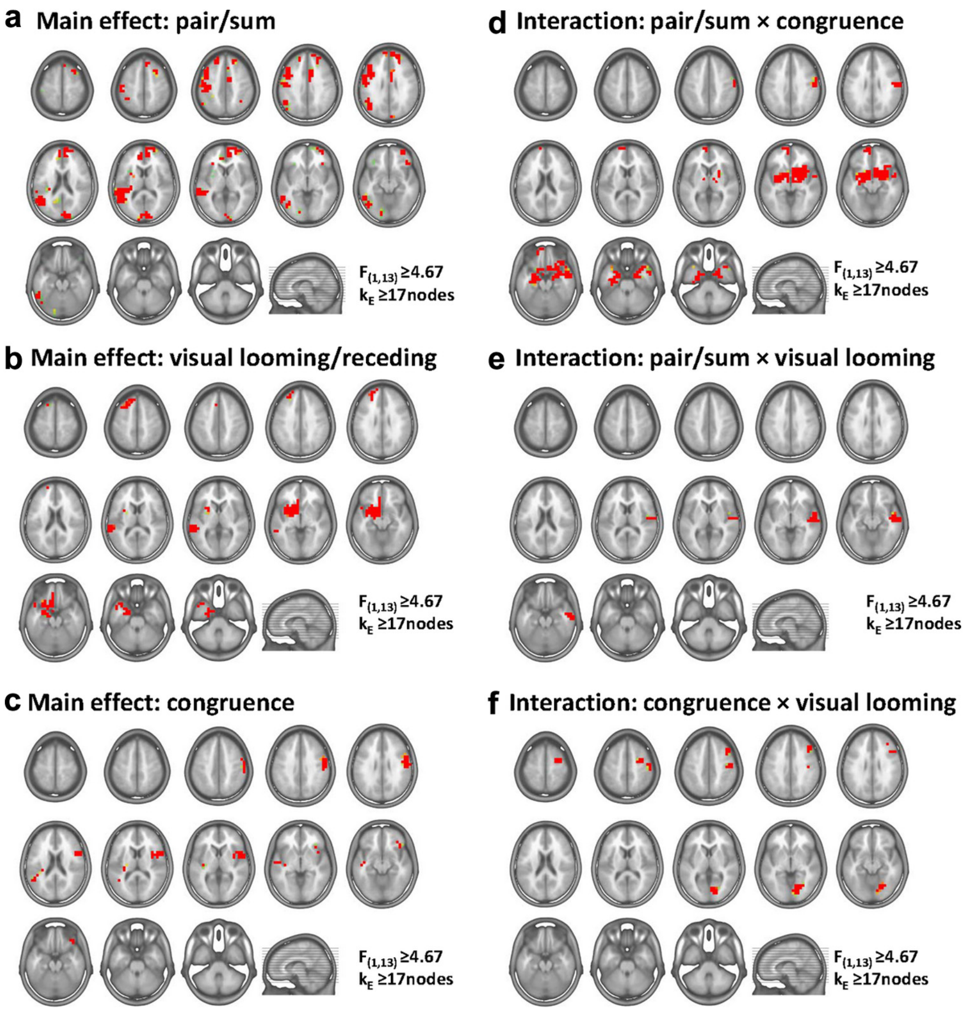


Figure 7. Statistical analyses of source estimations: main effects and two-way interactions. Group-averaged source estimations were calculated over the 73–113 ms poststimulus period for each experimental condition and submitted to a three-way ANOVA. Only nodes meeting the $p \leq 0.05$ α criterion as well as a spatial extent criterion of at least 17 contiguous nodes were considered reliable (see Materials and Methods for details). Regions exhibiting significant main effects are shown in ***a–c*** on axial slices of the MNI template brain. Regions exhibiting significant two-way interactions are shown in ***d–f*** on axial slices of the MNI template brain.

for a main effect of pair versus sum conditions within a widespread network of regions that included the bilateral cuneus extending along the calcarine sulcus, the left superior temporal gyrus extending superiorly into the angular gyrus, the left inferior frontal gyrus, and bilateral medial frontal gyrus (Fig. 7*a*). This main effect was the consequence of subadditive interactions, consistent with prior EEG and fMRI findings (Bizley et al., 2007; Martuzzi et al., 2007; Besle et al., 2008; Kayser et al., 2009; Cappe et al., 2010; Raij et al., 2010). The main effect of visual looming versus receding produced differences within the left superior temporal gyrus as well as left claustrum/insula extending inferiorly into the amygdala (Fig. 7*b*). Responses were stronger to visual looming versus receding stimuli, despite our counterbalancing contrast polarity (and therefore mean stimulus energy) across blocks of trials. This suggests that this main effect is driven by an analysis of the perceived direction of motion. Finally, there was evidence for a main effect of congruent versus incongruent multisensory combinations within the right inferior frontal gyrus such that responses were stronger to congruent combinations (Fig. 7*c*), consistent with studies implicating these regions in processing multisensory object congruence/familiarity (Doehrmann and Naumer, 2008). There was likewise evidence for a significant interaction between pair and sum conditions and congruent versus incongruent multisensory combinations within bilateral limbic and subcortical structures, including the amygdala and putamen (Fig. 7*d*). There was evidence for a significant interaction between pair and sum conditions and visual looming versus receding stimuli within the right superior and middle temporal gyri (Fig. 7*e*). Finally, there was evidence for a significant interaction between congruent versus incongruent multisensory combinations and visual looming versus receding stimuli within the right cuneus and right inferior frontal gyrus (Fig. 7*f*).

Discussion

This study provides the first demonstration that the human brain preferentially integrates multisensory looming signals. Such findings complement observations in non-human primates of preferential looking behavior with multisensory looming stimuli (Maier et al., 2004) as well as enhanced neural synchrony between auditory and superior temporal cortices (Maier et al., 2008). The present observation of selective superadditive interactions of responses to multisensory looming signals during early poststimulus onset periods (73–113 ms) that were moreover positively correlated with behavioral facilitation argues for synergistic interplay in humans between principles of multisensory integration established from single-neuron recordings in animals (Stein and Meredith, 1993). These results, in conjunction with the extant literature, highlight the challenge of directly transposing models of multisensory interactions from single-neuron to population-level responses and perception (Krueger et al., 2009; Ohshiro et al., 2011). They likewise suggest that multisensory interactions can facilitate the processing and perception of specific varieties of ethologically significant environmental stimuli; here those signaling potential collisions/dangers.

Synergistic interplay between principles of multisensory integration

Looming and receding stimuli provide an effective means for investigating the interplay between established principles of multisensory interactions, because the perceived distance and motion direction are higher-order indices that follow from first-order changes in the visual size or auditory intensity of the stimuli. In both senses a dynamic change in size/intensity (and by extension effectiveness) is interpreted at least perceptually and presumably coded neurophysiologically as a source varying in its distance from the observer. Likewise, because stimulus intensity at trial onset was equated across all conditions, participants necessarily treated stimulus dynamics. Also, no differential processing of looming versus receding signals was required by the task, but rather the differentiation of moving versus stationary stimuli. This allowed for the same task-response requirements for congruent and incongruent multisensory conditions. As such, the present differences can be considered implicit.

It is likewise indispensable to consider the suitability of transposing the spatial and inverse effectiveness principles to studies of multisensory interactions in humans (the temporal principle is not at play here as the stimuli were always synchronously covarying). Direct transposition of the spatial principle, particularly within the auditory modality, is challenged by evidence for population-based coding of sounds' positions rather than a simple spatio-topic mapping (Stecker and Middlebrooks, 2003; Murray and Spierer, 2009). Instead, recent single-unit recordings within auditory fields along the superior temporal plane in macaque monkeys indicate that these neurons are responsive to the full 360° of azimuth (Woods et al., 2006). With regard to rising versus falling intensity sound processing, there is evidence for the involvement of core auditory fields as well as for a general neural response bias (in terms of spiking rate, but not latency) for rising-intensity sounds regardless of their specific frequency or volume (Lu et al., 2001). Such findings suggest that unisensory looming stimuli may receive preferential processing and may in turn be one basis for the enhanced salience of looming stimuli (Kayser et al., 2005) that in turn cascades to result in selective integration of multisensory looming stimuli. The present results also run counter to a simple instantiation of the principle of inverse effectiveness, wherein receding stimuli would have been predicted to yield the largest enhancement of behavior and brain activity (though not forcibly the largest absolute amplitude responses). This was clearly not the case either with regard to the facilitation of reaction times (Cappe et al., 2009b) or ERPs (Fig. 4). In agreement, the extant literature in humans provides several replications from independent laboratories of early-latency (i.e., <100 ms poststimulus onset) nonlinear neural response interactions between high-intensity auditory-visual stimulus pairs (Giard and Peronnet, 1999; Teder-Sälejärvi et al., 2002; Gondan and Röder, 2006; Cappe et al., 2010; Raij et al., 2010; but see Senkowski et al., 2011), although it remains to be detailed under which circumstances effects are superadditive versus subadditive (cf. Cappe et al., 2010 for discussion).

Evidence for synergy between principles of multisensory interactions is likewise accumulating in studies of single-unit spiking activity within the cat anterior ectosylvian sulcus. The innovative discovery is that the firing rate within individual receptive fields of neurons is heterogeneous and varies with stimulus effectiveness in spatially and temporally dependent manners (for review, see Krueger et al., 2009). Superadditive and subadditive hotspots are not stationary within the neuron's receptive field either in cortical (Carriere et al., 2008) or subcortical (Royal et al., 2009) structures and furthermore are not straightforwardly predicted by unisensory response patterns. These features were further evident when data were analyzed at a population level, such that the percentage of integration was higher (in their population of neurons) along the horizontal meridian than for other positions, even though response profiles were uniformly distributed (Krueger et al., 2009). Regarding potential functional consequences of this organization of responsiveness, Wallace and colleagues postulate that such heterogeneity could be efficient in encoding dynamic/moving stimuli and in generating a "normalized" response profile (at least during multisensory conditions) across the receptive field (Krueger et al., 2009). The present results may be highlighting the consequences of such architecture (to the extent it manifests in humans) on the discrimination and population-level neural response to dynamic looming stimuli.

Mechanisms subserving the integration of looming signals

Mechanistically, we show that the selective integration of multisensory looming cues manifests as a superadditive nonlinear interaction in GFP over the 73–113 ms poststimulus period in the absence of significant topographic differences between responses to multisensory stimulus pairs and summed responses from the constituent unisensory conditions. Stronger GFP is consistent with greater overall synchrony of the underlying neural activity. In this regard, our finding is therefore in keeping with observations of enhanced inter-regional synchrony between auditory core and STS regions (Maier et al., 2008), though their limited spatial sampling cannot exclude the involvement of other regions, including the claustrum/insula as well as cuneus identified in the present study. Prior research suggests that the right claustrum/insula is sensitive to multisensory congruency during object processing (Naghavi et al., 2007) as well as when determining multisensory onset (a)synchrony (Bushara et al., 2001; Calvert et al., 2001). Such functions may similarly be at play here. Over the same time interval, significant effects with multisensory looming stimuli were observed in the cuneus bilaterally; regions observed during early-latency multisensory interactions (Cappe et al., 2010; Raij et al., 2010) and during multisensory object processing (Stevenson and James, 2009; Naumer et al., 2011). This network of regions is thus in keeping with synergistic (and dynamic) processing of multisensory features present in looming stimuli, as well as with evidence for the differential processing of looming signals in the amygdala (Bach et al., 2008).

While evidence is increasingly highlighting the role of oscillatory activity in multisensory phenomena (Lakatos et al., 2007, 2008, 2009; Senkowski et al., 2008), such signal analysis methods have yet to be optimized for application to single-trial source estimations based on scalp-recorded EEG (Van Zaen et al., 2010; Ramírez et al., 2011). Such notwithstanding, our results therefore indicate there to be phase-locked and stimulus-locked activities at early poststimulus latencies that exhibit nonlinear multisensory interactions. It will be particularly informative to ascertain which oscillatory components, as well as their potential hierarchical interdependencies and prestimulus contingencies, engender the selective effects observed with multisensory looming stimuli. However, such investigations must currently await further analytical developments.

In addition to these selective interactions following multisensory looming stimuli, we also demonstrate a robust positive linear correlation between behavioral and neural indices of multisensory facilitation that was not evident for any of the other multisensory conditions in this study (Fig. 5). This further highlights the behavioral relevance of early-latency and low-level multisensory interactions in humans (Romei et al., 2007, 2009; Sperdin et al., 2009, 2010; Noesselt et al., 2010; Van der Burg et al., 2011) as well as monkeys (Wang et al., 2008). Such a linear relationship also provides further support to the suggestion that looming signals are on the one hand preferentially processed neurophysiologically (Maier et al., 2008) and on the other hand subject to perceptual biases (Maier et al., 2004). Our findings provide a first line of evidence for a causal link between these propositions.

Aside from this looming-selective effect, there was also a generally earlier transition from one stable ERP topography (and by extension configuration of active brain regions) to another one following multisensory stimuli, regardless of the direction and congruence of perceived stimulus motion, over the 73–145 ms poststimulus period. The overall timing of our effects generally concurs with prior studies using stationary stimuli that were task-relevant, task-irrelevant (but nonetheless attended) or passively presented (Giard and Peronnet, 1999; Molholm et al., 2002; Vidal et al., 2008; Cappe et al., 2010; Raij et al., 2010). While the use of dynamic stimuli may conceivably result in delayed effects relative to these prior studies, this was not the case for multisensory looming stimuli. In these prior studies, nonlinear neural responses interactions were consistently observed over the 50–100 ms poststimulus period and oftentimes thereafter within near-primary cortices.

References

Bach DR, Schächinger H, Neuhoff JG, Esposito F, Di Salle F, Lehmann C, Herdener M, Scheffler K, Seifritz E (2008) Rising sound intensity: an intrinsic warning cue activating the amygdala. Cereb Cortex 18:145–150.

Beer AL, Plank T, Greenlee MW (2011) Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. Exp Brain Res 213:200–308.

Besle J, Fort A, Giard MH (2004) Interest and validity of the additive model in electrophysiological studies of multisensory interactions. Cogn Process 5:189–192.

Besle J, Fischer C, Bidet-Caulet A, Lecaignard F, Bertrand O, Giard MH (2008) Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. J Neurosci 28:14301–14310.

Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb Cortex 17:2172–2189.

Brunet D, Murray MM, Michel CM (2011) Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput Intell Neurosci 2011:813870.

Bushara KO, Grafman J, Hallett M (2001) Neural correlates of auditory-visual stimulus onset asynchrony detection. J Neurosci 21:300–304.

Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage 14:427–438.

Calvert G, Spence C, Stein BE (eds) (2004) The handbook of multisensory processes. Cambridge, MA: MIT.

Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. Eur J Neurosci 22:2886–2902.

Cappe C, Rouiller EM, Barone P (2009a) Multisensory anatomic pathway. Hear Res 258:28–36.

Cappe C, Thut G, Romei V, Murray MM (2009b) Selective integration of auditory–visual looming cues by humans. Neuropsychologia 47:1045–1052.

Cappe C, Thut G, Romei V, Murray MM (2010) Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. J Neurosci 30:12572–12580.

Carriere BN, Royal DW, Wallace MT (2008) Spatial heterogeneity of cortical receptive fields and its impact on multisensory interactions. J Neurophysiol 99:2357–2368.

De Lucia M, Clarke S, Murray MM (2010) A temporal hierarchy for conspecific vocalization discrimination in humans. J Neurosci 30:11210–11221.

Doehrmann O, Naumer MJ (2008) Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. Brain Res 1242:136–150.

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749–5759.

Falchier A, Schroeder CE, Hackett TA, Lakatos P, Nascimento-Silva S, Ulbert I, Karmos G, Smiley JF (2010) Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. Cereb Cortex 20:1529–1538.

Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cog Sci 10:278–285.

Ghazanfar AA, Neuhoff JG, Logothetis NK (2002) Auditory looming perception in rhesus monkeys. Proc Natl Acad Sci U S A 99 15755–15757.

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

Gondan M, Röder B (2006) A new method for detecting interactions between the senses in event-relate potentials. Brain Res 1073–1074:389–397.

Gonzalez Andino SL, Murray MM, Foxe JJ, de Peralta Menendez RG (2005) How single-trial electrical neuroimaging contributes to multisensory research. Exp Brain Res 166:298–304.

Grave de Peralta Menendez R, Gonzalez Andino S, Lantz G, Michel CM, Landis T (2001) Noninvasive localization of electromagnetic epileptic activity: I. Method descriptions and simulations. Brain Topogr 14:131–137.

Grave de Peralta Menendez R, Murray MM, Michel CM, Martuzzi R, Gonzalez Andino SL (2004) Electrical neuroimaging based on biophysical constraints. Neuroimage 21:527–539.

Graziano MS, Cooke DF (2006) Parieto-frontal interactions, personal space, and defensive behavior. Neuropsychologia 44:845–859.

Guthrie D, Buchwald JS (1991) Significance testing of difference potentials. Psychophysiology 28:240–244.

Kayser C, Petkov CI, Lippert M, Logothetis NK (2005) Mechanisms for allocating auditory attention: an auditory saliency map. Curr Biol 15:1943–1947.

Kayser C, Petkov CI, Logothetis NK (2009) Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. Hear Res 258:80–88.

Knebel JF, Murray MM (2012) Towards a resolution of conflicting models of illusory contour processing in humans. Neuroimage 58:2808–2817.

Knebel JF, Javitt DC, Murray MM (2011) Impaired early visual response modulations to spatial information in chronic schizophrenia. Psychiatry Res 193:168–176.

Koenig T, Melie-García L (2010) A method to determine the presence of averaged event-related fields using randomization tests. Brain Topogr 23:233–242.

Krueger J, Royal DW, Fister MC, Wallace MT (2009) Spatial receptive field organization of multisensory neurons and its impact on multisensory interactions. Hear Res 258:47–54.

Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53:279–292.

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320:110–113.

Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE (2009) The leading sense: supramodal control of neurophysiological context by attention. Neuron 64:419–430.

Lehmann D (1987) Principles of spatial analysis. In: Handbook of electroencephalography and clinical neurophysiology, Vol 1: Methods of analysis of brain electrical and magnetic signals (Gevins AS, Remond A, eds), pp 309–405. Amsterdam: Elsevier.

Lehmann D, Skrandies W (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr Clin Neurophysiol 48:609–621.

Leo F, Romei V, Freeman E, Ladavas E, Driver J (2011) Looming sounds enhance orientation sensitivity for visual stimuli on the same side as such sounds. Exp Brain Res 213:193–201.

Lu T, Liang L, Wang X (2001) Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. J Neurophysiol 85:2364–2380.

Maier JX, Neuhoff JG, Logothetis NK, Ghazanfar AA (2004) Multisensory integration of looming signals by rhesus monkeys. Neuron 43:177–181.

Maier JX, Chandrasekaran C, Ghazanfar AA (2008) Integration of bimodal looming signals through neuronal coherence in the temporal lobe. Curr Biol 18:963–968.

Martuzzi R, Murray MM, Michel CM, Thiran JP, Maeder PP, Clarke S, Meuli RA (2007) Multisensory interactions within human primary cortices revealed by BOLD dynamics. Cereb Cortex 17:1672–1679.

Martuzzi R, Murray MM, Meuli RA, Thiran JP, Maeder PP, Michel CM, Grave de Peralta Menendez R, Gonzalez Andino SL (2009) Methods for determining frequency- and region-dependent relationships between estimated LFPs and BOLD responses in humans. J Neurophysiol 101:491–502.

Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, Grave de Peralta R (2004) EEG source imaging. Clin Neurophysiol 115:2195–2222.

Michel CM, Koenig T, Brandeis D, Gianotti LR, Wackermann J (2009) Electrical neuroimaging. Cambridge UK: Cambridge UP.

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res 14:115–128.

Murray MM, Spierer L (2009) Auditory spatio-temporal brain dynamics and their consequences for multisensory interactions in humans. Hear Res 258:121–133.

Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ (2005) Grabbing your ear: rapid auditory somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex 15:963–974.

Murray MM, Brunet D, Michel CM (2008) Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr 20:249–264.

Naghavi HR, Eriksson J, Larsson A, Nyberg L (2007) The claustrum/insula region integrates conceptually related sounds and pictures. Neurosci Lett 422:77–80.

Naumer MJ, van den Bosch JJ, Wibral M, Kohler A, Singer W, Kaiser J, van de Ven V, Muckli L (2011) Investigating human audio-visual object perception with a combination of hypothesis-generating and hypothesis-testing fMRI analysis tools. Exp Brain Res 213:309–320.

Neuhoff JG (1998) Perceptual bias for rising tones. Nature 395:123–124.

Neuhoff JG (2001) An adaptive bias in the perception of looming auditory motion. Ecol Psychol 13:87–110.

Noesselt T, Tyll S, Boehler CN, Budinger E, Heinze HJ, Driver J (2010) Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. J Neurosci 30:13609–13623.

Ohshiro T, Angelaki DE, DeAngelis GC (2011) A normalization model of multisensory integration. Nat Neurosci 14:775–782.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113.

Perrin F, Pernier J, Bertrand O, Giard MH, Echallier JF (1987) Mapping of scalp potentials by surface spline interpolation. Electroencephalogr Clin Neurophysiol 66:75–81.

Pourtois G, Delplanque S, Michel C, Vuilleumier P (2008) Beyond conventional event-related brain potential (ERP): exploring the time-course of visual emotion processing using topographic and principal component analyses. Brain Topography 20:265–277.

Raij T, Ahveninen J, Lin FH, Witzel T, Jääskeläinen IP, Letham B, Israeli E, Sahyoun C, Vasios C, Stufflebeam S, Hämäläinen M, Belliveau JW (2010) Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. Eur J Neurosci 31:1772–1782.

Ramírez RR, Kopell BH, Butson CR, Hiner BC, Baillet S (2011) Spectral signal space projection algorithm for frequency domain MEG and EEG denoising, whitening, and source imaging. Neuroimage 56:78–92.

Remedios R, Logothetis NK, Kayser C (2010) Unimodal responses prevail within the multisensory claustrum. J Neurosci 30:12902–12907.

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50:19–26.

Romei V, Murray MM, Merabet LB, Thut G (2007) Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions. J Neurosci 27:11465–11472.

Romei V, Murray MM, Cappe C, Thut G (2009) Pre-perceptual and stimulus-selective enhancement of human low-level visual cortex excitability by sounds. Curr Biol 19:1799–1805.

Royal DW, Carriere BN, Wallace MT (2009) Spatiotemporal architecture of cortical receptive fields and its impact on multisensory interactions. Exp Brain Res 198:127–136.

Schiff W (1965) Perception of impending collision: a study of visually directed avoidant behavior. Psychol Monogr 79:1–26.

Schiff W, Caviness JA, Gibson JJ (1962) Persistent fear responses in rhesus monkeys to the optical stimulus of "looming". Science 136:982–983.

Seifritz E, Neuhoff JG, Bilecen D, Scheffler K, Mustovic H, Schächinger H, Elefante R, Di Salle F (2002) Neural processing of auditory looming in the human brain. Curr Biol 12:2147–2151.

Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. Trends Neurosci 31:401–409.

Senkowski D, Saint-Amour D, Höfle M, Foxe JJ (2011) Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. Neuroimage 56:2200–2208.

Sperdin HF, Cappe C, Foxe JJ, Murray MM (2009) Early, low-level auditory somatosensory multisensory interactions impact reaction time speed. Front Integr Neurosci 3:2.

Sperdin HF, Cappe C, Murray MM (2010) The behavioral relevance of multisensory neural response interactions. Front Neurosci 4:9.

Stecker GC, Middlebrooks JC (2003) Distributed coding of sound locations in the auditory cortex. Biol Cybern 89:341–349.

Stein BE, Meredith MA (1993) The merging of the senses. Cambridge, MA: MIT.

Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci 9:255–266.

Stevenson RA, James TW (2009) Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. Neuroimage 44:1210–1223.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system—an approach to cerebral imaging. New York: Thieme.

Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event related potential (ERP) recordings. Brain Res Cogn Brain Res 14:106–114.

Toepel U, Knebel JF, Hudry J, le Coutre J, Murray MM (2009) The brain tracks the energetic value in food images. Neuroimage 44:967–974.

Van der Burg E, Talsma D, Olivers CN, Hickey C, Theeuwes J (2011) Early multisensory interactions affect the competition among multiple visual objects. Neuroimage 55:1208–1218.

Van Zaen J, Uldry L, Duchêne C, Prudat Y, Meuli RA, Murray MM, Vesin JM (2010) Adaptive tracking of EEG oscillations. J Neurosci Methods 186:97–106.

Vidal J, Giard MH, Roux S, Barthélémy C, Bruneau N (2008) Crossmodal processing of auditory-visual stimuli in a no-task paradigm: a topographic event-related potential study. Clin Neurophysiol 119:763–771.

Wallace MT, Ramachandran R, Stein BE (2004) A revised view of sensory cortical parcellation. Proc Natl Acad Sci U S A 101:2167–2172.

Wang Y, Celebrini S, Trotter Y, Barone P (2008) Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. BMC Neurosci 9:79.

Woods TM, Lopez SE, Long JH, Rahman JE, Recanzone GH (2006) Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. J Neurophysiol 96:3323–3337.

NeuroImage 62 (2012) 1478–1488

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Electrical neuroimaging of memory discrimination based on single-trial multisensory learning

Antonia Thelen [a], Céline Cappe [b], Micah M. Murray [a,c,d,e,*]

[a] The Functional Electrical Neuroimaging Laboratory, Department of Clinical Neurosciences, Vaudois University Hospital Center and University of Lausanne, Lausanne, Switzerland
[b] Laboratory of Psychophysics, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[c] Radiology Department, Vaudois University Hospital Center and University of Lausanne, Lausanne, Switzerland
[d] Electroencephalography Brain Mapping Core, Center for Biomedical Imaging (CIBM) of Lausanne and Geneva, Switzerland
[e] Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

ARTICLE INFO

Article history:
Accepted 10 May 2012
Available online 18 May 2012

Keywords:
Multisensory
Cross-modal
Event-related potential (ERP)
Memory
Object recognition
Auditory cortex

ABSTRACT

Multisensory experiences influence subsequent memory performance and brain responses. Studies have thus far concentrated on semantically congruent pairings, leaving unresolved the influence of stimulus pairing and memory sub-types. Here, we paired images with unique, meaningless sounds during a continuous recognition task to determine if purely episodic, single-trial multisensory experiences can incidentally impact subsequent visual object discrimination. Psychophysics and electrical neuroimaging analyses of visual evoked potentials (VEPs) compared responses to repeated images either paired or not with a meaningless sound during initial encounters. Recognition accuracy was significantly impaired for images initially presented as multisensory pairs and could not be explained in terms of differential attention or transfer of effects from encoding to retrieval. VEP modulations occurred at 100–130 ms and 270–310 ms and stemmed from topographic differences indicative of network configuration changes within the brain. Distributed source estimations localized the earlier effect to regions of the right posterior temporal gyrus (STG) and the later effect to regions of the middle temporal gyrus (MTG). Responses in these regions were stronger for images previously encountered as multisensory pairs. Only the later effect correlated with performance such that greater MTG activity in response to repeated visual stimuli was linked with greater performance decrements. The present findings suggest that brain networks involved in this discrimination may critically depend on whether multisensory events facilitate or impair later visual memory performance. More generally, the data support models whereby effects of multisensory interactions persist to incidentally affect subsequent behavior as well as visual processing during its initial stages.

Introduction

Perceptions are often affected by the way in which information from several sensory modalities are combined, i.e. under multisensory conditions (Calvert, 2004; Stein and Meredith, 1993). In addition to their immediate effects on perception, multisensory experiences at one point in time can impact unisensory processing during later encounters (Gottfried et al., 2004; Kim et al., 2008; Lehmann and Murray, 2005; Meylan and Murray, 2007; Murray et al., 2004, 2005; Shams and Seitz, 2008). The circumstances under which such multisensory memory traces impact subsequent unisensory retrieval remain unresolved and were the focus of the present study. This issue falls within the larger framework of differential mechanisms and consequences of multisensory versus unisensory learning (Murray and Sperdin, 2010; Shams and Seitz, 2008; Shams et al., 2011).

The impact of multisensory learning upon unisensory (visual) recognition has been studied mainly through two paradigms. In one paradigm, subjects underwent extensive multisensory training before unisensory recognition performance was assessed in a separate retrieval session (Nyberg et al., 2000; Wheeler and Petersen, 2000). The task was to explicitly remember the context in which stimuli had been presented during encoding (unisensory or multisensory). In another paradigm, the impact of multisensory experiences upon subsequent unisensory recognition was assessed through single-trial learning during a continuous recognition task (Lehmann and Murray, 2005; Murray et al., 2004, 2005). Encoding and retrieval were separated by (relatively) short time intervals in the same experimental block, and subjects indicated whether or not images were being presented for the first or repeated time.

These paradigms have led to discordant results both in terms of performance and brain activity. For example, Nyberg et al. (2000) observed relatively poorer memory performance accuracy for words that had been paired with sounds during encoding vs. words that had not (76% vs. 84%, respectively). In terms of brain activity during

---

[*] Corresponding author at: Department of Clinical Neurosciences, CHUV, BH08.078, Rue du Bugnon 46, 1011 Lausanne, Switzerland. Fax: +41 21 314 1319.
E-mail address: Micah.Murray@chuv.ch (M.M. Murray).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.05.027

the retrieval phase, these authors found that auditory cortices were active in response to visually presented words that had been paired with corresponding sounds during the encoding phase (see also Wheeler and Petersen, 2000 for similar findings with pictures of objects). They took this activity pattern as evidence that networks active during encoding are re-activated during retrieval; a pattern more generally consistent with theories of redintegration (Hamilton, 1859). Under the framework of redintegration, a component part of a consolidated memory is sufficient to (re)activate the whole experience's representation. If the consolidated memory, for example, entails both auditory and visual components, then stimulating with sounds would lead to both auditory and visual cortical activity (see also Rugg et al., 2008 for a similar framework).

By contrast, our group has consistently observed improved memory performance for images that had been paired with their corresponding sounds versus those images only ever presented in a unisensory manner (Lehmann and Murray, 2005; Murray et al., 2004, 2005; Murray and Sperdin, 2010). Plus, both electrical and hemodynamic imaging studies indicated that regions of the lateral occipital cortices differentiate between image repetitions that had been previously paired with sounds or not (Murray et al., 2004, 2005). Stronger responses were observed for visual stimuli that had been previously presented with a semantically congruent sound. There was no evidence for effects within auditory cortices in these studies by Murray and colleagues. Moreover, the timing of effects observed in Murray et al. (2004) would suggest that past multisensory experiences impact the initial stages of visual processing (i.e. within the initial 100 ms post-stimulus onset). Effects within visual cortices do not appear to be limited to tasks requiring visual discrimination. For example, von Kriegstein and Giraud (2006) observed activation changes within the fusiform face area in response to voices as a function of prior learned associations with faces. Others have observed activations within visual cortices in response to meaningless complex sounds previously paired with meaningless images (though activations were smaller than in response to visual stimuli) (Butler and James, 2011). In this study, the magnitude of the activation did not appear to be directly linked to subjects' accuracy in indicating if a given sound/image had been previously paired with a sound vs. image.

These discrepant patterns of brain activity (i.e. differential effects within visual vs. auditory cortices) may stem from paradigmatic differences, including whether or not subjects must explicitly discriminate the context in which a given stimulus had been paired. Another alternative is that the activation of auditory regions in response to visual stimuli during memory-related tasks is a hallmark of performance impairments rather than redintegration per se. A continuous recognition task has been used to identify conditions under which memory performance is impaired for image repetitions that were initially presented in a multisensory context. While performance was enhanced for stimuli that had been initially encountered in a semantically congruent multisensory context, it was unchanged if initially encountered in a semantically incongruent context, and was impaired if initially encountered with a meaningless sound (i.e. pure tone) (Lehmann and Murray, 2005). All of these effects were relative to performance with images that were initially encountered in a unisensory context (i.e. the same image repeated once). It is likewise important to note that these effects on accuracy did not co-occur with parallel modulations in reaction times (in fact, reaction times did not differ) and were not the consequence of carry-over effects from encoding (reviewed in Murray and Sperdin, 2010; see also Baier et al., 2006 for similar work on multisensory expectancy). The performance impairment we observed with meaningless sounds was confounded by the fact that the same sound was paired with multiple objects across trials, thereby making it possible that the representation of any given visual object is effectively contaminated by a constant distracter (here the auditory stimulus).

The present study sought to resolve the role of episodic pairings on single-trial multisensory memory processes by combining psychophysics with electrical neuroimaging of VEPs. On the one hand, if single-trial episodic events are sufficient to generate a perceptual/memory trace then performance accuracy would be predicted to be impaired for repetitions of images that had been paired with sounds on their initial encounter. This would suggest that the observations of Lehmann and Murray (2005) are due to the episodic pairing rather than to the repeated association/dissociation of the same sound with multiple images across trials. Alternatively, no performance differences (vs. images that had only been encountered visually) would be predicted if the unique, meaningless sounds are treated as if they were an incongruent event akin to pairing the sound of one (known) object with the image of another object. On the other hand, prior brain imaging data would suggest that differential responses to incoming visual stimuli can manifest as early as ~100 ms post-stimulus onset and are indicative of the incidental discrimination of stored object representations according to past (semantic) multisensory experiences (Murray et al., 2004). To the extent that single-trial episodic pairings are similarly effective in establishing distinct object representations, we should observe differential VEPs at equally early latencies. Because electrical neuroimaging analyses distinguish between effects due to modulations in response strength and response topography (the latter of which is forcibly indicative of generator changes), we could likewise assess if and when distinct generator configurations respond to visual stimuli previously encountered in a unisensory vs. multisensory manner. When combined with distributed source estimations, we can then disentangle whether or not differential activity, putatively arising within auditory cortices and perhaps elsewhere, is linked to performance decrements or to redintegration processes.

Materials and methods

Participants

The experiment included 22 (11 women) volunteers aged between 23 and 30 years (mean age ± SD = 26.45 ± 3.1 years). The last 12 of these completed the psychophysics paradigm while EEG was simultaneously recorded (detailed below). The study was conducted in accordance with the Declaration of Helsinki, and all subjects provided their written 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. Nineteen subjects were right-handed and the remaining 3 left-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 as well as reported normal hearing.

Task

Subjects performed a continuous recognition task, which required the discrimination of initial from repeated presentations of line drawings that were pseudo-randomized within a block of trials. They were instructed to perform as quickly and as accurately as possible. Further, each object (irrespective of whether it was initially presented in a unisensory or multisensory context) was only repeated once throughout the duration of the experiment.

The pictures were subdivided into two groups. Initial presentations were either unisensory or multisensory. Repeated presentations were always unisensory. Thus, half of the repeated presentations had been multisensory when initially encountered and the other half had been unisensory when initially encountered. We will refer to our experimental conditions as V, for the visual-only initial condition and V− for the same stimuli when repeated. AV will refer to the initial auditory-visual presentation, whereas V+ will refer to the repeated presentation of the visual component of these stimuli (Fig. 1a).

The line drawings were taken from a standardized set (Snodgrass and Vanderwart, 1980) or obtained from an online library (dgl.microsoft.com), and included a mix of living and non-living stimuli (see Appendix for full list). The pictures were equally subdivided over experimental conditions and blocks as described above. Plus, the different categories of objects were equally intermixed. In order to minimize the possibility that the observed effects were due to low-level visual features, we analyzed the spatial frequency spectra and the luminance between the two image groups (AV and V), and we did not find any difference between image groups. The full details of these procedures have been reported by Knebel et al. (2008). The images were presented centrally and appeared black on a white background. On initial presentations these visual stimuli could (50%) or could not be paired with a meaningless sound (created with Adobe Audition 1.0). These sounds differed in their spectral composition, ranging from 100 Hz to 4700 Hz and were sometimes modulated in terms of amplitude envelopes and/or waveform types (triangular and sinusoid). All sounds were 500 ms duration (10 ms rise/fall, in order to avoid clicks; 16bit stereo; 44100 Hz digitization).

All stimuli were presented for 500 ms, followed by a randomized inter-stimulus interval (ISI) ranging from 900 to 1500 ms. The mean (± SD) number of trials between the initial and the repeated presentation of the same image was 9 ± 4 pictures for either presentation condition (V and AV). Also the distribution of old and new pictures throughout the length of the block was controlled, so as to avoid response-decision bias and to maintain an equal probability of a "new" object across quartiles within a block. 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. Within a block there were 136 trials, equally divided between V, AV, V−, and V+ conditions (i.e. 34 trials each). This is identical to the block length used in our prior studies (Lehmann and Murray, 2005; Murray et al., 2004).

The experiment took place in sound-attenuated chamber, where subjects were seated centrally in front of a 20" computer monitor (HP LP2068), located about 140 cm away from them (visual angle ~4°). The auditory stimuli were presented over insert earphones (Etymotic model: ER4S), and the volume was adjusted to a comfortable level (~62 dB). The stimuli were all presented and controlled by E-Prime 2.0, and all behavioral data were recorded in conjunction with the serial response box (Psychology Software Tools, Inc.; www.pstnet.com). All participants completed 2 blocks of trials. While this paradigm indeed introduces a degree of spatial disparity between the auditory and visual modalities (when stimuli are multisensory), we do not believe that such would be impeding any interactions here. Prior research has reliably documented integrative effects on behavior and brain responses using similar setups (e.g. Cappe et al., 2010, 2012; Raij et al., 2010).

The behavioral data were treated as follows: Mean RT (in [ms]) and accuracy ([%] of correct responses) were calculated for each subject and condition (V, AV, V−, and V+) separately. We then performed paired *t*-tests, specifically comparing the initial encoding conditions (V vs. AV) or the repeated retrieval conditions (V− vs. V+). Finally, we sought to assess whether effects during encoding (i.e. differences between V vs. AV) were directly linked to and/or predictive of effects during retrieval (i.e. differences between V− vs. V+), given that prior studies would suggest that these are dissociable (reviewed in Murray and Sperdin, 2010). As will be made clearer in the Results below, we

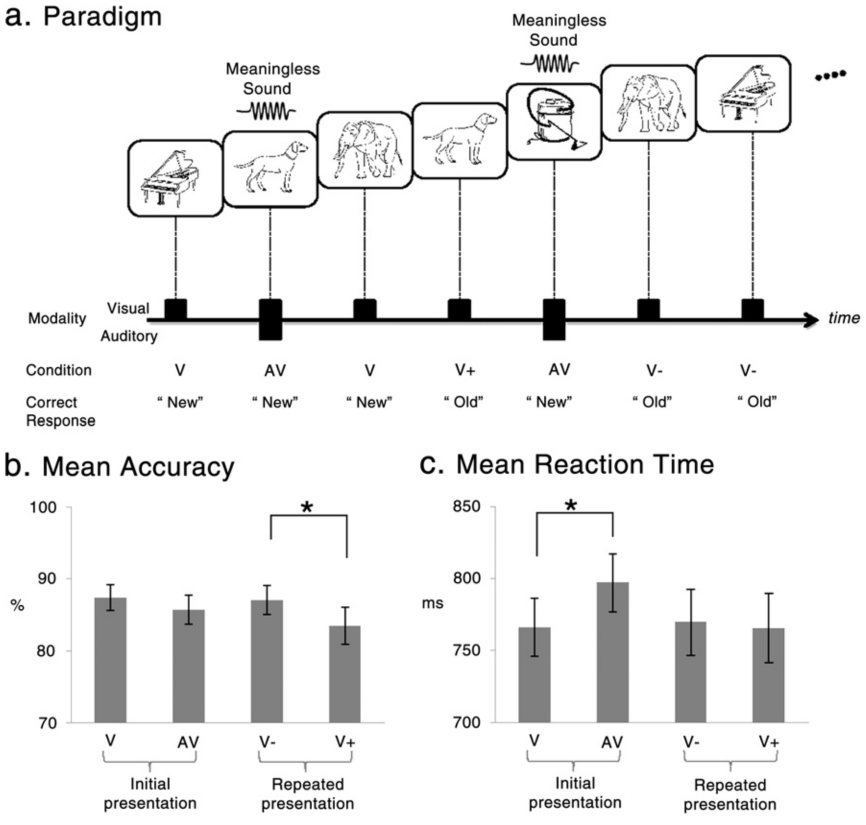


Fig. 1. (a) Illustration of the paradigm used to investigate multisensory encoding and recall in working memory. (b) Group-averaged accuracy (± s.e.m.) for the four experimental conditions: initial presentations of visual (V) and auditory-visual (AV) and their repeated presentation (V− and V+, respectively). (c) Group-averaged mean reaction times (± s.e.m.) for the same experimental conditions as in (b). Significant effects ($p<0.05$) are marked with an asterisk.

calculated the correlation coefficient between the differences in reaction time during encoding and the difference in performance accuracy during retrieval.

EEG acquisition and pre-processing

Continuous EEG was acquired from 160 scalp electrodes (sampling rate at 1024 Hz) using a Biosemi ActiveTwo system from a subset of 12 subjects (3 women; mean age ± SD = 27.1 ± 3.5 years) who were the last of the abovementioned 22 participants to take part in the psychophysics paradigm. Data pre-processing and analyses were performed using Cartool (Brunet et al., 2011; http://sites.google.com/site/fbmlab/cartool). Epochs from 100 ms pre-stimulus to 500 ms post-stimulus onset were averaged for each of the four experimental conditions and from each subject in order to calculate the VEPs. In addition to a ±80 μV artifact rejection, EEG epochs containing eye blinks or other noise transients were removed based on a trial-by-trial visual inspection of the data. Before group averaging, data from artifact electrodes of each subject were interpolated using 3-D splines (Perrin et al., 1987). On average, 5 of the 160 channels were interpolated (range 2–12). ERP data were baseline corrected using the pre-stimulus period, band-pass filtered (0.1–60 Hz including a notch at 50 Hz) and recalculated against the average reference. On average, there were 60 (range 43–68) and 61 (range 42–68) accepted epochs for the V− and V+ conditions, respectively.

VEP analyses

The VEP analyses were based on the hypothesis that a differential neural response would be found between the V+ and V− conditions (Murray et al., 2004, 2005). The approach we used here has been referred to as electrical neuroimaging and is based largely on the multivariate analysis of global features of the electric field at the scalp that in turn inform the selection of time periods for analyses of source estimations (Michel and Murray, 2012; Murray et al., 2008; Tzovara et al., in press). These electrical neuroimaging analyses allowed us to differentiate effects following from modulations in the strength of responses of statistically indistinguishable brain generators from alterations in the configuration of these generators (viz. the topography of the electric field at the scalp), as well as latency shifts in brain processes across experimental conditions. Additionally, we applied the local auto-regressive average distributed linear inverse solution (LAURA; Grave de Peralta Menendez et al., 2001, 2004) to visualize and statistically contrast the likely underlying sources of effects identified during the preceding analysis steps of the surface-recorded VEPs.

As a first level of analysis, we analyzed waveform data from all electrodes as a function of time post-stimulus onset in a series of pair-wise comparisons (*t*-tests) between responses to the V+ and V− conditions. Temporal auto-correlation at individual electrodes was corrected through the application of an 11 contiguous data-point temporal criterion (~10 ms at 1024 Hz sampling) for the persistence of differential effects (Guthrie and Buchwald, 1991). Similarly, spatial correlation was addressed by considering as reliable only those effects that entailed at least 5% of the electrodes from the 160-channel montage (i.e. 8). These combined criteria were applied to correct for multiple comparisons, though we note that our conclusions are based on reference-independent global measures of the electric field at the scalp. We would therefore also emphasize that the number of electrodes exhibiting an effect at a given latency will depend on the reference, and this number is not constant across choices of reference because significant effects are not simply re-distributed across the montage (discussed in Tzovara et al., in press). Likewise, the use of an average reference receives support from biophysical laws as well as the implicit re-centering of VEP data to such when performing source estimations (discussed in Brunet et al., 2011). Analyses of VEP voltage waveform data (vs. the average reference) are presented here to provide a clearer link between canonical VEP analysis approaches and electrical neuroimaging. The results of this VEP waveform analysis are presented as an area plot that shows the number of electrodes exhibiting a significant effect as a function of time (relative to stimulus onset). This type of display was chosen to provide a sense of the dynamics of a statistical effect between conditions. While these analyses give a visual impression of specific effects within the dataset, our conclusions are principally based on reference-independent global measures of the electric field at the scalp that are described below.

The global electric field strength was quantified using global field power (GFP) (Lehmann and Skrandies, 1980). This measure is equivalent to the standard deviation of the voltage potential values across the entire electrode montage at a given time point and represents a reference-independent measure of the VEP strength (Murray et al., 2008; Koenig and Melie-Garcia, 2010). GFP was statistically contrasted using a millisecond-by-millisecond paired *t*-test in conjunction with the abovementioned temporal criterion for significant effects to correct for multiple contrasts. While this dependent measure provides an assay of VEP strength, it is inherently insensitive to spatial (i.e. topographic) variation in the VEP across conditions.

In order to test the VEP topography independently of its strength, we used Global Dissimilarity (DISS) (Lehmann and Skrandies, 1980). DISS is equivalent to the square root of the mean of the squared difference between the potentials measured at each electrode for different conditions, normalized by the instantaneous GFP. It is also directly related to the (spatial) correlation between two normalized vectors (cf. Appendix in Murray et al., 2008). We then performed a non-parametric randomization test (TANOVA, Murray et al., 2008). The DISS value at each time point is compared to an empirical distribution derived from permuting the condition label of the data from each subject. Because changes in topography forcibly follow from changes in the configuration of the underlying active sources (Lehmann, 1987), this analysis reveals when the experimental conditions activated distinct sets of brain networks.

The collective post-stimulus group-average VEPs were subjected to a topographic cluster analysis based on a hierarchical clustering algorithm (Murray et al., 2008). This clustering identifies stable electric field topographies (hereafter template maps). The clustering is exclusively sensitive to topographic modulations, because the data are first normalized by their instantaneous GFP. The optimal number of temporally stable VEP clusters (i.e. the minimal number of maps that accounts for the greatest variance of the dataset) was determined using a modified Krzanowski-Lai criterion (Murray et al., 2008). The clustering makes no assumption on the orthogonality of the derived template maps (De Lucia et al., 2010a, 2010b; Pourtois et al., 2008). Template maps identified in the group-average VEP were then submitted to a fitting procedure wherein each time point of each single-subject VEP is labeled according to the template map with which it best correlated spatially (Murray et al., 2008) so as to statistically test the relative presence of each template map in the moment-by-moment scalp topography of the VEP and the differences in such across conditions. These values can be expressed as the probability of a given template map yielding a higher spatial correlation in the single-subject data from each condition. Statistical analysis of these values was performed with an analysis of variance (ANOVA).

Source estimations

We estimated the localization of the electrical activity in the brain using a distributed linear inverse solution (minimum norm) applying the LAURA regularization approach comprising biophysical laws as constraints (Grave de Peralta Menendez et al., 2001, 2004; see also Michel et al., 2004 for review). LAURA selects the source configuration that better mimics the biophysical behavior of electric vector fields (i.e. activity at one point depends on the activity at neighboring points according to electromagnetic laws). In our study, homogenous regression coefficients in all directions and within the whole solution space were used. LAURA uses a realistic head model, and the solution space included 4024 nodes, selected from a 6×6×6 mm grid equally distributed within the gray matter of the Montreal Neurological Institute's average brain (courtesy of R. Grave de Peralta and S. Gonzalez Andino; http://www.electrical-neuroimaging.ch/). Prior basic and clinical research from members of our group and others has documented and discussed in detail the spatial accuracy of the inverse solution model used here (e.g. Gonzalez Andino et al., 2005; Grave de Peralta Menendez et al., 2004; Martuzzi et al., 2009; Michel et al., 2004). In general, the localization accuracy is considered to be along the lines of the matrix grid size (here 6 mm). The results of the above topographic pattern analysis defined time periods for which intracranial sources were estimated and statistically compared between conditions (here 73–113 ms post-stimulus). Prior to calculation of the inverse solution, the VEP data were down-sampled and affine-transformed to a common 111-channel montage. Statistical analyses of source estimations were performed by first averaging the VEP data across time to generate a single data point for each participant and condition. This procedure increases the signal-to-noise ratio of the data from each participant. The inverse solution was then estimated for each of the 4024 nodes. These data were then submitted to a paired *t*-test.

We combined two statistical criteria for concluding that an effect was reliable. First, a spatial extent criterion of at least 17 contiguous significant nodes was applied (see also Cappe et al., 2010, 2012; De Lucia et al., 2010a, 2010b; Knebel and Murray, 2012; Knebel et al., 2011; Toepel et al., 2009 for a similar spatial criterion). This spatial criterion was determined using the AlphaSim program (available at http://afni.nimh.nih.gov) and assuming a spatial smoothing of 6 mm full-width half maximum. This criterion indicates that there is a 3.54% probability of a cluster of at least 17 contiguous nodes, which gives an equivalent node-level *p*-value of $p \leq 0.0002$. Second and because distributed source models yield non-zero values in all solution points, it is conceivable that statistical effects will be obtained in nodes that are weakly responsive (i.e. have current density values close to zero or alternatively well below the mean across the entire set of nodes in the brain volume). To minimize the contribution of such "erroneous" or "ghost" sources, we removed all nodes with current density values less than or equal to two standard deviations below the volume's mean within each condition (here, V+: mean ± SD = $0.0073 \pm 0.0025\ \mu\text{A/mm}^3$ and V−: = $0.0071 \pm 0.0026\ \mu\text{A/mm}^3$). In this way, we sought to limit statistical effects to nodes that could reasonably be described as "active" sources. The results of the source estimations were rendered on the Montreal Neurologic Institute's average brain with the Talairach and Tournoux (1988) coordinates of the largest statistical differences within a cluster indicated.

Correlation analysis

To test whether there was a linear relationship between brain activity within source estimations and behavior, we correlated (Pearson's correlation after first testing for normality of the distributions using a Kolmogorov–Smirnov test) the difference in performance accuracy between V+ and V− conditions with the corresponding difference in source strength. Specifically, we subtracted the percent correct performance on the V+ condition from that from the V− condition for each subject. The [(V+) − (V−)] difference was calculated for the mean activity within the significant clusters (pSTG/STS and mSTG/MTG; see Results for details) for every subject.

Results

Behavioral data

Analysis of performance accuracy (Fig. 1b) revealed that participants were equally capable of indicating the initial presentation of stimuli both when presented in a multisensory auditory-visual context (AV = 85.7 ± 2%; mean ± s.e.m.) and also when presented in a unisensory visual context (V = 87.4 ± 1.8%; $t_{(21)} = 1.452$; $p = 0.161$; $\eta_p^2 = 0.681$). There was therefore no indication of encoding differences in terms of discrimination accuracy. By contrast, performance when indicating image repetitions was significantly impaired when the initial presentation had entailed a multisensory vs. unisensory context (i.e. V+ vs. V−; 83.5 ± 2.6% vs. 87 ± 2%; $t_{(21)} = -2.38$; $p = 0.027$; $\eta_p^2 = 0.674$). This difference indicates that image repetitions are incidentally discriminated according to past multisensory experiences. Though not a priori part of our research aims, we also tested the role of object category (i.e. living vs. man-made) on image repetition discrimination as a function of prior multisensory vs. unisensory pairings. The accuracy data were subjected to a 2×2 within subjects ANOVA with factors of object category (living and man-made) and experimental condition (V− and V+). There was a significant main effect of experimental condition ($F_{(1,21)} = 4.522$; $p = 0.045$; $\eta_p^2 = 0.177$). Neither the main effect of object category ($p = 0.058$) nor the interaction ($p = 0.797$) reached the 0.05 significance criterion.

In contrast with the pattern observed with accuracy rates (but highly consistent with prior work; e.g. Lehmann and Murray, 2005), mean reaction times differed for initial presentations, but not for image repetitions (Fig. 1c). Specifically, reaction times to initial presentations under multisensory conditions were significantly slower than those to initial presentations under unisensory conditions (AV vs. V = 797 ± 20 ms vs. 766 ± 20 ms; $t_{(21)} = -4.233$; $p < 0.001$; $\eta_p^2 = 0.873$). Reaction times to image repetitions did not significantly differ (V+ vs. V− = 765 ± 24 ms vs. 769 ± 23 ms; $t_{(21)} = 0.800$; $p = 0.433$; $\eta_p^2 = 0.955$). Finally, we examined if there was a linear relationship between reaction time differences during encoding and accuracy differences during retrieval. There was no evidence of a significant correlation ($r_{(20)} = -0.008$; $p > 0.9$), providing no evidence for a carry-over effect.

This pattern of results rules out explanations in terms of auditory capture of attention, selective attention, or novel context detection facilitating perceptual memory trace formation (Ranganath and Rainer, 2003), as well as general alerting. If auditory capture of attention were driving our effects, then a significant correlation would have been expected between the magnitude of this capture (quantified as the reaction time difference between visual and multisensory conditions during initial image presentations) and the magnitude of its effect on memory discrimination (i.e. accuracy levels during repeated imaging presentations). Similarly, if selective attention could account for our findings, then faster reaction times would have been expected for initial presentations of multisensory vs. unisensory stimuli (i.e. AV vs. V). Rather, our results suggest that participants were not overtly attending to the auditory channel, which in principle could have cued participants to respond "new." Slower reaction times for the multisensory condition suggest that participants did not expect auditory events (Spence et al., 2001), and accuracy did not reliably differ. We also rule out possible bias due to general arousal and fatigue, because the relative distribution of "new" and "old" pictures was maintained throughout a block of trials (see Materials and methods).

Surface VEP data

The VEP analyses focused on differences between the V+ and V− conditions (see Materials and methods). Fig. 2 displays VEPs from the V+ and V− conditions (Fig. 2a) as well as the results of electrical neuroimaging analyses (Fig. 2b–f). Analyses of VEP voltage waveforms from the entire electrode montage as a function of time revealed two time periods of significant differences – the first at ~100 ms and a second at ~270 ms (Fig. 2b) post-stimulus onset. We would remind the reader of the reference-dependent nature of analyses of voltage waveforms. Nonetheless, we supply such here to provide a clearer link between electrical neuroimaging and more traditional analysis approaches. Fig. 2c displays the group-average GFP waveforms as well as results of *t*-tests on the GFP as a function of time (note that 1 minus *p*-value is displayed). There was no evidence of GFP differences between conditions, providing no evidence for differences in response strength as a function of past multisensory vs. unisensory experience. Fig. 2d displays the DISS between the group-average VEPs from the V+ and V− conditions. Significant topographic differences were observed over the 100–130 ms and 270–310 ms post-stimulus intervals. We next determined whether these topographic effects stem from the predominance of different stable map configurations in each condition or instead from latency shifts across conditions. We subjected the cumulative group-average VEPs from both conditions to a clustering analysis. This procedure identified 19 template maps that explained 97.26% of variance of the concatenated group-averaged VEP data set. While during most of the post-stimulus period one template map could account for the

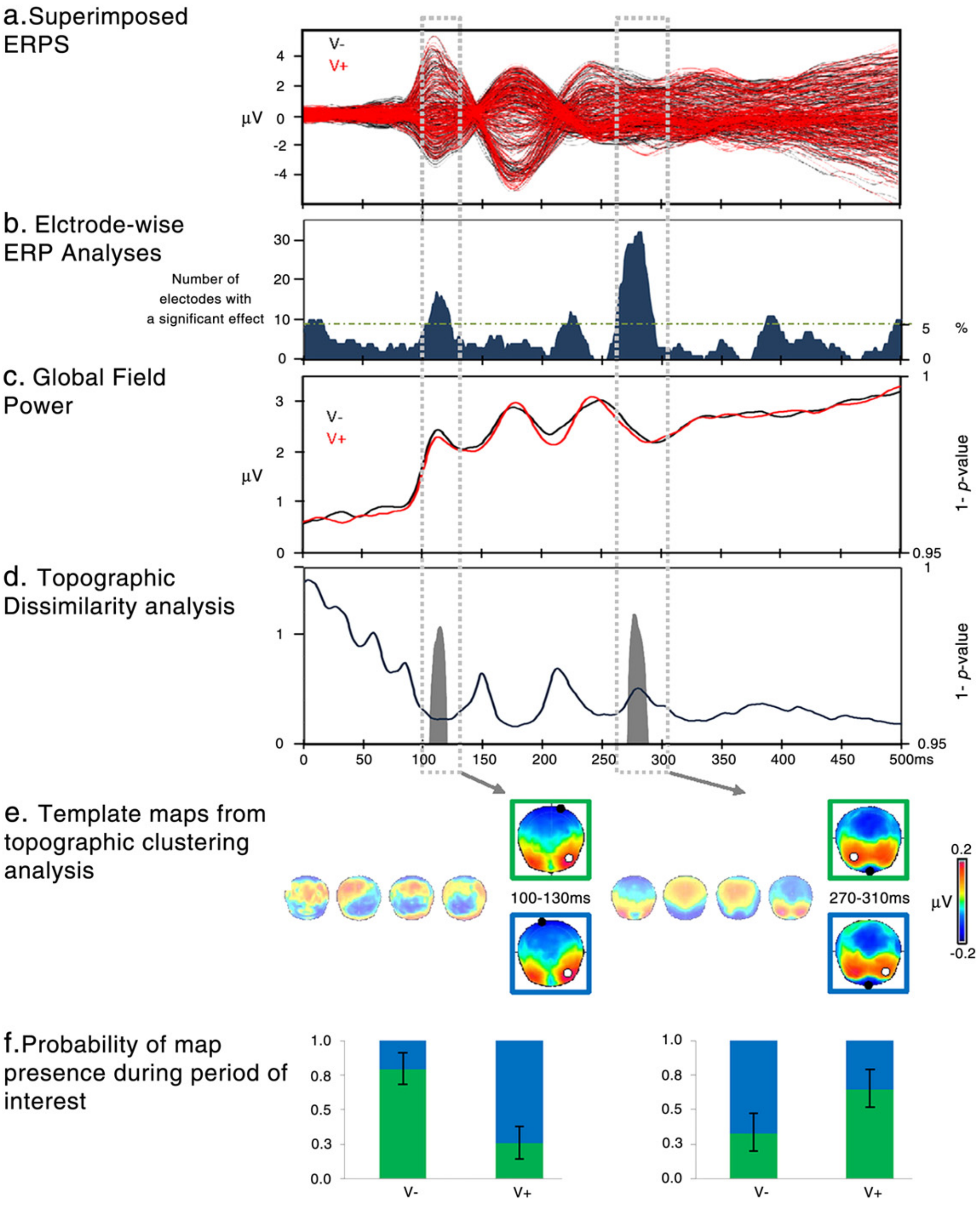


Fig. 2. Visual evoked potential (VEP) data and results of the stepwise analyses (only $p<0.05$ with an 11 consecutive time frame criterion are shown). (a) Group-averaged ($n=12$) VEP waveforms are shown superimposed across all electrodes for both experimental conditions of repeated image presentations (V− in black and V+ in red). (b) The results of the millisecond-by-millisecond paired *t*-test at each of the scalp electrodes is shown. The total number of electrodes expressing a significant difference at each TF is plotted. The dotted line marks the 5% ($n=8$) threshold of total electrode montage. (c) GFP analysis across time did not reveal significant modulations between the V− (black trace) and V+ (red trace) conditions. (d) Global dissimilarity analysis revealed two periods of differential responses at 100–130 ms and at 270–310 ms. The dark blue line displays the DISS as a function of time. Significant differences based on a non-parametric permutation test are shown in gray. The scale on the right indicates 1 minus *p*-value (range: 0.95–1). (e) The template maps resulting from the topographic pattern analysis are shown. During most of the post-stimulus period the same template map was observed in the group-average VEPs from both conditions. During two periods, different maps were observed as a function of experimental condition in the group-averaged VEPs. (f) The histograms depict the result of the spatial correlation fitting for both periods. During each period one of the two template maps fitted better to one of the experimental conditions.

responses of both conditions, during both the 100–130 ms and 270–310 ms post-stimulus time periods two distinct maps were identified that appeared to differentially account for each condition (see Fig. 2e). This pattern observed in the group-averaged VEPs was statistically assessed in the single-subject VEPs using a spatial-correlation fitting procedure. There was a significant condition × map interaction for both the 100–130 ms ($F_{(1,11)}=16.643$; $p=0.002$; $\eta_p^2=0.602$) and the 270–310 ms ($F_{(1,11)}=6.348$; $p=0.028$; $\eta_p^2=0.366$) post-stimulus periods. In both time periods one map predominated the responses to one of the conditions while the other predominated the responses to the other condition (see Fig. 2f).

The results to this point indicate that visual objects are incidentally discriminated according to the context of their initial presentation (i.e. in a unisensory or a meaningless multisensory context). Further they suggest that this differential discrimination takes place during early stages (~100 ms) of visual processing and engages different generator configurations.

Source estimations

Source estimations from both time periods were statistically analyzed to identify the likely brain regions contributing to these differential effects. During the early time period (100–130 ms post-stimulus onset) both conditions included prominent sources within the occipital, temporo-parietal, and frontal lobes (upper portion of Fig. 3a). The statistical contrast of these source estimations identified a single cluster of solution points meeting our criteria (lower portion of Fig. 3a; see Materials and methods for details of statistical criteria). This cluster was located within the right posterior insular cortex (BA 13; coordinates of maximal *t*-value = 41, −28, 14 mm) and extended into the superior temporal gyrus (STG; BA 21). Although it did not meet our spatial extent criterion, there was evidence of a small 6-node cluster (BA 18; coordinates of maximal *t*-value 17, −87, −9 mm) within the right inferior lateral occipital cortex that exhibited stronger activity in response to the V− than V+ condition. We mention this cluster here, given the previous evidence of differential activity within this region during this type of task (Murray et al., 2004, 2005). During the 270–310 ms post-stimulus period, both conditions again included prominent sources within the occipital, occipital-temporal and frontal lobes (upper portion of Fig. 3b). The statistical contrast again identified a single cluster meeting our significance criteria (lower portion of Fig. 3b). This cluster was found in the right STG (BA 22; coordinates of maximal *t*-value = 65, −17, 2 mm) and extended anteriorly to the middle temporal gyrus (MTG; BA 21). Source estimations in each of these clusters were then correlated with performance accuracy (Fig. 3c and d). There was no evidence for a significant correlation between source estimations over the 100–130 ms period and performance accuracy ($r=0.287$; $t_{(10)}=0.95$, $p=0.366$). By contrast, we observed a

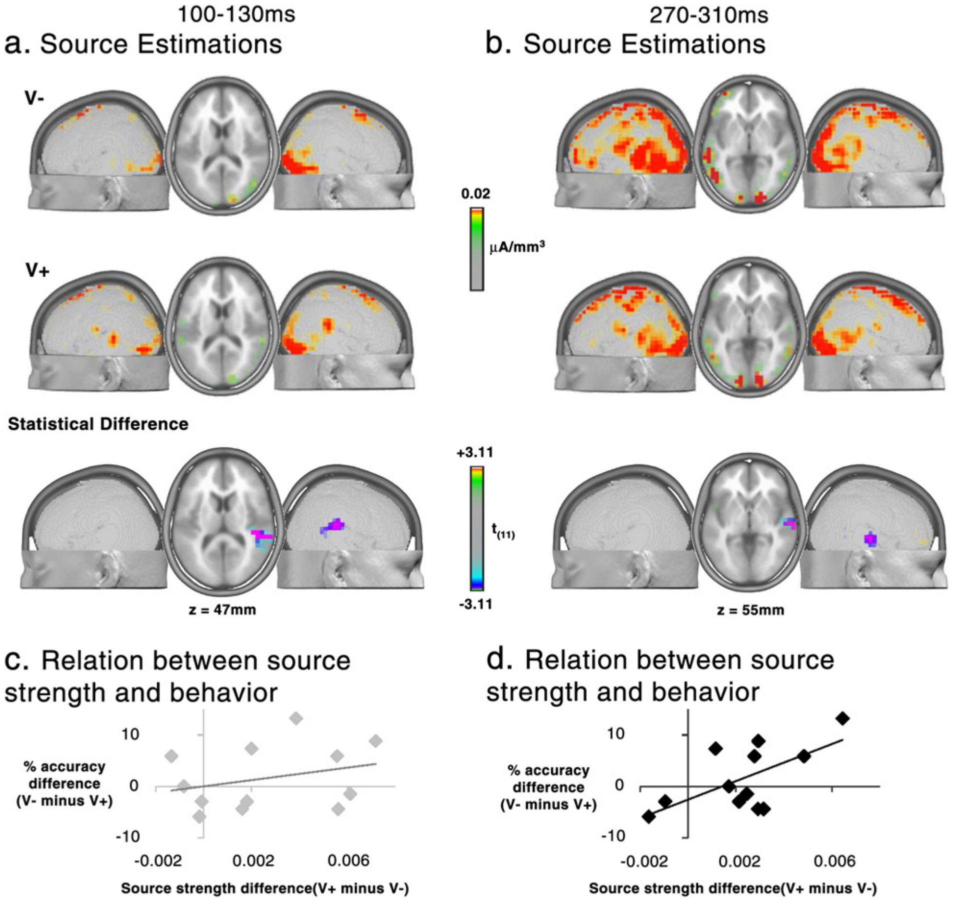


Fig. 3. The results of the source estimations are shown. The columns show the results from the 100–130 ms (a) and the 270–310 ms (b) post-stimulus onset time windows, respectively. For each time period the mean activation for each condition (V− and V+) are shown. Beneath, the *t*-value map of the contrast between conditions is shown. Note that the negative values (cooler colors) indicate that the clusters are more active in the V+ condition than in the V− condition. Over the 100–130 ms post-stimulus period the maximal *t*-value was located at 41, −28, 14 mm. Over the 270–310 ms post-stimulus period the maximal *t*-value was located at 65, −17, 2 mm. (c and d) Correlations between the observed cost in discrimination accuracy (*y*-axis, in [%]) in the V+ condition compared to the V− condition and the gain in source strength (predictor, *x*-axis, in [μA/mm^3]). Only the later period (270–310 ms) yielded a significant correlation ($r_{(10)}=0.627$; $p=0.029$) between source strength in the in the mSTG/MTG cluster and the cost in discrimination accuracy over subjects.

significant positive correlation between the V+ vs. V− difference in source strength within the mSTG/MTG cluster and the cost in discrimination accuracy (V− minus V+) ($r = 0.627$; $t_{(10)} = 2.55$, $p = 0.029$).

Discussion

The present study demonstrates that the discrimination of image repetitions is affected by whether or not the initial experience was multisensory versus solely visual. These effects occur even though the auditory information was both incidental for the task at hand and also devoid of any semantic content (i.e. is purely episodic in nature). This extends prior research showing this type of phenomenon following semantically congruent single-trial exposures to multisensory stimuli (Lehmann and Murray, 2005; Murray et al., 2004, 2005). We show that memory performance with visual objects is impaired by prior episodic multisensory pairings and cannot be explained by carry-over effects from the initial memory encoding or by an affect of differential attention (for a review see Murray and Sperdin, 2010). Electrical neuroimaging analyses revealed differential processing of repeated visual stimuli starting at ~100 ms post-stimulus onset and following from changes in the topography of the electric field at the scalp. Source estimations localized these effects to regions of the auditory cortex (and more subtly within the inferior LOC). Subsequent effects (270–310 ms) were again driven by topographic differences and we localized within auditory cortices of the STG/MTG. These later effects in turn positively correlated with behavior; individuals with larger differential STG/MTG responses exhibited larger costs in performance accuracy in indicating image repetitions. In what follows, we discuss our findings in terms of incidental effects of multisensory interactions on memory processes.

Our results show that single-trial multisensory learning occurs with the pairing of visual objects with meaningless sounds. Two factors contributing to this conclusion can be distinguished. First, semantic congruence between the senses is not a determinant factor. This conclusion was tentatively drawn by Lehmann and Murray (2005) who used a paradigm identical to that here, with the exception that the same sound (pure tone) was used for all multisensory pairings. This paradigm did not allow the authors to discern whether the effects derived from the episodic nature of the combination or instead from a potential “noising” of any single object representation (and its re-activation with image repetition) by the association of the same sound with multiple objects across the course of the experiment. The pairing of unique meaningless sounds in the present study allowed us to rule out such a “noising” mechanism and instead supports the role of episodic pairing in engendering distinct multisensory representations that can be incidentally accessed upon image repetition (something to which we return in our discussion of the electrical neuroimaging results). In this way, our findings somewhat challenge the hypothesis that single-trial multisensory interactions impact subsequent unisensory retrieval only in specific semantically congruent situations or when information across the senses is concordant. For example, some emphasize the role of either object familiarity (van der Linden et al., 2010) or ethological validity (von Kriegstein and Giraud, 2006) as principal factors in whether performance (and in some instances brain responses) would be affected by prior multisensory exposure.

Effects were reliably obtained following single-trial exposure and despite the task-irrelevance of the auditory stimulus. In this regard, effects of multisensory learning appear to occur in an incidental fashion. This conclusion originally came out from prior works using this paradigm (Lehmann and Murray, 2005; Murray et al., 2004, 2005). Similarly, Nyberg et al. (2000) showed that effects on auditory activity in response to repeated visual stimuli were not affected by the strength of the association between the senses (quantified by the number of repetitions of a given multisensory pairing). Another example of the effects of single-trial multisensory exposure on subsequent unisensory processing is the recalibration of auditory spatial representations by vision (Wozny and Shams, 2011). More generally, these findings contribute to a growing literature emphasizing the both short- and long- lasting effects of multisensory interactions on subsequent unisensory processing (Meylan and Murray, 2007; Naue et al., 2011; Shams et al., 2011). One important aspect that will require additional data is the duration over which single-trial multisensory learning persists in its effects on later visual processing. Resolving this will undoubtedly impact the clinical/developmental applicability of our findings. It will also be beneficial for future research to clarify the determinants of whether multisensory pairings result in memory performance enhancements or decrements. On the one hand, our collective findings with a continuous recognition task suggest that semantic pairings reliably enhance memory performance whereas episodic pairings result in decrements. By contrast, explicit memory tasks are less consistent, such that performance decrements were observed with semantically congruent pairings (e.g. Nyberg et al., 2000). Clarifying the source(s) of this discrepancy will be necessary to derive potential utility of the single-trial and incidental nature of these effects in clinical and developmental populations; something at the focus of ongoing research within our group. It will likewise be important to determine the potential contribution of active learning of the multisensory associations. Prior studies using active learning sessions with similar paradigms have yielded mixed effects on performance, with some observing enhancements (von Kriegstein and Giraud, 2006) and others decrements (Butler and James, 2011; Nyberg et al., 2000; Wheeler and Petersen, 2000) for multisensory vs. unisensory (or within-modal) pairings.

The results also support there being a general time window wherein incoming visual stimuli are first incidentally discriminated according to past multisensory experiences. In the present as well as our prior study (Murray et al., 2004) VEPs to the V+ and V− conditions first differed at ~100 ms post-stimulus onset. Moreover, and in both studies, this effect was due to modulations in the VEP topography that in turn must follow from changes in the underlying generator configuration. This timing suggests that relatively early stages of visual object processing are subject to influences from past (single-trial) multisensory exposure and raises the question of the precise visual processes underway at this latency. Because the critical contrast in this study was between two sub-types of image repetitions that themselves required the same task-related analysis and motor response (i.e. both had to be recognized as “old”), any difference would presumably reflect processes subsequent to and/or complementing an initial (and potentially coarse-level) object recognition stage. Rapid visual object discrimination has indeed been reported, with effects as early as 100 ms (e.g. Liu et al., 2009; Rousselet et al., 2002; Thorpe et al., 1996; VanRullen and Thorpe, 2001), and effects of image repetition have been documented as early as ~50 ms post-stimulus onset (e.g. Michel et al., 2004; Mouchetant-Rostaing et al., 2000; Murray et al., 2009; Seeck et al., 1997). The present results extend these findings to show that discrimination of repeated objects can be impacted by whether or not their prior exposure occurred in a multisensory vs. unisensory context that was also completely task-irrelevant (and in fact likely ignored by subjects as suggested by their slower reaction times to the AV than V condition). Moreover, we show that this context can be purely episodic and need not be linked to the processing of semantic congruence or in fact any semantic information contained within the auditory channel (see also Butler and James, 2011). It is also worth noting that all images from both the V+ and V− condition were highly familiar, suggesting that object familiarity is not driving the present effects (van der Linden et al., 2010). In addition to the effect over the 100–130 ms post-stimulus period, we also observed significant topographic VEP modulations over the 270–310 ms post-stimulus period the sources of which significantly correlated with performance metrics.

The electrical neuroimaging analyses applied in this study also identified the likely sources contributing to the incidental discrimination of image repetitions as well as the link between brain activity and performance accuracy. In terms of source estimations, significant differences were observed within the right auditory cortices (STG) and to a lesser degree with object-related cortices of the right LOC during the 100–130 ms post-stimulus period. Source activity was significantly stronger for the V+ than V− condition within the STG and significantly weaker for the V+ than V− condition within the LOC. Differential responses within the LOC in a similar paradigm involving semantically congruent multisensory pairings were observed by Murray et al. (2004, 2005). Interestingly, in these studies responses were consistently stronger for the V+ than V− condition, which is the opposite directionality of what was observed in the present study. One way of reconciling these differences is that stronger activity may be associated with the condition resulting in more accurate discrimination performance. Such being said, we would note that there was no evidence of a reliable correlation between differences in performance and differences in source strength within the LOC ($r_{(10)} = -0.287$; $p > 0.35$). Over the 270–310 ms post-stimulus period, significantly stronger activity in response to the V+ condition was observed within auditory cortices (MTG). Effects within the auditory cortices (STG at 100–130 ms and MTG and 270–310 ms) were not reliably observed in our prior investigations.

The present observation of effects within nominally auditory regions in response to visual stimuli may be linked to the episodic nature of the multisensory pairings and/or the impaired performance for the V+ vs. V− condition. Support for the former can be found in fMRI studies showing that activity within superior temporal regions is inversely related to the strength of the association between arbitrary auditory-visual multisensory stimulus combinations (e.g. Tanabe et al., 2005; see also Naghavi et al., 2011 for effects during encoding that are in turn linked with subsequent memory performance). Support for the latter possibility can be gleaned from the results of Nyberg et al. (2000). These authors observed stronger responses within auditory cortices in response to visually presented words that had been explicitly learned and remembered as being previously paired with their corresponding sounds. Interestingly and somewhat downplayed by Nyberg et al., there was an 8% performance decrement for words previously paired with sounds vs. unpaired words that was not statistically assessed. It may therefore be the case that the enhanced auditory activity reflects this performance cost and by extension failed retrieval processes. Another possibility is that these enhanced auditory cortex responses reflect memory processes linked to the reactivation of the initial context (i.e. redintegration; Hamilton, 1859; see also ; Wheeler and Petersen, 2000). As we found no evidence of a correlation between performance decrements and differential source activity within the STG over the 100–130 ms post-stimulus period, we can neither support nor refute any causal link between differential activity within the STG and performance decrements. By contrast, there was a significant correlation between performance decrements and differential source activity within the MTG over the 270–310 ms post-stimulus period. The more strongly this cluster was active for the V+ than the V− condition, the larger the performance difference was between the V− vs. V+ condition.

Despite this link between brain activity and performance, our use of episodic pairings makes an interpretation in terms of pure redintegration unlikely. Participants were engaged in a visually demanding task and showed no benefit from the auditory information, but rather performance impairments. Additionally, because the sounds were unique and meaningless, it is unlikely that participants were able to establish a distinct representation for each sound based on single-trial exposure. Instead, it may be the case that the auditory cortex is activated in an unspecific manner in response to images that had been paired with sounds; a form of echoic memory elicited by images. A more convincing demonstration of stimulus-specific redintegration is found in von Kriegstein and Giraud (2006), who showed enhanced activity within the fusiform face area in response to voices that had been associated with faces but not voices that had been paired with names. More recently, it has been shown that responses to visual stimuli within auditory cortices differ according to the semantic category of the visual stimulus (e.g. animals, musical instruments, etc.; Meyer et al., 2010). Although the role of mental imagery cannot be fully discounted, these types of data nonetheless suggest that low-level cortices may respond in a semi-selective manner to stimuli from other sensory modalities and in particular to stimuli with multisensory associations. An alternative account of the present results is that episodic multisensory experiences of the kind used here elicit recursive activity within auditory cortices that is disadvantageous for the discrimination of image repetitions. In this way, the present findings may offer a potential compromise in that varieties of redintegration processes may elicit distinct networks as a function of performance accuracy. That is, situations leading to improved performance (e.g. following semantically congruent exposures as in Murray et al., 2004, 2005) would recruit a network predominantly within lateral occipital cortices (and perhaps also intraparietal sulcus; e.g. Werner and Noppeney, 2010). By contrast, situations leading to impaired performance (e.g. following meaningless and episodic exposures) would recruit a network predominantly within auditory cortices (both STG and MTG). To more fully assess this possibility will require contrasting V+ and V− conditions as a function of performance accuracy; something that would require sufficient numbers of trials leading to inaccurate memory discrimination. This was unfortunately not the case in the present study.

Multiple temporal phases and/or levels of differential activity have been previously observed in studies of multisensory object discrimination (e.g. Diaconescu et al., 2011; Kayser, 2010; Molholm et al., 2004; Schneider et al., 2008; Werner and Noppeney, 2010) and multisensory interactions between simple stimuli (e.g. Cappe et al., 2010, 2012). In one model, Werner and Noppeney (2010) proposed that responses within primary auditory cortices are enhanced by (corresponding) visual stimuli independently of task-context and without a direct link to performance metrics. These effects were considered as a general mechanism of multisensory enhancement of stimulus salience without a direct relationship with behavior. Effects within the superior temporal sulcus, planum temporale, and inferior parietal sulcus were significantly correlated with performance gains on object classification and were thus considered by Werner and Noppeney (2010) to be involved in the integration of object-specific features. The extent to which our results can be grafted onto this model is not immediately forthcoming and will undoubtedly require further experimental data. This is particularly the case because responses within auditory cortices were not observed in our prior studies involving semantically congruent multisensory learning (Murray et al., 2004, 2005); something that the Werner and Noppeney (2010) model might otherwise have predicted. Likewise, the above model is based on interactions between externally presented and semantically congruent objects stimuli, whereas the present results reflect the influence of past multisensory experiences on current unisensory visual processing. Nonetheless, it is interesting to note the parallel between the effects obtained by Werner and Noppeney (2010) and our observations of behaviorally independent responses in relatively low-level auditory cortices at 100–130 ms followed by behaviorally coupled responses in higher-order auditory regions at 270–310 ms post-stimulus onset. One possibility is that prior multisensory exposures (at least those that are episodic in nature) manifest themselves on current visual processing in a manner similar to what occurs following the presentation of actual multisensory stimuli. Assuming such, the responses within auditory cortices (both STG and MTG) may serve to facilitate the differentiation of incoming visual stimuli even if ultimately to the detriment of memory performance accuracy.

Supplementary materials related to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2012.05.027.

Acknowledgments

This work has been supported by the Swiss National Science Foundation (grant 310030B-133136 to MMM). The Cartool software has been programmed by Denis Brunet, from the Functional Brain Mapping Laboratory, Geneva, Switzerland, and is supported by the EEG Brain Mapping Core of the Center for Biomedical Imaging (www.cibm.ch) of Geneva and Lausanne. It is freely available at: http://sites.google.com/site/fbmlab/cartool/cartooldownload. Jean-François Knebel provided input on the statistical analysis on spectra and luminance of our visual stimuli. Marzia De Lucia assisted with additional computer programming. Rosanna De Meo provided helpful comments on a preliminary version of this manuscript.

References

Baier, B., Kleinschmidt, A., Muller, N.G., 2006. Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. J. Neurosci. 26 (47), 12260–12265.

Brunet, D., Murray, M.M., Michel, C.M., 2011. Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput. Intell. Neurosci. 2011, 813–870.

Butler, A.J., James, K.H., 2011. Cross-modal versus within-modal recall: differences in behavioral and brain responses. Behav. Brain Res. 224, 387–396.

Calvert, G.A., 2004. The handbook of multisensory processes. MIT Press, Cambridge, MA.

Cappe, C., Thut, G., Romei, V., Murray, M.M., 2010. Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. J. Neurosci. 30 (38), 12572–12580.

Cappe, C., Thelen, A., Romei, V., Thut, G., Murray, M.M., 2012. Looming signals reveal synergistic principles of multisensory integration. J. Neurosci. 32 (4), 1171–1182.

De Lucia, M., Clarke, S., Murray, M.M., 2010a. A temporal hierarchy for conspecific vocalization discrimination in humans. J. Neurosci. 30 (33), 11210–11221.

De Lucia, M., Michel, C.M., Murray, M.M., 2010b. Comparing ICA-based and single-trial topographic ERP analyses. Brain Topogr. 23 (2), 119–127.

Grave de Peralta Menendez, R., Gonzalez Andino, S., Lantz, G., Michel, C.M., Landis, T., 2001. Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. Brain Topogr. 14 (2), 131–137.

Grave de Peralta Menendez, R., Murray, M.M., Michel, C.M., Martuzzi, R., Gonzalez Andino, S.L., 2004. Electrical neuroimaging based on biophysical constraints. Neuroimage 21 (2), 527–539.

Diaconescu, A.O., Alain, C., McIntosh, A.R., 2011. The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. J. Neurophysiol. 106, 2896–2909.

Gonzalez Andino, S.L., Murray, M.M., Foxe, J.J., de Peralta Menendez, R.G., 2005. How single-trial electrical neuroimaging contributes to multisensory research. Exp. Brain Res. 166 (3–4), 298–304.

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

Guthrie, D., Buchwald, J.S., 1991. Significance testing of difference potentials. Psychophysiology 28 (2), 240–244.

Hamilton, W., 1859. Lectures on Metaphysics and Logic. Gould & Lincoln, Boston.

Kayser, C., Logothetis, N.K., Panzeri, S., 2010. Visual enhancement of the information representation in auditory cortex. Curr. Bio. 20 (1), 19–24.

Kim, R.S., Seitz, A.R., Shams, L., 2008. Benefits of stimulus congruency for multisensory facilitation of visual learning. PLoS One 3 (1), e1532.

Knebel, J.F., Murray, M.M., 2012. Towards a resolution of conflicting models of illusory contour processing in humans. Neuroimage 59 (3), 2808–2817.

Knebel, J.F., Toepel, U., Hudry, J., le Coutre, J., Murray, M.M., 2008. Generating controlled image sets in cognitive neuroscience research. Brain Topogr. 20 (4), 284–289.

Knebel, J.F., Javitt, D.C., Murray, M.M., 2011. Impaired early visual response modulations to spatial information in chronic schizophrenia. Psychiatry Res. 193 (3), 168–176.

Koenig, T., Melie-Garcia, L., 2010. A method to determine the presence of averaged event-related fields using randomization tests. Brain Topogr. 23 (3), 233–242.

Lehmann, D., 1987. Principles of spatial analysis. In: Gevins, A.S., Remond, A. (Eds.), Handbook of electroencephalography and Clinical Neurophysiology. : Methods of Analysis of Brain Electrical and Magnetic Signals, vol. 1. Elsevier, Amsterdam, pp. 309–354.

Lehmann, S., Murray, M.M., 2005. The role of multisensory memories in unisensory object discrimination. Brain Res. Cogn. Brain Res. 24 (2), 326–334.

Lehmann, D., Skrandies, W., 1980. Reference-free identification of components of checkerboard evoked multichannel potential fields. Electroencephalogr. Clin. Neurophysiol. 48, 609–621.

Liu, H., Agam, Y., Madsen, J.R., Kreiman, G., 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62 (2), 281–290.

Martuzzi, R., Murray, M.M., Meuli, R.A., Thiran, J.P., Maeder, P.P., Michel, C.M., Grave de Peralta Menendez, R., Gonzalez Andino, S.L., 2009. Methods for determining frequency- and region-dependent relationships between estimated LFPs and BOLD responses in humans. J. Neurophysiol. 101 (1), 491–502.

Meyer, K., Kaplan, J.T., Essex, R., Webber, C., Damasio, H., Damasio, A., 2010. Predicting visual stimuli on the basis of activity in auditory cortices. Nat. Neurosci. 13 (6), 667–668.

Meylan, R.V., Murray, M.M., 2007. Auditory-visual multisensory interactions attenuate subsequent visual responses in humans. Neuroimage 35 (1), 244–254.

Michel, C.M., Murray, M.M., 2012. Towards the utilization of EEG as a brain imaging tool. Neuroimage, http://dx.doi.org/10.1016/j.neuroimage.2011.12.039.

Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., Grave de Peralta, R., 2004. EEG source imaging. Clin. Neurophysiol. 115 (10), 2195–2222.

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

Mouchetant-Rostaing, Y., Giard, M.H., Delpuech, C., Echallier, J.F., Pernier, J., 2000. Early signs of visual categorization for biological and non-biological stimuli in humans. Neuroreport 11 (11), 2521–2525.

Murray, M.M., 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, http://dx.doi.org/10.1007/978-1-4419-5615_11.

Murray, M.M., Michel, C.M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., 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., Wylie, G.R., 2005. The brain uses single-trial multisensory memories to discriminate without awareness. Neuroimage 27 (2), 473–478.

Murray, M.M., Brunet, D., Michel, C.M., 2008. Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr. 20 (4), 249–264.

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

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

Naue, N., Rach, S., Struber, D., Huster, R.J., Zaehle, T., Korner, U., Herrmann, C.S., 2011. Auditory event-related response in visual cortex modulates subsequent visual responses in humans. J. Neurosci. 31 (21), 7729–7736.

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

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

Perrin, F., Pernier, J., Bertrand, O., Giard, M.H., Echallier, J.F., 1987. Mapping of scalp potentials by surface spline interpolation. Electroencephalogr. Clin. Neurophysiol. 66 (1), 75–81.

Pourtois, G., Delplanque, S., Michel, C., Vuilleumier, P., 2008. Beyond conventional event-related brain potential (ERP): exploring the time-course of visual emotion processing using topographic and principal component analyses. Brain Topogr. 20 (4), 265–277.

Raij, T., Ahveninen, J., Lin, F.H., Witzel, T., Jääskeläinen, I.P., Letham, B., Israeli, E., Sahyoun, C., Vasios, C., Stufflebeam, S., Hämäläinen, M., Belliveau, J.W., 2010. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. Eur. J. Neurosci. 31 (10), 1772–1782.

Ranganath, C., Rainer, G., 2003. Neural mechanisms for detecting and remembering novel events. Nat. Rev. Neurosci. 4 (3), 193–202.

Rousselet, G.A., Fabre-Thorpe, M., Thorpe, S.J., 2002. Parallel processing in high-level categorization of natural images. Nat. Neurosci. 5 (7), 629–630.

Rugg, M.D., Johnson, J.D., Park, H., Uncapher, M.R., 2008. Encoding-retrieval overlap in human episodic memory: a functional neuroimaging perspective. Prog. Brain Res. 169, 339–352.

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), 1244–1254.

Seeck, M., Michel, C.M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., Landis, T., Schomer, D.L., 1997. Evidence for rapid face recognition from human scalp and intracranial electrodes. Neuroreport 8 (12), 2749–2754.

Shams, L., Seitz, A.R., 2008. Benefits of multisensory learning. Trends Cogn. Sci. 12 (11), 411–417.

Shams, L., Wozny, D.R., Kim, R., Seitz, A., 2011. Influences of multisensory experience on subsequent unisensory processing. Front. Psychol. 2 (264).

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

Spence, C., Nicholls, M.E., Driver, J., 2001. The cost of expecting events in the wrong sensory modality. Percept. Psychophys. 63 (2), 330–336.

Stein, B.E., Meredith, M.A., 1993. The merging of the senses. The MIT Press, Cambridge Mass.; London.

Talairach, J., Tournoux, P., 1988. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. G. Thieme Medical Publishers, Stuttgart; New York.

Tanabe, H.C., Honda, M., Sadato, N., 2005. Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. J. Neurosci. 25 (27), 6409–6418.

Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381 (6582), 520–522.

Toepel, U., Knebel, J.F., Hudry, J., le Coutre, J., Murray, M.M., 2009. The brain tracks the energetic value in food images. Neuroimage 44 (3), 967–974.

Tzovara, A., Murray, M.M., Michel, C.M., De Lucia, M., in press. A tutorial review of electrical neuroimaging from group-average to single-trial event-related potentials. Dev. Neuropsychol.

van der Linden, M., van Turennout, M., Indefrey, P., 2010. Formation of category representations in superior temporal sulcus. J. Cogn. Neurosci. 22 (6), 1270–1282.

VanRullen, R., Thorpe, S.J., 2001. Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. Perception 30 (6), 655–668.

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

Wheeler, M.E., Petersen, S.E., 2000. Memory's echo: vivid remembering reactivates sensory-specific cortex. Proc. Natl. Acad. Sci. U. S. A. 97 (20), 11125–11129.

Werner, S., Noppeney, U., 2010. Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J. Neurosci. 30 (7), 2662–2675.

Wozny, D.R., Shams, L., 2011. Computational characterization of visually induced auditory spatial adaptation. Front. Integr. Neurosci. 5 (75).

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

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

[1]The Functional Electrical Neuroimaging Laboratory, Department of Clinical Neurosciences, Vaudois University Hospital Center and University of Lausanne, Lausanne, Switzerland
[2] Department of Experimental Psychology, Ghent University, Ghent, Belgium
[3]Radiology Department, Vaudois University Hospital Center and University of Lausanne, Lausanne, Switzerland
[4]Electroencephalography Brain Mapping Core, Center for Biomedical Imaging of Lausanne and Geneva, 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
Antonia.Thelen@chuv.ch

Abstract pages	2
Number of text pages	29
Reference pages	6
Figure Legend pages	2
Number of figures	2
Number of Tables	3
Number of words	8387

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. The implication is that multisensory exposures establish distinct object representations that are accessible during later unisensory processing. The generalization of such effects to auditory object recognition has yet to be established and was the focus of the present study. First, we demonstrate that visual object recognition 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 recognition. Auditory object recognition 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 the unisensory object recognition 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 independent object processing mechanisms between modalities. We discuss these findings in terms of the *conceptual short term memory* and the *modality appropriateness* models. 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). Moreover, these interactions have been correlated with behavior (Cappe, Thelen, Romei, Thut, & Murray, 2012; Romei, Murray, Merabet, & Thut, 2007; van den Brink et al., 2013; Van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011). Cappe and colleagues found that increases in neuronal response strength at early latencies were positively correlated with multisensory gains in a motion discrimination task (Cappe et al. 2012). On the other hand, Romei and colleagues found correlations between multisensory events and the impact of a TMS pulse delivered over the occipital pole on auditory detection response speed (Romei et al. 2007). In another study, van der Burg et al (2011) showed auditory facilitation effects in a visual search task modulating activity within parieto-occipital cortices. In a subsequent study, van den Brink et al. (2013) found that this facilitation was predicted by the strength of anatomical connections between sub-cortical and cortical auditory structures. While these and similar data have an impact on models of sensory and perceptual processes at a given moment in time, others have focused on how multisensory interactions taking place at one point in time have an impact on subsequent unisensory processing. For example, Meylan and Murray (2007) showed that processing of visual stimuli was significantly attenuated within occipital cortices when the stimuli were preceded by a multisensory stimulus at intervals of 52ms. Others 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). Our group has instead 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). 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 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 the advances that the above studies have brought to our understanding of the impact of multisensory memories upon subsequent visual object recognition, one unresolved question concerns whether or not auditory object recognition also benefits from single-trial multisensory memories. To date, this question has been addressed in a paradigm entailing the recognition of voices and ringtones that had been learned during a separate session in multisensory contexts varying in whether the visual features were images (i.e. faces or pictures of phones) or written labels (first or brand names) (von Kriegstein & Giraud, 2006). These authors found that recognition accuracy for voices and ringtones was improved by learning in a multisensory context and that the magnitude of this improvement was larger for voice-face pairs versus both voice-name pairs and also ringtone-phone and ringtone-brand name pairs. Consequently, the results are suggestive of larger gains for ethologically valid pairings. The particular gain for voices learned as voice-face pairs was shown to involve enhanced activity (as measured by functional magnetic resonance imaging; fMRI) within the fusiform face area.

Though not the primary focus of their work, others have failed to observe a reliable benefit of multisensory learning on auditory recognition (Cohen, Horowitz, & Wolfe, 2009). Thus, it remains unclear how auditory object recognition is affected by previous exposure to multisensory events.

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. 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 a line drawing or a sound had been presented for the first time or not. The image and sound discrimination tasks were presented in separate experimental sessions and the stimuli themselves were pseudo-randomized within a block of trials. They were instructed to perform as quickly and as accurately as possible. Furthermore, each target object (irrespective of whether it was initially presented in a unisensory or multisensory context) was only repeated once throughout each experimental block (see Fig. 1 for a schematic representation of the paradigm).

In both recognition tasks, half of the initial presentations were auditory-visual multisensory parings, which were semantically congruent, incongruent or meaningless (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. Of the remaining stimuli, half of the previously multisensory stimuli were presented in a unisensory manner. The remaining initially unisensory stimuli were paired with either with a meaningful congruent, incongruent or meaningless sound (or image) where each variety of pairing was equally probable (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. In addition, we created a series of meaningless (scrambled) pictures from the 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 Fig. 1b). Since these stimuli were created in view of an EEG study, this procedure ensured that differences found between meaningful and meaningless visual object processing were in fact due to object recognition per se, rather than to differences due to low-level visual features (Knebel, Toepel, Hudry, le Coutre, & Murray, 2008).

The auditory objects were obtained from online libraries (http://www.freesound.org; www.findsounds.com) and were chosen to correspond to the objects presented in the line drawings. 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. 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).

Data Analysis

Accuracy and RT data were computed for each condition for every subject over all blocks of trials. Subjects completed two visual blocks. 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 constant level of accuracy, subjects completed three auditory blocks. Because there was no evidence for a learning effect between the first and the latter two auditory blocks, all three blocks were collapsed in the analyses. In order to directly compare performance between the visual and the auditory task, we computed the gain/cost index for each subject and for each condition. This index was calculated as the difference of accuracy (and response speed) upon repeated presentations of previously multisensory presentations minus the accuracy (and response speed) upon repeated presentations of previously unisensory condition. This resulted in a comparable measure of the impact of multisensory memory traces on subsequent auditory and visual object recognition, avoiding the caveat of introducing differences due to task-related performance differences.

Furthermore, we investigated the impact of the semantic relationship between auditory and visual objects. Thus, the following multisensory pairings were taken into account in the analyses reported here: unisensory repetition of previously unisensory presentations (V- and A- respectively) and unisensory repetitions of objects that had been coupled with an object in the other sensory modality (V+ and A+ respectively; c for previous, semantically congruent pairings; i for previous, semantically incongruent pairings; and m for previous, semantically 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 specifically directed at responding to three research questions. First, accuracy and reaction time (RT) data were analyzed independently for the visual and the auditory modalities, in order to facilitate comparisons with our previous results (Lehmann & Murray, 2005; Murray et al., 2004; Murray, Foxe, & Wylie, 2005; Thelen, Cappe, & Murray, 2012). Second, in order to directly compare the impact of multisensory memory traces upon subsequent visual and auditory object recognition, we analyzed gain/cost indices. Third, we sought to investigate possible correlations and/or predictive relationships between encoding and retrieval mechanisms as well as correlations between sensory modalities.

In the first step, the raw data (accuracy and RTs) from the visual and the auditory discrimination tasks were entered into separate one-way ANOVAs, with the factor of Encounter Context (previously unisensory, multisensory congruent, multisensory incongruent, multisensory meaningless presentations of unisensory repetitions). The RT data were then submitted to a 2x4 repeated-measures ANOVA with within-subjects factors of Task Modality (Visual and Auditory) and Encounter Context (same as reported before) in order to test for modality related differences (Yuval-Greenberg & Deouell, 2007, 2009).

Second, in order to directly compare the impact of multisensory memory traces on subsequent visual and auditory object recognition, gain/cost indices (accuracy and RTs) were submitted to a 2x3 ANOVA with factors of Task Modality and Multisensory Pairing similar to the raw data. We then entered the data into two separate one-way ANOVAs with the within-subject factor of Multisensory Pairing, in order to investigate the impact of multisensory memory traces on subsequent visual and auditory object recognition, separately. Post-hoc dependent t-tests were performed for all significant main effects and interactions and significance levels were corrected for multiple comparisons (Holm-Bonferroni method; Holm, 1979). Since 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 control for specific differences between multisensory pairings for the visual task modality. We also compared gain/cost indices for visual and auditory RTs into a 2x3 ANOVA. Lastly, gain/cost index matrices were compared to zero matrices in order to test for significant benefit/cost effects.

Third, we sought to investigate linear and/or predictive relationships in our data set. To this end, we first investigated whether effects observed during encoding were predictive of subjects' performance upon retrieval. In other words, we tested whether the cost in terms of RTs upon initial multisensory presentations as compared to unisensory presentations was predictive of the difference in discrimination accuracy observed upon repeated presentations. We did not expect to find significant correlations between these measures, since our previous results have repeatedly shown that these processes are dissociable (see Murray & Sperdin, 2010). Additionally, we tested whether behavioral outcome in one sensory modality was directly linked to and/or predictive of the behavioral outcome in the other modality. The existence of such a direct link between object recognition in the auditory and the visual modalities, would suggest that it relies on an, at least partially, common processing mechanism.

Results

Raw data

The one-way ANOVA for the accuracy data from the visual task revealed a main effect of Encounter Context ($F_{(3,23)}=7.990$; $p=0.001$; $\eta_p^2=0.510$)(Fig. 2a and Table 2a), indicating differential impact of prior encounter contexts upon subsequent unisensory retrieval. To further investigate these differences, we performed a series of 1-tailed t-tests, given the strong a-priori hypotheses. The Holm-Bonferroni method was applied to correct for multiple comparisons, though we report uncorrected p-values below. The results revealed that recognition accuracy for images that had been paired with a congruent sound (V+c) upon initial encounter was significantly higher than that for images initially paired with a semantically incongruent (V+i) (V+c vs. V+i = 94.87 ± 1.3% vs. 88.62 ± 1.9%; $t_{(25)}=4.554$; $p<0.001$) or meaningless sound (V+m) (V+c vs. V+m = 94.87 ± 1.3% vs. 89.74 ± 1.7%; $t_{(25)}=3.192$; $p=0.002$). Likewise, images that had been presented in a unisensory context upon initial presentations, were significantly better recognized than images previously coupled with incongruent sounds (V- vs. V+i = 92.52 ± 1% vs. 88.62 ± 1.9%; $t_{(25)}$=2.418; p=0.012). Although significant differences in recognition accuracy were observed between images that had been initially paired with semantically congruent sounds (V+c) and those that had been initially presented in a unisensory context(V-) (V- vs. V+c = 92.52 ± 1% vs. 94.87 ± 1.3%; $t_{(25)}$=-2.032; p=0.027), this difference did not survive the Holm-Bonferroni correction. Similarly, there was a significant difference in recognition accuracy between images that had been initially presented in a unisensory context (V-) and those that had been initially paired with a meaningless sound (V+m) (V- vs. V+m = 92.52 ± 1% vs. 89.74 ± 1.7%; $t_{(25)}$=2.057; p=0.025), which again did not survive the Holm-Bonferroni correction. Finally, the recognition accuracy for images that been presented either with a semantically incongruent or meaningless sound upon their initial encounter did not significantly differ (V+i vs. V+m = 88.62 ± 1.9% vs. 89.74 ± 1.7%; $t_{(25)}$=-0.595; p=0.557).

The one-way ANOVA on discrimination accuracy in the auditory modality also revealed a main effect of Encounter Context ($F_{(3,23)}$=21.685; p<0.001; η_p^2=0.739) (Fig. 2b and Table 2a), indicating that recognition accuracy for sounds is differentially affected by the previous encounter context. Post-hoc two-tailed t-tests were conducted to further investigate these differences and Holm-Bonferroni corrections were applied, and are reported below. The results revealed that previously unisensory presentations of sound objects (A-) were discriminated less accurately than sounds that had been paired with a congruent image upon initial encounter (A+c), but were more accurately discriminated than sounds that had been coupled to an incongruent image (A+i) (A- vs. A+c = 67.68 ± 2.9% vs. 74.02 ± 3%; $t_{(25)}$=-3.244; p=0.007; A- vs. A+i = 67.68 ± 2.9% vs. 56.53 ± 3.2%; $t_{(25)}$=6.258; p<0.001). These results suggest that the semantic contingency between sensory presentations impact the retrieval of implicitly formed sensory memory traces. Further, previously multisensory congruent presentations were discriminated more accurately than previously incongruent and meaningless (A+m) presentations (A+c vs. A+i = 74.02 ± 3% vs. 56.53 ± 3.2%; $t_{(25)}$=8.055; p<0.001; A+c vs. A+m = 74.02 ± 3% vs. 67.58 ± 2.57%; $t_{(25)}$=3.882; p=0.002). Again, these results suggest that rather than the mere presentations of a meaningful image upon initial encounter, only semantically congruent presentations are beneficial for subsequent recognition. Finally, previously incongruent presentations led to worse discrimination accuracy than initially meaningless presentations (A+i vs. A+m = 56.53 ± 3.2% vs. 67.58 ± 2.57%; $t_{(25)}$=-6.455; p<0.001), suggesting that only meaningful images coupled to initial sounds impact subsequent recognition accuracy of the latter. Surprisingly, there was no difference in recognition accuracy for sounds that had been coupled with a meaningless images upon initial encounter and sounds initially encountered in a unisensory context (A- vs. A+m = 67.68 ± 2.9% vs. 67.58 ± 2.57%; $t_{(25)}$=0.064; p=0.949), indicating that only meaningful images impact auditory sensory memory trace retrieval.

In terms of RTs, we first analyzed the data in the same way as the accuracy results and then 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 task-specific one-way ANOVAs did not reveal any difference in terms of response speed (Table 2a). On the other hand the 2x4 repeated-measures ANOVA 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). Post-hoc analyses revealed that subjects were generally faster in visual blocks as compared to auditory blocks (overall mean ± s.e.m.: visual blocks = 769 ± 16.36ms; vs. auditory blocks = 954 ± 11.5ms; p<0.001), similarly to what has been reported by Yuval-Greenberg and Deouell (2007; 2009). Additionally, subjects were slower when responding to previously incongruent and meaningless as compared to previously unisensory presentations (unisensory vs.

incongruent = 856 ± 12.51ms vs. 867 ± 12.34ms; $t_{(25)}=-2.508$; $p=0.019$; unisensory vs. meaningless = 856 ± 12.51ms vs. 864 ± 13.44ms; $t_{(25)}=-2.277$; $p=0.032$).

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 overall sensory modality related differences. In order to directly compare the impact of single-trial multisensory memories upon auditory and visual object recognition, we entered gain/cost indices from both tasks into a 2x3 repeated-measures ANOVA. There was no main effect of Task Modality when computing analyses on gain/cost indices (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 recognition. There was a main effect of Multisensory Pairing ($F_{(2,24)}=40.507$; $p<0.001$; $\eta_p^2=0.771$) and a significant interaction between the factors Task Modality and Multisensory Pairing ($F_{(2,24)}=11.548$; $p<0.001$; $\eta_p^2=0.490$).

Given this interaction, additional ANOVAs were conducted. The task-specific one-way ANOVA on the percentage of gain/cost for visual discriminations revealed a significant effect of Multisensory Pairing ($F_{(2,24)}=12.504$; $p<0.001$; $\eta_p^2=0.510$) (Fig. 2c). To further investigate the differential impact of semantic contingencies, we performed post-hoc 2-tailed t-tests, and applied Holm-Bonferroni corrections for multiple comparisons. The analyses 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; $t_{(25)}=3.192$; $p=0.008$). Additionally, 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 same one-way ANOVA on the gain/cost indices from the auditory task also revealed a significant effect of Multisensory Pairing ($F_{(2,24)}=32.252$; $p<0.001$; $\eta_p^2=0.729$) (Fig. 2d). Again, Post-hoc 2-tailed t-tests were performed and p-values were corrected for multiple comparisons. The results showed that the main effect of Multisensory Pairing was due to previously incongruent presentations, which led to significantly more important performance decreases than all other conditions (A+c vs. A+i = 6.35 ± 1.95% vs. -11.15 ± 1.78%; $t_{(25)}=8.054$; $p<0.001$; A+i vs. A+m = -11.15 ± 1.78% vs. -0.09 ± 1.44; $t_{(25)}=-6.454$; $p<0.001$). On the other hand, 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; $t_{(25)}=3.882$; $p=0.001$).

In order to ensure that the gain/cost indices were significantly different 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 multisensory pairings in the visual modality (V+c vs. Zero = 2.35 ± 1.16% vs. 0; $t_{(25)}=2.03$; $p=0.027$; V+i vs. Zero = -3.9 ± 1.61% vs. 0; $t_{(25)}=-2.419$; $p=0.012$; V+m vs. Zero = -2.77 ± 1.35% vs. 0; $t_{(25)}=-2.057$; $p=0.025$), suggesting that visual object recognition is generally affected by single-trial multisensory encounters. However, we found that only the A+m gain/cost index did not differ from zero within the auditory modality (A+c vs. Zero = 6.35 ± 1.95% vs. 0; $t_{(25)}=3.244$; $p=0.002$; A+i vs. Zero = -11.15 ± 1.78% vs. 0; $t_{(25)}=-6.257$; $p<0.001$; A+m vs. Zero = -0.09 ± 1.44% vs. 0; $t_{(25)}=-0.065$; $p=0.949$). Significant differences are marked with an asterisk on the error bars in Fig. 2c and 2d.

After having investigated gain/cost indices for discrimination accuracy, we submitted the gain/cost indices of response speed into the same 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 impact the response speed of subsequent unisensory object recognition.

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. Nonetheless, 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 recognition 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).

Finally, we assessed whether subjects' performance in one modality could predict the recognition accuracy in the other modality (A vs. V). For this, we tested whether discrimination accuracies were correlated across conditions and across modalities. The results suggest that there was no linear relationship either between initial and repeated presentations within a sensory modality or between modalities (Table 3). Interestingly, response accuracies were only significantly correlated within modalities (gray boxes in Table 3) and only within presentation type (initial vs. repeated).

Discussion

The present study demonstrated that auditory object recognition 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 independence of multisensory influences on visual and auditory object recognition. Further, since similar patterns of performance were observed for unisensory visual and auditory object recognition, we discuss the potential involvement of common memory processes and/or sources, proposing how the present findings are compatible with a more general auditory-visual object association framework, involving an occipital-temporal-frontal network.

The primary finding of this study is that auditory object recognition is affected by prior multisensory contexts (Fig. 2b and 2d). More precisely, recognition accuracy is 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 or in a unisensory manner upon their initial presentation. 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 recognition. Our results are 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. This discrepancy may be linked to the type of stimuli that were presented to subjects. While in our study, we presented sounds belonging to various object categories (living and man-made objects), von Kriegstein and Giraud (2006) investigated speaker recognition in light of either voice-face or voice-name associations. While voice-face associations are unique, voice-name associations are arbitrary in nature (many people carry the same name, but have a unique voice). Contrariwise, in the present study subjects saw schematized black-and-white drawings of category-representative objects and heard unique sounds. 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. Further, there has been 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, interpretability of the findings of von Kriegstein and Giraud may be limited to specific material (voice and faces) and not readily generalized.

The results of our visual recognition task showed that recognition accuracy was enhanced for images that had been paired with a congruent sound upon their initial encounter, whereas recognition was impaired for images that had been paired with an incongruent or a meaningless sound upon their initial encounter. This was compared to recognition for images only encountered visually (Fig. 2a and 2c). Consequently, we replicated our previous findings in visual object recognition being incidentally affected by past multisensory encounters (Fig. 2a and 2c) (Murray et al. 2004; 2005; Thelen et al. 2012). This further emphasizes the robustness of the impact of single-trial multisensory memories upon subsequent unisensory object recognition, while also addressing 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). Unisensory object recognition accuracy upon repeated encounters differed according to the semantic pairing upon initial presentations. Thus, there effects in the present study suggest the involvement of a perceptual memory mechanism rather than effects arising through mere attentional capture. Additionally, the semantic contingencies upon multisensory trials were manipulated on a trial-by-trial manner, rather than in a blocked context, again preventing any statistical prediction about upcoming trials. It has been argued that the magnitude of the congruency effect (i.e. faster reaction times (RTs) and higher accuracy upon congruent auditory-visual trails as compared to incongruent trials) is highly context-dependent (Botvinick, Cohen, & Carter, 2004; Egner & Hirsch, 2005; 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. 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). Such mechanisms are thought to occur in an automatic fashion, and subjects are usually unaware of such interference resolutions. Although these findings were always relative to simultaneously presented multisensory pairs, this 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. Consequently, eventual context-dependent conflict resolution mechanisms, which could have been differentially involved in our past studies, leading to enhanced recognition for initially congruent presentations as compared to initially incongruent presentations can be excluded.

The collective 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. Although these studies focused on sequential visual-only presentations, this can likely be transposed to simultaneous auditory-visual presentations in the present study. In fact, Intraub proposed that rapid presentations (125ms per image) 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.

The similar impact of multisensory memories upon subsequent visual and auditory object recognition suggests the involvement of partially common object processing related mechanisms. Prior studies examining neural correlates for this variety of effects would suggest that multisensory memories impact subsequent unisensory recognition differently according to whether semantically congruent, incongruent or meaningless stimulus associations are presented upon initial encounters (Murray et al. 2004; 2005; Naghavi et al. 2011; van der Linden et al. 2011; Thelen et al. 2012). Interestingly, despite differences in experimental design and/or whether these studies focus on encoding or retrieval-related brain activations, these studies consistently report the lateral occipital, superior and middle temporal as well as inferior frontal regions to be involved in multisensory object formation/unisensory object retrieval. In our previous research, we focused upon the incidental impact of multisensory presentations upon visual object recognition (Murray et al. 2004; 2005; Thelen et al. 2012). In those studies, we found that brain activity upon unisensory recognition differs according to the semantic contingencies of the initial coupling. While initial semantically congruent auditory-visual pairings lead to enhanced activity within lateral occipital regions upon recognition of these images, initial meaningless auditory-visual pairings lead to greater activity within superior and middle temporal areas upon retrieval. Naghavi and colleagues (2011) found similar activation upon initial auditory-visual presentations for the semantically congruent and incongruent couplings. Van der Linden and colleagues (2011) also found greater activations upon presentation of semantically incongruent presentations within the right middle temporal gyrus. These findings, together with the work of von Kriegstein and colleagues (2006; 2010) suggest that the underlying neural correlates for auditory-visual object association formation and the retrieval of the latter might involve lateral and ventral occipital, superior and middle temporal as well as inferior frontal regions. Nonetheless, the absence of explicit correlations between performance in the two modalities in the present study suggests that this higher-order object association network is differentially modulated by the sensory modality required to perform the task. Further research is necessary to investigate the brain dynamics and connectivity changes induced by task demands upon this network.

While unisensory object recognition is similarly affected by initially congruent (recognition enhancement) and incongruent (recognition impairment) multisensory pairings, we also observed some notable distinctions between sensory modalities. First, effects in one modality did not correlate with those in the other (Table 3). 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 the impact of explicit attentional orienting towards either a concurrent visual, auditory or auditory-visual stimulus stream upon rapid serial visual presentation (RSVP) elicited steady-state visual evoked potentials (SSVEP). SSVEPs amplitudes were significantly decreased when subjects had to pay attention to concurrent visual and auditory-visual stimuli. Interestingly, SSVEP amplitudes were less affected when subjects were asked to attend to concurrent auditory stimuli. These authors 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, interference from the semantically incongruent task-irrelevant stimuli was greater for subsequent auditory recognition as compared to visual object recognition (Fig. 2c and 2d). This, along with generally higher recognition accuracy in the visual task compared to the auditory task, can be explained in the light of the assumption that vision is the more appropriate and thus dominant sense in object recognition at least under the conditions used here (but see Suied & Viaud-Delmon, 2009; Welch & Warren, 1980; Yuval-Greenberg & Deouell, 2007, 2009). 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.

Conclusion

Taken together, the present study shows that memory traces formed after a single multisensory encounter impact subsequent auditory object recognition. To our knowledge this is the first account in the literature of such effects within the auditory modality. Also, we directly compare the impact of such single-trial encounters upon visual and auditory object recognition within the same group of subjects. We were able to exclude attentional explanations to the present impact, by addressing paradigmatic shortcomings in our previous studies. We discuss our results in the context of a model of the underlying neural generators involved in such object recognition, although differentiating between more general/supramodal and sensory-specific modulations of the before mentioned auditory-visual object association network.

References:

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.

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

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.

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

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.

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.

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., 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., 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., 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.

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.

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.

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; doi:10.1163/22134808-00002426

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. doi:10.1093/cercor/bht069

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.

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.

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

Figure Legends:

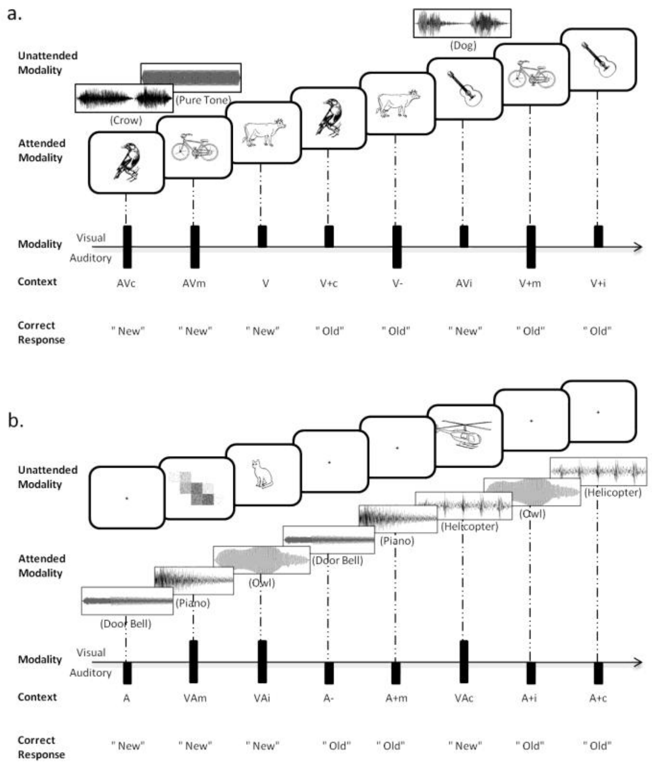


Figure 1

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.

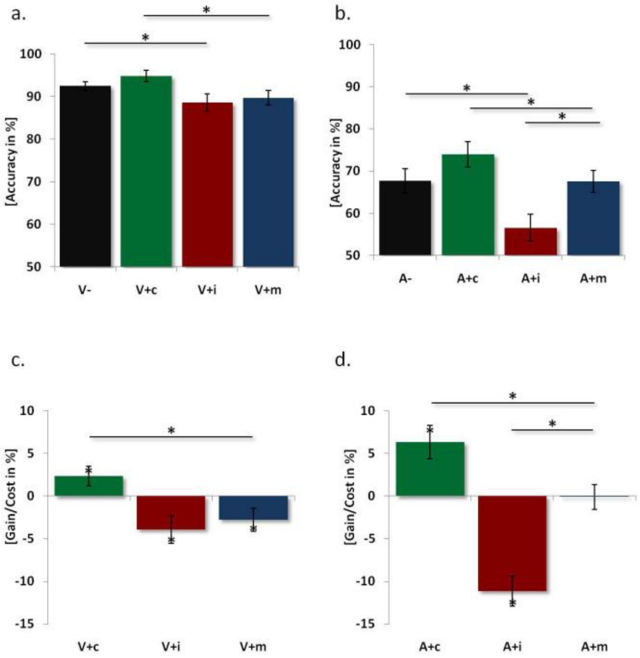
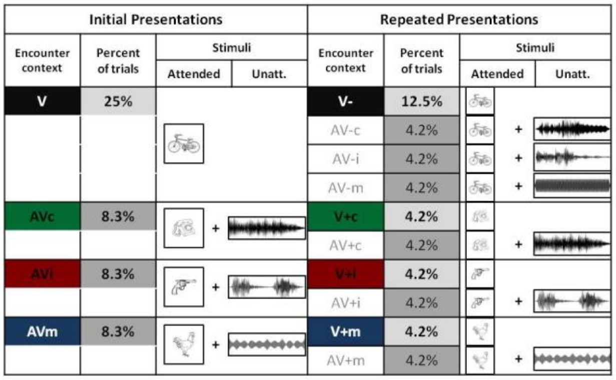


Figure 2

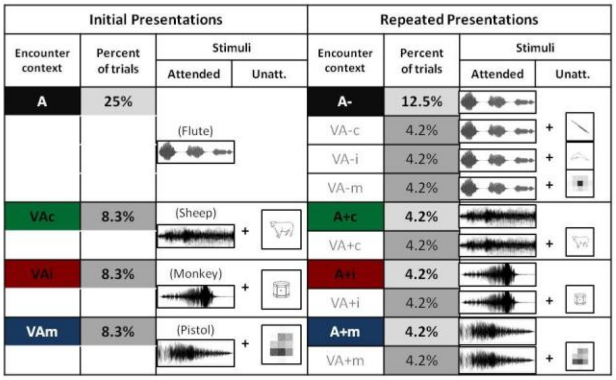
Figure. 2. Behavioral data from the visual and the auditory tasks (mean ± s.e.m.). **a.** and **b.** show the discrimination accuracy for the visual and auditory tasks, respectively. **c.** and **d.** show the performance gain/cost in percentage (V+/A+ minus V-/A-), for the visual and auditory tasks, 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 in **c.** and **d.**).

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%		
				AV-c	4.2%		+
				AV-i	4.2%		+
				AV-m	4.2%		+
AVc	8.3%		+	V+c	4.2%		
				AV+c	4.2%		+
AVi	8.3%		+	V+i	4.2%		
				AV+i	4.2%		+
AVm	8.3%		+	V+m	4.2%		
				AV+m	4.2%		+

b.

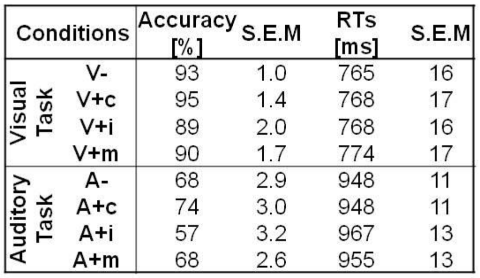


Initial Presentations				Repeated Presentations			
Encounter context	Percent of trials	Stimuli		Encounter context	Percent of trials	Stimuli	
		Attended	Unatt.			Attended	Unatt.
A	25%	(Flute)		A-	12.5%		
				VA-c	4.2%		+
				VA-i	4.2%		+
				VA-m	4.2%		+
VAc	8.3%	(Sheep)	+	A+c	4.2%		
				VA+c	4.2%		+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

Table 1

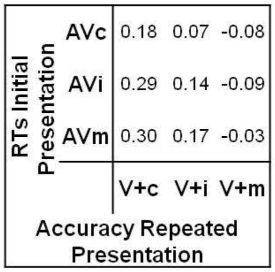
Table 1. Illustration of the probability of unisensory (light gray boxes) and multisensory stimuli (dark gray boxes) over trials within a block. The color code denotes initial encounter contexts (unisensory=black; congruent=green; incongruent=red; meaningless=blue). Only stimulus presentations discussed 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.

a.



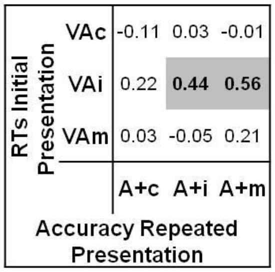
Conditions		Accuracy [%]	S.E.M	RTs [ms]	S.E.M
Visual Task	V-	93	1.0	765	16
	V+c	95	1.4	768	17
	V+i	89	2.0	768	16
	V+m	90	1.7	774	17
Auditory Task	A-	68	2.9	948	11
	A+c	74	3.0	948	11
	A+i	57	3.2	967	13
	A+m	68	2.6	955	13

b.



RTs Initial Presentation		Accuracy Repeated 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

c.



RTs Initial Presentation		Accuracy Repeated 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

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.

	V	AVc	AVi	AVm	V-	V+c	V+i	V+e	A	VAc	VAi	VAm	A-	A+c	A+i
AVc	**0.67**														
AVi	**0.64**	0.35													
AVm	**0.51**	0.33	**0.65**												
V-	0.03	0.09	-0.02	0.31											
V+c	0.12	0.30	-0.14	0.20	**0.55**										
V+i	0.33	**0.39**	0.14	0.37	**0.58**	**0.72**									
V+m	0.08	0.12	-0.08	0.01	**0.62**	**0.47**	**0.48**								
A	0.04	0.25	0.19	-0.08	-0.32	-0.26	-0.11	-0.24							
VAc	0.12	0.31	0.38	0.14	-0.30	-0.26	-0.25	-0.36	**0.69**						
VAi	0.07	0.13	0.20	-0.08	-0.25	-0.30	-0.05	-0.14	**0.79**	**0.67**					
VAm	0.15	0.37	0.13	0.09	-0.08	0.19	0.30	-0.16	**0.83**	**0.51**	**0.59**				
A-	-0.06	0.08	0.02	-0.20	0.02	0.03	0.00	0.11	0.07	0.13	0.11	-0.04			
A+c	0.19	0.17	0.24	-0.18	-0.14	0.02	0.11	0.15	0.17	0.14	0.15	0.03	**0.78**		
A+i	-0.07	0.02	-0.04	-0.22	-0.02	0.06	0.15	0.15	0.14	0.05	0.11	0.10	**0.83**	**0.76**	
A+m	0.00	0.08	-0.01	-0.31	-0.19	-0.01	0.05	0.07	0.08	-0.03	0.08	-0.03	**0.87**	**0.83**	**0.85**

Table 3. Correlation coefficient matrix for discrimination accuracy data over both tasks and all encounter contexts (initial unisensory V/A, initial multisensory AV/VA; repeated unisensory presentations of V-/A- =initially unisensory presentations; V+/A+ =initially multisensory presentations; c=congruent; i=incongruent; m=meaningless). The gray boxes indicate significant correlations ($-0.39 \leq r_{(26)} \geq +0.39$).