Exp Brain Res (2016) 234:1307–1323
DOI 10.1007/s00221-016-4590-z

RESEARCH ARTICLE

The COGs (context, object, and goals) in multisensory processing

Sanne ten Oever[1] · Vincenzo Romei[2] · Nienke van Atteveldt[1,3] · Salvador Soto-Faraco[4,5] · Micah M. Murray[6,7,8] · Pawel J. Matusz[6,9]

Received: 15 April 2015 / Accepted: 30 January 2016 / Published online: 1 March 2016

Abstract Our understanding of how perception operates in real-world environments has been substantially advanced by studying both multisensory processes and "top-down" control processes influencing sensory processing via activity from higher-order brain areas, such as attention, memory, and expectations. As the two topics have been traditionally studied separately, the mechanisms orchestrating real-world multisensory processing remain unclear. Past work has revealed that the observer's goals gate the influence of many multisensory processes on brain and behavioural responses, whereas some other multisensory processes might occur independently of these goals. Consequently, other forms of top-down control beyond goal dependence are necessary to explain the full range of multisensory effects currently reported at the brain and the cognitive level. These forms of control include sensitivity to stimulus context as well as the detection of matches (or lack thereof) between a multisensory stimulus and categorical attributes of naturalistic objects (e.g. tools, animals). In this review we discuss and integrate the existing findings that demonstrate the importance of such goal-, object- and context-based top-down control over multisensory processing. We then put forward a few principles emerging from this literature review with respect to the mechanisms underlying multisensory processing and discuss their possible broader implications.

Keywords Attention · Multisensory · Control · Object · Top-down · Bottom-up · Audio-visual · Brain mapping

Introduction

Research from the past 30 years has demonstrated a whole range of behavioural benefits engendered by integrating information across the senses (multisensory integration, MSI), including faster motor responses and facilitated object recognition in noisy environments (e.g. Stein 2012). In a separate and independent manner, studies

✉ Pawel J. Matusz
pawel.matusz@gmail.com; pawel.matusz@chuv.ch

1 Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

2 Department of Psychology, Centre for Brain Science, University of Essex, Colchester, UK

3 Department of Educational Neuroscience, Faculty of Psychology and Education and Institute LEARN!, VU University Amsterdam, Amsterdam, The Netherlands

4 Multisensory Research Group, Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain

5 Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

6 The Laboratory for Investigative Neurophysiology (The LINE), Neuropsychology and Neurorehabilitation Service and Department of Radiology, Centre Hospitalier Universitaire Vaudois (CHUV), University Hospital Center and University of Lausanne, BH7.081, rue du Bugnon 46, 1011 Lausanne, Switzerland

7 EEG Brain Mapping Core, Center for Biomedical Imaging (CIBM) of Lausanne and Geneva, Lausanne, Switzerland

8 Department of Ophthalmology, Jules-Gonin Eye Hospital, University of Lausanne, Lausanne, Switzerland

9 Attention, Brain, and Cognitive Development Group, Department of Experimental Psychology, University of Oxford, Oxford, UK

employing unisensory stimuli have been critically advancing our understanding of the nature and the importance of top-down mechanisms that control information processing. The *top-down* nature of these mechanisms lies in that they shape perceptual processing of new inputs by activating information stored in higher-order brain areas (e.g. Summerfield and Egner 2009).

Studies of top-down control have traditionally focused on attentional (i.e. goal-dependent) mechanisms, which promote the processing of stimuli or objects in the environment that are important to the current behavioural goals of the observer (e.g. Desimone and Duncan 1995). These mechanisms enhance the processing of stimuli appearing in task-relevant spatial locations and/or moments in time. These mechanisms likewise facilitate the processing of those stimuli whose attributes (e.g. colour red), feature dimensions (e.g. shape), or identity (e.g. a particular face) match the observer's goals. Simultaneously, it has been increasingly recognised in the literature that information processing is sensitive to other types of top-down processes, principally those gauged by the memory of past stimulation and one's expectations (see Fig. 1a for a summary of top-down control processes). This discovery has advanced our understanding of top-down control in several important ways. First, the role that attentional control mechanisms based on memory of objects and scenes play in naturalistic environments has been frequently investigated in recent years (see Nobre and Kastner 2014). Second, cognitive sciences have increasingly recognised the role of the brain in information processing as that of a proactive agent rather than that of a passive receiver. Approaches such as predictive coding and Bayesian models (Fries 2005; Schroeder et al. 2010; Summerfield and Egner 2009; Summerfield and de Lange 2014; Rohe and Noppeney 2015) have highlighted the importance of this form of top-down control, based on the context, or the "immediate situation in which the brain operates" (van Atteveldt et al. 2014a). The matching of features of particular objects (i.e. stimulus pairings related by meaning vs. arbitrarily linked) has been likewise shown to influence object processing, based on factors, such as the evolutionary relevance of some objects (e.g. Schiff et al. 1962; Maier et al. 2004; Bach et al. 2009; Matusz et al. 2015a). These advances, however, are only starting to impact our understanding of multisensory processing (e.g. Schroeder et al. 2010; Arnal and Giraud 2012; Fetsch et al. 2013; Talsma 2015).

This relative lack of systematic investigation of multisensory processing at the intersection with top-down control processes has led to persistent uncertainty as to whether the multisensory effects reported in the literature are a consequence of purely bottom-up, stimulus-driven mechanisms or, instead, are a result of a combination of stimulus-driven and of top-down mechanisms. This has recently been changing; an increasing number of studies has been investigating how MSI changes across paradigms and varying levels of task demands. Talsma et al. (2010) integrated that body of research within a framework that proposes a continuum in which different multisensory processes are more or less dependent on the current goals and/or available attentional resources of the observer. This framework significantly advanced our understanding of the mechanisms orchestrating multisensory processing. However, new challenging questions have recently been emerging from studies that investigate how the processing of multisensory stimuli is modulated by the stimulus context or by their match with attributes of naturalistic objects (tools, animals, etc.). How profound is the control of these object-based and context-based modulations over multisensory processing? If a stimulus represents a familiar multisensory object (e.g. a cat meowing), does your brain detect (and benefit from) this familiarity irrespective of what you are currently doing? Or would the top-down nature of such facilitation render it dependent on your goals? Are there multisensory processes whose occurrence is impervious to the context in which they are elicited?

Here, we review the existing literature, with a focus on studies employing audio-visual (AV) stimuli, which suggests that multisensory processes are influenced, to a differential degree, by goals as well as by the attributes of the eliciting stimuli (i.e. objects) and the context in which these stimuli occur. Defining what constitutes a multisensory process is a challenge. Some processes are linked to the matching of features of objects within the stimuli, e.g. when the brain detects that speech sounds and lip movements arrive from the same speaker (Fig. 1b). Other types of multisensory processes focus on the detection of congruence across low-level features, most notably, simultaneity. The existence of neurons sensitive to a simultaneous onset/offset of stimuli across the senses is supported by the pioneering work of Stein and colleagues (e.g. Meredith et al. 1987). As is discussed below, multisensory processes seem to depend to differing degrees also on the current goals as well as the stimulus context, with some processes perhaps occurring independently of all sources of top-down control (Fig. 1c). We conclude this review by proposing several emerging principles regarding how bottom-up multisensory processes interact with top-down control, and then discuss the possible broader implications of these emerging principles.

Top-down control of multisensory processes by goals

The influence of top-down attention, such as the observer's current goals, has been previously recognised as one principal source of control over multisensory processing (Talsma et al. 2010). At the same time, the framework proposed

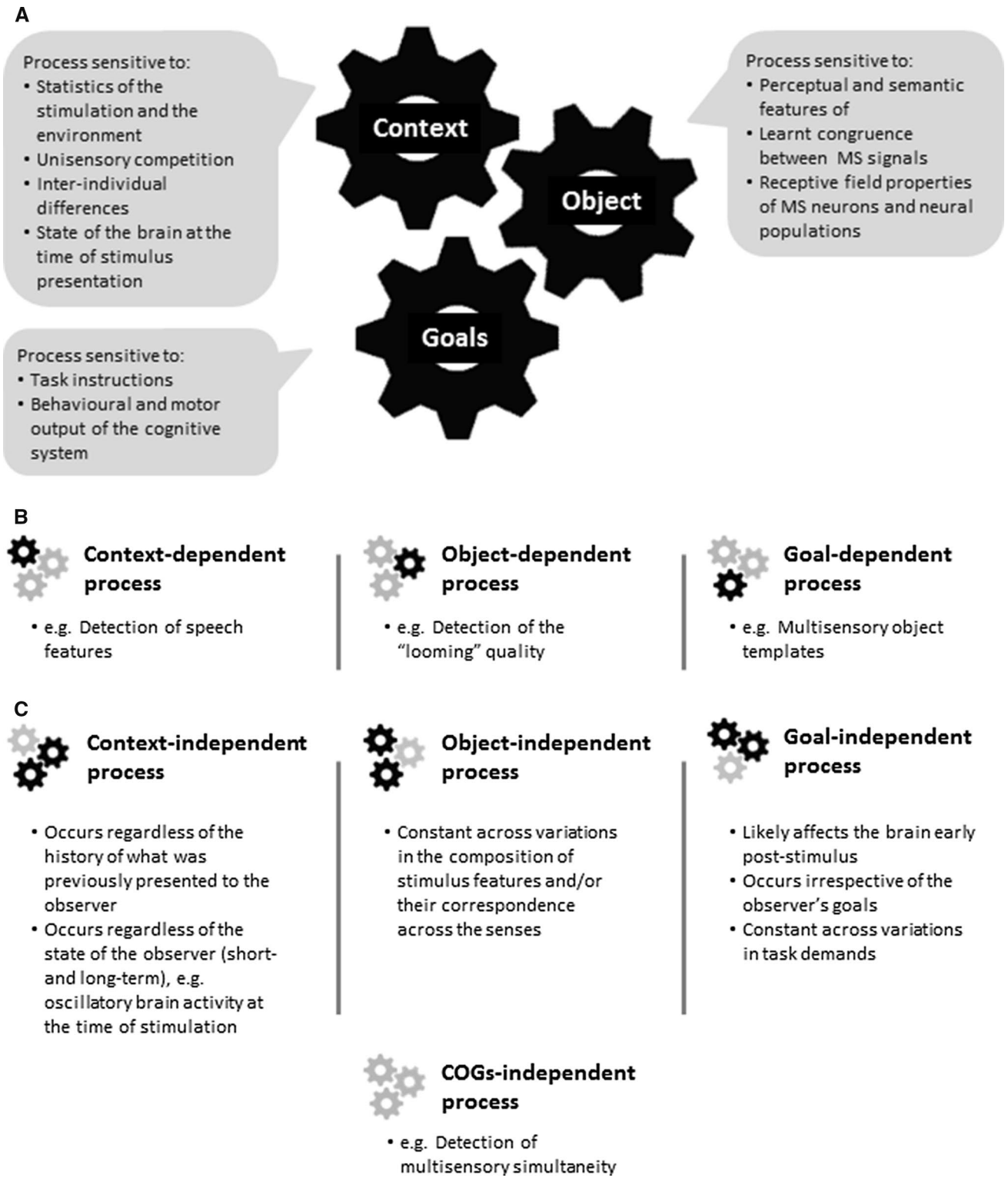

Fig. 1 **a** A schematic depiction of how multisensory processes might be defined by their relative dependence on each of the three types of top-down control. In *rounded boxes*, a summary of influences that a multisensory process should be sensitive to in order to be classified as dependent on context, object, and the observer's goals, respectively. **b** An example of a context-, object-, and goal-dependent multisensory process, respectively. **c** Attributes related to the dependence on context, object, and goals that should characterise a multisensory process for it to be classified as independent of each of these processes, respectively, together with an example of a multisensory process perhaps independent of all three top-down processes. *Note*: By "independence" we understand here independence of the *occurrence* of a particular multisensory process from top-down control

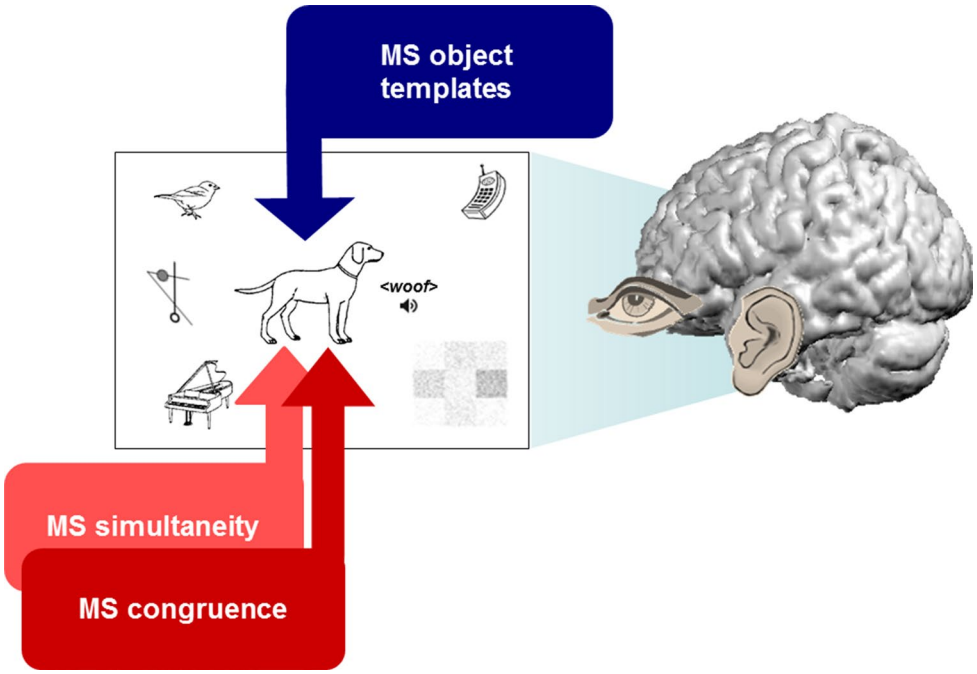

Fig. 2 The likely main types of multisensory processes that would jointly contribute to the perception of an exemplary naturalistic, task-relevant multisensory stimulus presented within a multi-stimulus unisensory visual display: goal-dependent multisensory object templates (in *blue*), detection of multisensory congruence of object-matching AV features (*dark red*), and detection of multisensory simultaneity (*lighter red*)

by Talsma et al. (2010) delineated the situations in which stimuli from different senses could interact independently of one's current goals. At least two types of multisensory processes likely influence the processing of stimuli in many of the currently employed perception paradigms (Fig. 2): those determined by goals and those independent of such control. Below we review the current evidence in support of the existence of both types of process.

MS processes whose presence is independent of one's current goals

The strongest evidence for the idea that MSI can occur independently of one's current goals would be provided by results demonstrating the presence of multisensory processes in response to stimuli defined by task-irrelevant features or feature dimensions and appearing in task-irrelevant locations in space or moments in time (Fig. 1; Desimone and Duncan 1995). The ability to affect information processing despite complete irrelevance has been revealed for simultaneous pairings of auditory and visual stimuli. In a multisensory adaptation of a visual attention task (the spatial cueing paradigm: Folk et al. 1992; Matusz and Eimer 2011), participants were instructed to search for targets defined by a visual feature (e.g. blue bars) and assess their orientation. Search arrays always followed displays with visual distracters (sets of four dots). The ability of the visual distracters to capture visual attention was measured by a difference in reaction times (RTs) to targets appearing in same versus different locations as those just occupied by the distracters (i.e. spatial cueing effects). There were stronger distraction effects on trials where visual distracters were paired with a tone. Notably, these multisensory enhancements were observed irrespective of the relevance of the visual feature (i.e. for distracters of the target colour and those of another, non-target colour, e.g. red) and despite the irrelevance of the sounds (i.e. they possessed no target-defining features). Additionally, neither visual nor auditory distracters provided any information about the location of the target, and both appeared at task-irrelevant moments in time. Thus, despite top-down attention likely suppressing the detection of the multisensory nature of the stimulus, the latter still influenced information processing (even if weakly). Other studies confirm that it is the simultaneity of the multisensory stimulus that mediates the goal independence of this effect (see "Early MSI as a hallmark of a bottom-up multisensory process" section).

At the same time, studies involving explicit judgements about multisensory simultaneity might be interpreted as suggesting that simultaneity detection is sensitive to one's current goals. In one such study, Stevenson and Wallace (2013) showed that the particular demands of the task (instructing participants to focus on the presence of simultaneity versus the order of auditory–visual presentations) modulated the tolerance in participants' judgements to inter-stimulus delays; this sensitivity generalised across both arbitrarily associated and semantically (i.e. related to identity) congruent pairings. However, in such tasks simultaneity is task relevant, which would likely render the reported effects a combination of bottom-up and goal-dependent influences. Notably, some evidence (e.g. Santangelo and Spence 2007) suggests that spatiotemporally

aligned multisensory stimuli trigger attention shifts in spatial tasks even when an additional, attention-demanding task is involved. However, the relevance of the stimuli in these studies was not fully eliminated. Such notwithstanding, while the detection of multisensory simultaneity by the brain is likely controlled by feature-, modality-, space-, and time-based goal-dependent mechanisms, it might continue to exert influence over stimulus processing in a goal-independent manner (even if weakly). In the next section we review neuroimaging evidence that also points to the goal independence of simultaneity detection by the brain.

Early MSI as a hallmark of a bottom-up multisensory process

Studies employing temporally resolved brain mapping methods, i.e. electroencephalography/magnetoencephalography (EEG/MEG) and transcranial magnetic stimulation (TMS), corroborate the idea that detection of multisensory simultaneity influences brain and behavioural responses despite top-down attentional suppression. Early event-related potential (ERP) studies in humans have demonstrated that detection of such simultaneity improves perception and that these benefits are frequently accompanied by brain responses whose amplitudes differ from the amplitude of the sum of brain responses to unisensory stimuli presented alone (Giard and Peronnet 1999; Molholm et al. 2002; Fort et al. 2002; Teder-Sälejärvi et al. 2002). These early-latency, i.e. occurring within the first 100 ms after stimulus onset, "nonlinear" brain responses have since been reported across a variety of experimental paradigms, from no-task setups through detection and discrimination tasks to multi-stimulus and multi-array paradigms that necessitate increased top-down attentional control (reviewed in De Meo et al. 2015; Murray et al. 2016). These early multisensory integration (eMSI; <100 ms post-stimulus) effects have been found for both task-relevant (alike the early ERP studies) and task-irrelevant MS stimuli. Thus, the eMSI could reflect a multisensory process relatively robust against (variations in) top-down goal-dependent control, as mere temporal coincidence between fully irrelevant AV stimuli suffices for it to occur.

Studies where the eMSI was directly compared across attended and unattended conditions are well suited to verify whether the eMSI can retain its presence despite top-down attentional suppression. For example, Talsma et al. (2007) instructed participants to detect rare target stimuli within either of two centrally presented streams of letter and digits or basic stimuli (beeps and flashes). The eMSI effects were found to be goal dependent, with respect to their quality; while attended AV stimuli elicited enhanced, "super-additive" eMSI, unattended AV pairings triggered attenuated, "sub-additive" (when compared to summed responses to attended unisensory stimuli) eMSI. In other studies, the eMSI were reported in response to multisensory stimuli appearing in irrelevant locations or moments in time (Talsma and Woldorff 2005; van der Burg et al. 2011). Thus, while both strength and quality of the eMSI seem under top-down attentional control, the presence of the eMSI in response to both attended and unattended stimuli suggests the brain's sensitivity to the presence of multisensory simultaneity might not be completely eliminated by top-down attention.

The idea that at least the presence of eMSI might be impervious to one's current goals is corroborated by the early (50–100 ms) latency of this process and by low-level sensory-perceptual cortices reported as its likely source (Cappe et al. 2010; Raij et al. 2000, 2010; De Meo et al. 2015; Murray et al. 2016). Existing results suggest a surprisingly extensive early crosstalk between inputs from different senses, where auditory-based responses within visual cortices co-occur with or even precede visually based responses to the same multisensory stimulus (animal models: Schroeder et al. 2004; Musacchia and Schroeder 2009; humans: Raij et al. 2010; Brang et al. 2015). Thus, information is transferred across different senses at latencies still considered as characterising the initial stimulus-driven brain activity, which is thought to be largely independent of top-down control (see, for example, Desimone and Duncan 1995; Lamme and Roelfsema 2000; Ding et al. 2014; for evidence for top-down control affecting brain responses >100 ms post-stimulus). In line with this idea, sounds can activate visual cortices <100 ms post-stimulus, control behaviour directly, and do so outside of the observer's awareness (e.g. Spierer et al. 2013; Sutherland et al. 2014). For example, visual cortex excitability, driven by a TMS pulse administered over occipital areas and as measured by phosphene perception, can be enhanced by concomitant presentation of sounds as early as 60–75 ms before the TMS stimulation onset (Romei et al. 2007).

Continuous flash suppression studies consistently demonstrate that faint unisensory stimuli (e.g. colour-changing discs) can be consciously perceived when paired with a coincident supra-threshold input into another sense (e.g. Palmer and Ramsey 2012; Alsius and Munhall 2013; Lunghi et al. 2014; Aller et al. 2015). Some studies report that top-down attention did not completely suppress MSI between irrelevant supra-threshold stimuli. However, the manipulations in these studies diverting attentional focus away from multisensory events might have not been fully effective. Thus, the reports of the brain detecting simultaneity within multisensory pairings involving unconsciously perceived inputs strengthen the idea that at least the detection *per se* (even if strongly attenuated) occurs independently of one's goals. Particularly compelling support for this argument comes from electrophysiological studies that

report the eMSI in anaesthetised animals. For example, Barth et al. (1995) observed the eMSI in multi-unit activity within low-level visual and auditory cortices (see Rowland and Stein 2007 for similar results in cats). As discussed in detail in "Top-down control of multisensory processes by stimulus context" section, expectation-based mechanisms might affect sensory-driven multisensory processing even at early latencies. Thus, the reports of eMSI in anaesthetised preparations, where pre-stimulus top-down modulations are absent, offer important support for the idea that the presence of simultaneity detection alone occurs independently of one's goals. Recent studies (e.g. Parise and Ernst 2015) are starting to shed light on the neural computations enabling the brain to detect temporal congruence across the senses in such, bottom-up manner.

To summarise, there is substantial evidence to suggest that multisensory stimuli can interact based on as little as simultaneity. Moreover, the eMSI reflect the brain's sensitivity to this simultaneity (see also "General differences in multisensory processing related to object matching" and "MS processes whose occurrence is independent of stimulus context" sections) and also occur via a goal-independent, bottom-up mechanism (Fig. 1c). This notwithstanding, the pursuit of bottom-up multisensory processing as well as the effective experimental setups that were aimed at reducing goal-based (unisensory) control left us with little knowledge of goal-dependent processes that are multisensory in nature but which likely affect processing of multisensory stimuli in real-world environments (Fig. 2). Next, we summarise findings on such multisensory processes.

Multisensory processes whose occurrence depends on goals

A growing number of studies reveal how the task relevance of features of multisensory pairings enhances stimulus processing of the latter (e.g. Iordanescu et al. 2009; van Ee et al. 2009; Orchard-Mills et al. 2013a, b; Nardo et al. 2014; Masterberdino et al. 2015). While these findings are in line with the influence of feature-based unisensory attention (Desimone and Duncan 1995), whether this particular explanation applies to multisensory situations remains unclear. To investigate this possibility, Matusz and Eimer (2013) employed multi-stimulus visual displays and instructed participants to search for targets defined by a visual feature alone (e.g. blue bars) or by an arbitrary conjunction of visual and auditory features (e.g. blue bars accompanied by high-pitch tones). The search array was always preceded by unisensory visual distracters matching the target-defining feature value (e.g. its colour). Across three experiments, the same, identical unisensory distracters captured attention reliably during visual search but showed a reduced or completely eliminated ability to do so during multisensory search. These attenuations, visible in both behaviour (RTs spatial cueing effects) and the ERPs (the N2pc component; Luck and Hillyard 1994), can be explained by goal-dependent control mechanisms suppressing the processing of visual distracters in conditions where they did not *fully* match the representation of the (multisensory) target object (cf. top-down object templates; Duncan and Humphreys 1989). These "arbitrary multisensory object templates" will likely facilitate the processing of any, otherwise arbitrary, multisensory pairings if the current task renders them task-relevant "objects" (e.g. task-determined colour–pitch combinations). The existence of such a *par excellence* goal-dependent multisensory process (Fig. 1a, b), triggered by unfamiliar but task-relevant combinations of AV features, seems intuitive: In real-world environments, we routinely search for objects defined *ad hoc*, by arbitrary multisensory associations, e.g. when searching for our flatmate's ringing mobile in the shared living room. In such situations, distraction by every object matching a visual or auditory feature of the phone would be highly disruptive to our behaviour.

Top-down control of multisensory processes by object matching

As already suggested, some multisensory processes might be triggered only when all features of the eliciting multisensory stimulus match a particular object. We define "object" broadly, in line with previous proposals: "≪something [oftentimes] material that may be perceived by the senses≫ (Merriam-Webster online dictionary) (…) including not only concrete objects such as vehicles, tools, and persons, but also more abstract objects such as letters or speech with its accompanying lip movements" (Amedi et al. 2005, pp. 559–560; see also Fig. 1a). Early studies in the area have typically shown that congruent (e.g. in colour: a visual form paired with a verbal colour label) AV pairings elicit faster responses than unisensory stimuli, which in turn trigger responses faster than incongruent multisensory stimuli. Unisensory stimuli with redundant attributes show no similar response benefits (e.g. Laurienti et al. 2004). Such findings have suggested that a crossmodal match of *semantic* features may be one further general "factor" determining MSI (alongside low-level spatiotemporal factors). In the following section we discuss several points that portray a more nuanced view on the interplay between MS processing and object matching.

General differences in multisensory processing related to object matching

Processes whose presence depends on matching features (traditionally, semantic) of an object and those whose

presence is independent of such (Fig. 1b, c) differ in their brain mechanisms, in terms of both where the effects occur and when in time they unfold. Functional magnetic resonance imaging (fMRI) studies typically point to a fronto-temporal network as mediating MSI of complex, meaningful AV object stimuli, with frontal cortices (the inferior and dorsolateral prefrontal cortex) typically engaged by incongruent and/or unfamiliar AV associations (for a comprehensive review, see Doehrmann and Naumer 2008). The temporal cortex, especially the bisensory-multisensory superior temporal cortex (bmSTC) subregion, has been repeatedly implicated in the processing of naturalistic multisensory stimuli (e.g. Beauchamp et al. 2004; Stevenson et al. 2010; see Perrodin et al. 2015 for a recent review indicating anterior temporal cortices as supporting identity-focused processes) and in driving the associated behavioural benefits (Werner and Noppeney 2010a, b). These findings can be taken as evidence for the existence of a semantic congruence area supporting MSI. However, naturalistic multisensory stimuli frequently activate the STC together with other areas, such as the planum temporale (speech/script, van Atteveldt et al. 2004) or the inferior parietal sulcus (IPS; tools/animals, Werner and Noppeney 2010a). Thus, substrates for separate object-dependent multisensory processes likewise seem to exist. Perceptual tasks involving arbitrarily linked multisensory stimuli seem to engage a somewhat different set of brain areas: the STC (likely the synchrony-STC subregion; Stevenson et al. 2010) and primary visual and auditory cortices (Martuzzi et al. 2007). The STC and low-level cortices have also been reported to be functionally coupled (Cappe et al. 2010; Werner and Noppeney 2010b).

The temporal precision of electrophysiological brain mapping methods has revealed that multisensory processes occurring independently of object matching modulate brain responses at earlier latencies compared to processes dependent on this match. The idea that the eMSI (detailed in "Early MSI as a hallmark of a bottom-up multisensory process" section) reflects the brain's sensitivity to multisensory simultaneity is supported by the wide range of arbitrarily linked multisensory stimulus pairings shown to trigger this process in both humans and non-human primates (NHPs) (Fig. 1b, c; see De Meo et al. 2015; Murray et al. 2016). Furthermore, the eMSI is likewise found in response to naturalistic objects (tools/animals in humans, conspecific communication signals in NHPs), with no evidence of modulation by stimulus feature congruence or task (Ghazanfar et al. 2005; Kayser et al. 2008; Diaconescu et al. 2011; cf. "Early MSI as a hallmark of a bottom-up multisensory process" section). Contrastingly, differential brain responses between congruent and incongruent multisensory pairings are typically observed only after 100 ms post-stimulus. For example, task-irrelevant stimuli representing naturalistic movement (e.g. clips of water drops; Senkowski et al. 2007) or speech (van Wassenhove et al. 2005) trigger earlier and/or nonlinear ERPs starting at 120 ms post-stimulus (albeit the latter effects seem to be of anticipatory, not integrative, nature; Stekelenburg and Vroomen 2007, Expt.3) when presented in multisensory contexts. Most evidence suggests that the brain is sensitive to (in)congruence of object semantic features in multisensory stimuli starting at approximately 150–200 ms post-stimulus (evoked responses: Raij et al. 2000; Molholm et al. 2004; Diaconescu et al. 2011; but see Naci et al. 2012 for effects <100 ms; induced responses: Yuval-Greenberg and Deouell 2007).

Task-based effects

In the majority of the studies discussed in "General differences in multisensory processing related to object matching" section, stimulus congruence was relevant to the task, often being at its very focus. The presence of a task, such as multisensory congruence matching, seems to "override" the brain network activated otherwise in no-task situations as shown for AV script (van Atteveldt et al. 2007). When there is no explicit task, congruent pairs are likely assigned the highest relevance (see "Stimulus-based effects" section). Task instructions can seemingly overrule this default relevance assignment and render congruent and incongruent letter–sound pairs equally relevant to the task, as evidenced by the comparable STC activations found by van Atteveldt et al. (2007). Similarly, the particular task choice can have a dramatic effect on the processing of naturalistic multisensory stimuli. For example, instructing participants to detect versus categorise naturalistic stimuli modulates how MSI transpires within low-level visual and auditory cortices (van Atteveldt et al. 2014b). As already discussed in the "MS processes whose presence is independent of one's current goals" section, observers' tolerance to inter-stimulus delays during judgements of simultaneity of AV stimuli is modulated by the particular demands of the simultaneity-judgement task (Stevenson and Wallace 2013; but this tolerance likewise depends on the stimulus category, see below). Collectively, these findings demonstrate that the role of matching multisensory stimulus features can be better understood in situations where the congruence of object features is task irrelevant (Masterberdino et al. 2015; Santangelo et al. 2015).

Stimulus-based effects

Do the additional activations from higher-order areas that characterise the processing of naturalistic multisensory stimuli render these stimuli more or less impervious to top-down attentional control, compared to arbitrary multisensory pairings? The detection of feature congruence in speech is one multisensory process that might be useful to

answer this question. Auditory and visual signals produced by the speaker are intrinsically related by the common communication source, and listeners capitalise on these correlations during perception, already from an early age (reviewed in Soto-Faraco et al. 2012). Some even portray speech processing as altogether distinct from the processing of other naturalistic objects (e.g. Belin et al. 2000; see Tuomainen et al. 2005 for this argument applied to the particular case of audio-visual speech processing). In a study of Stevenson and Wallace (2013), the temporal window within which participants perceived AV stimuli as synchronous was larger for speech fragments than for tools/animals or arbitrary stimulus pairings. This effect might be due to the inherent complexity of speech that incurs longer processing, which in turn might render it robust against larger stimulus-onset disparities. Do these qualities suffice for multisensory speech congruence to be detected and continue influencing perception despite suppression by top-down attentional control?

The McGurk illusion (i.e. perceiving a novel auditory syllable from mismatching auditory and visual syllables; McGurk and MacDonald 1976) has been found to be attenuated, albeit still present, when the observer's attention was diverted away from the (irrelevant) McGurk stimuli and onto a concurrent, attention-demanding task (Alsius et al. 2005, 2007). However, when ERPs were recorded to McGurk stimuli in such dual-task contexts, the typical reduction of ERP latencies to AV stimuli present under full attention (e.g. van Wassenhove et al. 2005) were found to be substantially reduced (or even eliminated; Alsius et al. 2014). Particularly strong support for the goal-*dependent* nature of the mechanisms orchestrating the mere *detection* of multisensory speech congruence is provided by studies employing multi-stimulus unisensory displays. Multi-stimulus setups necessitate stronger goal-based control than single-stimulus setups (Desimone and Duncan 1995). Consistent with this idea, peripheral visual distracters lose their ability to interfere with search carried out within a central array as the number of search items increases (reviewed in Lavie 2010). Similarly, in multi-speaker visual setups, the efficiency of locating a congruent AV face–voice match was found to decrease as the number of relevant talking faces/voices increased (e.g. Alsius and Soto-Faraco 2011; see also Fernández et al. 2015; see Iordanescu et al. 2009 for comparable findings during search for tools and animals). This detection is based on goals insofar as new goals are required when the participant must perform the task with increasing number of distracters. If multisensory speech congruence was detected independently of one's goals, the sounds should be effortlessly bound with the corresponding mouth in the array, making the multisensory pairing "pop out" of the array and reveal its location to the participants irrespective of the number of other faces.

The processing of script, an object category closely related to speech (Dehaene and Cohen 2007; van Atteveldt and Ansari 2014), has likewise been tested with respect to how strongly it depends on one's current goals. Studies employing multi-stimulus displays have suggested that multisensory speech/script congruence is detected for stimuli in unattended locations independently of the level of task demands (Matusz et al. 2015b, Supplemental Expt.). Other studies are inconsistent with these findings. While early-latency, automatic brain processes (mismatch negativity, MMN) are modulated by the detection of script congruence even within task-irrelevant AV stimuli, this ability develops only after years of reading instruction (Froyen et al. 2009; cf. the early life onset of speech congruence detection; Soto-Faraco et al. 2012). Additionally, as already mentioned, the brain networks supporting multisensory script processing are seemingly determined by task (van Atteveldt et al. 2007). Thus, goals seem to exert a stronger influence over the detection of multisensory script than of speech congruence.

Recently, the importance of the observer's goals has likewise been tested for the detection of correspondences across lower-level, perceptual features (Mondloch and Maurer 2004), such as those between visual size/elevation and auditory pitch/intensity (reviewed in, for example, Spence and Deroy 2013). Multisensory congruence across perceptual stimulus attributes seems to be detected and to influence behaviour predominantly when one or both attributes match the current goals of the observer (Fig. 1a, b). On the one hand, judgements of sound localisation (left/right) appear more erroneous if visual stimuli accompanying the sounds match the "intuitive" pitch–size association (Parise and Spence 2009). On the other hand, when a similar task was used in a joint ERP-TMS study (Bien et al. 2012), the auditory and visual stimuli interacted quite late, i.e. 250 ms post-stimulus (cf. <100 ms latencies of the eMSI; De Meo et al. 2015). Likewise, the search for bars changing in their brightness (from dim to bright) improves with presence of high-pitch sounds (i.e. ones that "correspond" with bright flashes), but only if participants are aware of the correspondence or if the task demands are low, i.e. the search occurs in small-size arrays (Klapetek et al. 2012, Expt. 2–3).

The few studies that directly compared the detection of multisensory congruence (typically between semantic features) and of multisensory simultaneity suggest that, when these two qualities are task irrelevant, combined congruence and simultaneity modulates both brain and behavioural responses more strongly compared to simultaneity alone. This was demonstrated, for example, by how strongly congruence and simultaneity affect memory when task irrelevant. In a continuous unisensory "old/new" task ("did you see this image before?"), naturalistic unisensory objects (e.g. a dog) are categorised as repeated more

accurately if they are initially paired with congruent (e.g. a bark) stimuli in the other, irrelevant sense (Murray et al. 2004, 2005; Matusz et al. 2015a; Thelen et al. 2015). When the same unisensory stimuli are initially paired with simple stimuli (e.g. a pure tone) in the other sense, memory benefits for repeated objects are still found, but only in the individuals exhibiting stronger responses to initial presentations of multisensory stimuli (Thelen et al. 2014).

While more research is required here, the boost in the processing of multisensory stimuli matching an object might be typically driven by simultaneous co-activations of over-learnt multisensory associations that trigger additional feedback from higher-order brain areas to lower-order brain areas (e.g. van Atteveldt et al. 2007). These processing enhancements might likewise arise from the expectations that congruent crossmodal signals likely share a common source (Vatakis and Spence 2007). The co-activation and expectation mechanisms are not mutually exclusive (Fig. 1a). Notably, the detection of congruence across certain perceptual features of multisensory stimuli might sometimes have a more hardwired nature, which is possibly based on the properties of receptive fields of multisensory neurons (Fig. 1a). The detection of the "looming" (i.e. approaching) quality within multisensory stimuli results in stronger MSI across autonomic, behavioural, and neural responses when compared to stationary multisensory stimuli (Cappe et al. 2009; Spierer et al. 2013; Tyll et al. 2013; Cecere et al. 2014; Finisguerra et al. 2015). For example, visual "go/nogo" movement detection judgements are faster if the visual stimuli are accompanied by irrelevant looming sounds, compared to stationary or receding sounds, and these selective benefits are linked to early brain response modulations within temporal, parietal, and occipital cortices as well as, notably, the amygdala (Cappe et al. 2009, 2012). These selective benefits also develop early in life in humans (Walker-Andrews and Lennon 1985) and are present even in insects (Rind and Simmons 1999).

To return to the question posed at the beginning of this section, the top-down nature of the mechanisms underlying multisensory processes triggered by the detection of matches with specific object categories might be responsible for the effective dependence of these processes on top-down attentional control (Iordanescu et al. 2009; Fairhall and Macaluso 2009; Fernández et al. 2015). Simultaneously, the detection of some forms of congruence, e.g. looming, within multisensory stimuli might possibly occur independently of the current goals.

Top-down control of multisensory processes by stimulus context

Compared to goal- and object-based control, the investigation of context-based top-down control over multisensory processing seems less straightforward. This difficulty stems in part from the broadness of control mechanisms categorised as context-based in traditional, unisensory research. Such mechanisms range from fine-grained changes in stimulus features (e.g. their colour or position; Bar 2004) to the observer's external or internal states (e.g. studying specific material in a particular setting; Baddeley et al. 2009). We define context here as the "immediate situation in which the brain operates" (van Atteveldt et al. 2014a). Naturally, this context can extend backwards in time across multiple timescales. In the following sections, we review the evidence demonstrating the importance of stimulus regularities, expectations as well as past experiences of the observer as sources of context-based top-down control over current multisensory processing (Fig. 1a). As already discussed in detail elsewhere (van Atteveldt et al. 2014a), we contend that the large majority of the studied multisensory processes is modulated by some type of context-based control. In addition, we review additional evidence suggesting that some multisensory processes might be more robust against context-based top-down control than others.

Stimulus statistics and beyond

One form of context-based top-down control that has received substantial interest over the years in the area of multisensory processing is statistical learning, i.e. a process whereby an individual learns the underlying structure of stimulation within the environment by extracting information about the distribution of these inputs across time and/or space. As detailed below, statistical learning is known to support a variety of mental functions, both within and across the senses (reviewed in Frost et al. 2015), with effects transpiring across multiple temporal scales (e.g. Baier et al. 2006; Beierholm et al. 2009; Chandrasekaran et al. 2009; Barakat et al. 2013; Barenholtz et al. 2014; Altieri et al. 2015; Sarmiento et al. 2012, 2016).

A single testing session is frequently sufficient for participants to learn a relationship between two or more stimuli and utilise this information to improve their task performance. Many effects in the literature are based on temporal expectations, e.g. those that one stimulus follows another after a constant time interval (Niemi and Näätänen 1981; Coull and Nobre 1998; Cravo et al. 2011; Los and Van der Burg 2013; ten Oever et al. 2014). The importance of other types of expectations is increasingly reported. On the one hand, expectations linking spatial locations with high/low incidence of multisensory *in*congruence have been shown to modulate the ability of irrelevant sounds to influence judgements on the duration of visual stimuli (Sarmiento et al. 2012), with the effects transpiring even at a single-trial level (Sarmiento et al. 2016). On the other hand, exposure to as few as five successive presentations of phonetically

incoherent (e.g. "ra-ka") AV syllables presented before the McGurk ("da") syllable can substantially reduce the prevalence of the illusion (Nahorna et al. 2012). Importantly, expectations can affect stimulus processing even when based on irrelevant stimuli: Lateralised targets are detected faster in unattended spatial locations if they appear in sync with a rhythmic irrelevant stimulus (Jones 2015).

Expectations might be fundamental for the processing of some multisensory pairings. For example, identification of specific syllables might be uniquely facilitated by the delay existing between the onset of the mouth movement and that of the following speech sound (ten Oever et al. 2013). At the same time, a consistent asynchrony between visual and auditory stimuli, even if small in size and experienced only for a short time (a few minutes), suffices to alter subsequent conscious judgements of simultaneity on the same AV stimuli, which judgements adjust to the exposed stimulus lag (Fujisaki et al. 2004; Vroomen et al. 2004; cf. "Top-down control of multisensory processes by goals" section). Other results, however, suggest that the recalibration affects only the task-relevant multisensory pairings (Heron et al. 2012; Ikumi and Soto-Faraco 2014). Training that is focused on building more explicit associations between specific crossmodal stimuli can have similarly dramatic effects on multisensory processing. The temporal binding window for simple AV pairings can be narrowed by 40 % following as little as 1 h of training of AV simultaneity judgements with feedback, with the previously discussed network involving STC and low-level sensory cortices being activated by the same stimuli less strongly post-training (Powers et al. 2012). In turn, several days of explicit, object discrimination training can increase the efficiency of distinguishing among pairings of Gabor patches with particular tilts paired with tones of specific frequencies, as evidenced by reductions in strength of the eMSI that these stimuli trigger (Altieri et al. 2015).

One important way in which task-based context controls multisensory processing, besides the influences driven by the stimulus history, is the level of competition within the task-*irrelevant* sense. As proposed by Talsma et al. (2010), frequency of stimulation within the task-irrelevant sensory modality can determine perceptual salience of signals appearing within this sense, thus modulating the likelihood of these signals to interact (effortlessly and involuntarily) with the relevant sensory modality inputs. This idea is supported by the results of, for example, Sanabria et al. (2005), who showed that auditory motion judgements are affected by concurrent irrelevant moving dots, but this influence is strongly attenuated in contexts where the number of dots is large.

Role of the observer

The effects of context engendered by the current task setup rarely impact multisensory processing in the vacuum. Observers themselves are one vital source of context-based top-down control. Some of these influences involve *intra*-individual variability. For example, while performance on a visual detection task always naturally fluctuates across trials, it will do so periodically (and in a time-locked fashion) in the presence of a temporally predictive sound (Fiebelkorn et al. 2011, 2013). These results highlight the importance of both the ongoing oscillatory brain activity and of crossmodal inputs for perception. Notwithstanding, many of the studies on the observer-based top-down control focused on *inter*-individual differences, predominantly the long-term observers' experiences.

Some context-based influences afforded by the observer's history are quite intuitive. As discussed, prolonged experience with the perceptual processing of particular multisensory pairings might result in the detection of their presence relatively independently of one's goals (e.g. Froyen et al. 2009). Notably, the benefits of *long-term experiences* might generalise to multi-stimulus settings, with the multisensory congruence across some types of object features being detected independently of the level of competition within the relevant sense. Matusz et al. (2015b) showed that the search within central arrays for visual targets that are defined by a single feature (e.g. red targets) is sensitive to interference from distracters simultaneously appearing in the periphery (i.e. at irrelevant locations). The distracters matched the target-defining feature visually (a red square), aurally (a verbal colour label), or in both senses. Only the multisensory distracters interfered equally effectively with the search involving both three search array items and those involving no distracters at all, which suggests that the multisensory congruence across the colour dimension was detected (and extracted) independently of the demands the task imposed on participants' attentional control. Critically, however, the demand-independent nature of this multisensory interference was shown to have a developmental trajectory, not yet reliably present in 6-year-olds when the task demands were high (Matusz et al. 2015b). Thus, sufficient experience with the perceptual processing of particular multisensory pairings results in their detection when they match the unisensory target even when (1) the task defines the location of these stimuli as irrelevant (2) features within one sense only (i.e. vision) are deemed task relevant, and (3) the task difficulty eliminates the efficacy of target-matching unisensory distracters.

The impact of the observer's experiences within a particular environment goes beyond equipping specific multisensory pairings with detectability independent of goals and task demands; these experiences influence both the brain areas involved in the processing of the multisensory pairings as well as the efficiency of learning novel multisensory associations. For example, the activity enhancements in the STC that are similar to those observed in

Dutch readers for letter–sound AV pairings (van Atteveldt et al. 2004) were found in English readers for congruent *number*-sound, but not letter–sound, pairings (Holloway et al. 2015). These disparities are likely driven by the distinct levels of transparency (i.e. the consistency of correspondence between a letter and a single sound) across the two languages: While letter–sound pairings in Dutch and number-sound pairings in both languages are relatively transparent, English letter–sound pairings are not. In turn, Barenholtz et al. (2014) has recently showed that multisensory associations consistent with the stimulus statistics of the observer's environment (e.g. faces and voices of particular gender or congruent images and vocalisations of animals) are learnt more efficiently than incongruent multisensory pairings (e.g. faces and voices of different genders). Notably, both sets of results might be indicative of another, already mentioned, context-based control mechanism based on one's expectations. The degree to which the observer expects/believes the two inputs originate from the same source has long been proposed to impact multisensory processing (the "unity assumption"; Welch and Warren 1980). The influence of such expectations has been demonstrated, e.g. by more erroneous temporal-order judgements on gender-matching than gender-mismatching AV speech clips (Vatakis and Spence 2007). Multisensory speech congruence detection might be particularly sensitive to expectations, as suggested by results indicating that the McGurk illusion does not occur if participants interpret the sounds as noise (Tuomainen et al. 2005; Fig. 1b, c).

Other observer-based influences of context over multisensory processing are more akin to traditionally defined inter-individual differences. For example, while the detection of the looming quality within multisensory stimuli controls perception outside of the observer's awareness (Romei et al. 2009), its influence over later stages of information processing, following the stimulus offset, is dependent on the observer's attentional preferences (assessed with a multisensory divided-attention task). Specifically, for individuals with "auditory attentional preferences", but not those with "visual attentional preferences", the modulation of phosphene perception by looming sounds follows the velocity of these sounds (Romei et al. 2013). A major source of control over multisensory processing might constitute also the duration of the alpha cycle of the individual's oscillatory brain activity (ranging 8–14 Hz; Romei et al. 2012). For example, Cecere et al. (2015) showed that the length of the temporal window for the perception of the double-flash illusion (i.e. perceiving a single flash as two flashes if it is accompanied by two sounds) correlates with the individual alpha cycle, and it can be shrunk/enlarged by occipital transcranial alternating current stimulation that is slower/faster in its frequency than the individual frequency peak.

MS processes whose occurrence is independent of stimulus context

One possible candidate for a process that can perhaps exert influence over multisensory stimuli despite the context-based control of stimulus statistics, the observer's history, and their expectations could be the detection of multisensory simultaneity by the brain (Fig. 1c). First, the attenuations of the amplitudes of the eMSI triggered by the trained Gabor-frequency multisensory pairings in the Altieri et al. (2015; Section. 3.1) were seemingly accompanied by eMSI in response to the non-trained pairings. Additionally, some findings (Heron et al. 2012; Ikumi and Soto-Faraco 2014) suggest that sensitivity to simultaneity of multisensory stimuli (as measured, notably, with explicit, subjective judgements; cf. "Top-down control of multisensory processes by goals" section) might be altered only in respect to the multisensory pairings used during the exposure, rather than globally, for all simultaneous multisensory stimuli. In turn, the independence of the eMSI from experiences is strongly supported by their reports across different species. In fixating monkeys, the eMSI were observed in the primary and secondary auditory fields, in both local field potentials and spiking activity (Ghazanfar et al. 2005; Kayser et al. 2008; see also Lakatos et al. 2008). Furthermore, the eMSI were observed across several, primate and non-primate, species; critically, as already discussed, they were reported also in anaesthetised preparations (e.g. in rats, Barth et al. 1995; in cats, Rowland and Stein 2007). Jointly, these findings suggest that the brain's sensitivity to multisensory simultaneity, as reflected by the eMSI, might persevere despite top-down context-based influences.

Discussion and outlook

While it is well established that information across the senses interacts to jointly influence perception and neural responses (Stein 2012; Murray and Wallace 2012), the top-down control mechanisms orchestrating these effects are still far from being fully understood. Some of the first advances in this domain were made by Talsma et al. (2010), who delineated the conditions in which multisensory processing will likely depend on the current goals of the observer, i.e. top-down attention. However, to fully explain the full range of currently known multisensory phenomena, other top-down control mechanisms need to be invoked, such as those based on the stimulus context or multisensory stimulus attributes matching naturalistic objects (Fig. 1a). Having reviewed the findings that have demonstrated the relative importance of these three types of top-down control for multisensory processing, we will now propose a few emerging principles and then discuss their possible broader implications.

Towards mechanistic investigations of multisensory processing

One general idea emerging from the reviewed literature is that the nature of multisensory processing is multidimensional, rather than unitary or unidimensional. One viewpoint, based in part on well-characterised neurophysiologic observations (Meredith et al. 2012; Stevenson et al. 2014), contends that multisensory processes vary in their nature and substrate depending on the context (see also van Atteveldt et al. 2014a; De Meo et al. 2015). While more research is required to specify what constitutes a multisensory process and how many processes there are exactly, the evidence suggests that many multisensory processes seem intimately linked to the presence of stimuli containing specific, perceptually (e.g. elevation, size, intensity, frequency, the "looming" quality) or semantically corresponding, attributes [matching a particular object category, e.g. speech, script, everyday objects (tools/animals)]. At the same time, some multisensory processes seem to be elicited based merely on the fulfilment of particular physical (i.e. temporal coincidence) or cognitive (i.e. task relevance) conditions, independently of whether the eliciting stimuli contain specific attributes. A second general idea supported by the evidence reviewed here is that each of these processes could perhaps depend on goal-, context- and object-based types of top-down control to a differing degree. We next summarise this evidence for each of the three types of top-down control.

Multisensory processes seem to differ in how critically their *presence* depends on the current goals. Simultaneity as well as congruence of features across the senses can play a role even within task-irrelevant multisensory stimuli if these appear alone. However, the existing findings demonstrate that the presence of object-dependent processes is ultimately dependent on top-down attentional control, as they are no longer observed when heightened unisensory competition triggers enhanced goal-dependent control. In contrast, the current behavioural and functional neuroimaging findings across single- and multi-stimulus setups converge to suggest that multisensory simultaneity is detected and affects information processing, even if only weakly, despite attenuation by top-down attention. Furthermore, while some multisensory processes might occur only if the eliciting stimuli match attributes of an object across the senses, other processes might be elicited independently of such matching, as long as specific conditions are fulfilled, e.g. temporal coincidence. The few studies that directly compared object matching and simultaneity detection (until now, only within serial, single-stimulus paradigms) suggest that, when triggered by task-irrelevant stimuli, the effects of processing congruent simultaneous multisensory stimuli are stronger than that of merely simultaneous multisensory stimuli. These two types of multisensory processes might also generally differ in how (i.e. when and where) they modulate brain responses. Some studies also report distinct behavioural effects for the processing of stimuli from particular naturalistic object categories. This idea, if further confirmed, would run against the hypothesis that congruence is a unitary factor modulating MSI. Lastly, the large majority of multisensory processes seem to depend in their presence and/or how they transpire in the brain on context-based control that ranges from within-sensory competition to the individual's neurocognitive developmental outcome. Simultaneously, some evidence suggests that certain multisensory processes occur independently of such influences.

Implications: multisensory processing and beyond

While the proposal that goal dependence is a dimension of multisensory processing has helped to reconcile a long-standing debate in the area (Talsma et al. 2010), it fails to explain other contradictory findings that continue to accumulate. For example, context (e.g. experience) seems to determine the presence of multisensory processing in some cases (e.g. observers' reading proficiency, Froyen et al. 2009), but not in others (e.g. temporal coincidence within multisensory stimuli is detected early in life, Lewkowicz 2014). Furthermore, even irrelevant arbitrarily linked multisensory pairings are processed more strongly than their unisensory counterparts, but some multisensory processes are triggered solely by the presence of specific stimulus features (e.g. selective benefits for looming stimuli, e.g. Cappe et al. 2012). The ideas proposed here, i.e. that multisensory processing is multi-dimensional and that multisensory processes might be distinctly influenced by different top-down control mechanisms, could help to reconcile these results. It also harkens a reconsideration of some of the seminal findings in the literature. For example, Giard and Peronnet (1999) have demonstrated behavioural facilitation and eMSI during a task where the participants discriminated between two objects defined by (arbitrary) conjunctions of specific auditory and visual features. However, the observed effects might have been driven by a combination of unisensory (feature- and space-based attentional control mechanisms) and multisensory processes (AV simultaneity, newly learnt perceptual-feature match, arbitrary multisensory object templates), all likely engaged by this experimental setup (cf. Figure 2). If a particular multisensory process, such as the one based on a newly learnt match between crossmodal perceptual features, is of interest, the influence of other multisensory processes (e.g. detection of multisensory simultaneity) should be considered and eliminated. Thus, to summarise, there are a few possible advantages of the delineated factors contributing to multisensory processing: (1) They enable us to reconcile the past

contradictory findings, (2) they advance our present understanding of multisensory processing as a whole, and (3) they could foster more mechanistically oriented investigations within the field in the future.

More generally, the aim of the present review was to shed more light on information processing that may extend to a fuller understanding of how it occurs in naturalistic environments, where the multitude of sources of possible stimulation across the senses is matched by the number of possible sources of top-down control. For this purpose, the focus has been put, on the one hand, on the stimulus-driven processes engendered by stimuli that are typical for everyday situations, i.e. ones that engage multiple senses at once. On the other hand, the top-down control mechanisms scrutinised here go beyond the traditional investigations of the role of the observer's goals. If anything, top-down control processes are typically studied in relative isolation (object attributes/semantics: Doehrmann and Naumer 2008; context: Bar 2004; goals: Nobre and Kastner 2014, but see Braver 2012). As the literature reviewed here indicates, multisensory processing is actually subserved by a wide variety of processes. While these processes are frequently linked to specific object categories, others occur independently of such, as long as specific conditions (physical coincidence in time/space or task relevance) are fulfilled. These conceptualisations fit with a growing consensus that mechanisms subserving multisensory interactions are *de facto* not special or otherwise distinct from the mechanisms at play in processing of any sensory information (van Atteveldt et al. 2014a). As such, our review provides a wider purview into dimensions important for understanding top-down control processes in general and not exclusively in cases of multisensory stimulation.

Future directions

A major issue, which this review could not resolve and which prevents proposing a full-fledged framework at the current stage, is how the three types of top-down control interact with each other within each multisensory process as well as more generally, with each other. For one, it remains unclear whether matching attributes of an object enables a multisensory stimulus to be processed more strongly (compared to mere AV simultaneity) when this detection occurs outside of the attentional focus. Do the additional top-down inputs from higher-order brain areas triggered by over-learnt associations render object-dependent processes more sensitive to goal-based control (e.g. Fernández et al. 2015)? Are bottom-up activations driven by the detection of multisensory simultaneity always sufficiently strong to counteract goal-based influences (Matusz and Eimer 2011)? The links between the processes' sensitivity to goals and context (most notably, expectations) also require further research. Does sensitivity to expectations render a process sensitive to goals? Initial results suggest that regularity in stimuli (even the irrelevant ones), which elicits expectations, can act as a "double-edged sword"(Matusz et al. 2016). When helpful to the goal-directed behaviour, the irrelevant stimuli are continuously processed and utilised by the brain (e.g. ten Oever et al. 2014), but when they are unlikely to be helpful, e.g. in no-task setups, the regularity enables the brain to suppress these inputs (Matusz et al. 2016). The interactions between context- and object-based sensitivity are equally under-investigated. While expectations (e.g. Tuomainen et al. 2005; Stevenson and Wallace 2013) and long-term experiences (e.g. Froyen et al. 2009; Matusz et al. 2015b; ten Oever et al. 2013) are vital to the presence and the effects of many multisensory processes that are dependent on object matching, sensitivity to both multisensory simultaneity and speech congruency seems to develop early in life (Soto-Faraco et al. 2012; Lewkowicz 2014). The full extent to which dependence of a multisensory process on one form of top-down control indeed impacts dependence on other types of control is further complicated by the likely contingence of top-down expectations and object-matching processes on goal-based control, especially during development (e.g. Astle and Scerif 2011; Thillay et al. 2015; Amso and Scerif 2015). Shedding more light on the interdependencies within multisensory processes as well as between respective forms of top-down control is a critical next step to advance our understanding of sensory processing in real-world environments.

Acknowledgments This research was supported by grants from the Ministerio de Economia y Competitividad (PSI2013-42626-P), AGAUR Generalitat de Catalunya (2014SGR856), and the European Research Council (StG-2010 263145) to S.S-F, and the Swiss National Science Foundation (Grant #320030-149982 as well as the National Centre of Competence in Research project "SYNAPSY, The Synaptic Bases of Mental Disease" [Project 51AU40-125759]) and the Swiss Brain League (2014 Research Prize) to MMM. StO receives support from the Dutch Organisation for Scientific Research (Grant 406-11-068).

References

Aller M, Giani A, Conrad V, Watanabe M, Noppeney U (2015) A spatially collocated sound thrusts a flash into awareness. Front Integr Neurosci 9:16. doi:10.3389/fnint.2015.00016

Alsius A, Munhall KG (2013) Detection of audiovisual speech correspondences without visual awareness. Psychol Sci 24:423–431

Alsius A, Soto-Faraco S (2011) Searching for audiovisual correspondence in multiple speaker scenarios. Exp Brain Res 213:175–183

Alsius A, Navarra J, Campbell R, Soto-Faraco S (2005) Audiovisual integration of speech falters under high attention demands. Curr Biol 15:839–843

Alsius A, Navarra J, Soto-Faraco S (2007) Attention to touch weakens audiovisual speech integration. Exp Brain Res 183:399–404

Alsius A, Möttönen R, Sams ME, Soto-Faraco S, Tippana K (2014) Effect of attentional load on audiovisual speech perception: evidence from ERPs. Front Psychol 5:727. doi:10.3389/fpsyg.2014.00727

Altieri N, Stevenson RA, Wallace MT, Wenger MJ (2015) Learning to associate auditory and visual stimuli: behavioral and neural mechanisms. Brain Topogr 28(3):479–493

Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. Exp Brain Res 166:559–571

Amso D, Scerif G (2015) The attentive brain: insights from developmental cognitive neuroscience. Nat Rev Neurosci 16:606–619

Arnal LH, Giraud AL (2012) Cortical oscillations and sensory predictions. Trends Cogn Sci 16:390–398

Astle DE, Scerif G (2011) Interactions between attention and visual short-term memory (VSTM): what can be learnt from individual and developmental differences? Neuropsychologia 49(6):1435–1445

Baart M, Stekelenburg JJ, Vroomen J (2014) Electrophysiological evidence for speech-specific audiovisual integration. Neuropsychologia 53:115–121

Bach DR, Neuhoff JG, Perrig W, Seirfritz E (2009) Looming sounds as warning signals: the function of motion cues. Int J Psychophysiol 74:28–33

Baddeley A, Eysensck AW, Anderson MC (2009) Memory: motivated forgetting. Psychology press, New York

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

Baker CI, Olson CR, Behrmann M (2004) Role of attention and perceptual grouping in visual statistical learning. Psychol Sci 15(7):460–466

Bar M (2004) Visual objects in context. Nat Rev Neurosci 5:617–629

Barakat BK, Seitz AR, Shams L (2013) The effect of statistical learning on internal stimulus representations: predictable items are enhanced even when not predicted. Cognition 129:205–211

Barenholtz E, Lewkowicz DJ, Davidson M, Mavica L (2014) Categorical congruence facilitates multisensory associative learning. Psychon Bull Rev 21(5):1346–1352

Barth DS, Goldberg N, Brett B, Di S (1995) The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. Brain Res 678:177–190

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 7:1190–1192

Beierholm UR, Quartz SR, Shams L (2009) Bayesian priors are encoded independently from likelihoods in human multisensory perception. J Vis 9:23

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312

Besle J et al (2011) Tuning of the human neocortex to the temporal dynamics of attended events. J Neurosci 31:3176–3185

Bien N, ten Oever S, Goebel R, Sack AT (2012) The sound of size: crossmodal binding in pitch-size synesthesia: a combined TMS, EEG and psychophysics study. NeuroImage 59:663–672

Brang D, Towle VL, Suzuki S, Hillyard SA, Di Tusa S, Dai Z, Grabowecky M (2015) Peripheral sounds rapidly activate visual cortex: evidence from electrocorticography. J Neurophys. doi:10.1152/jn.00728.2015

Braver TS (2012) The variable nature of cognitive control: a dual-mechanism framework. Trends Cogn Sci 16:106–113

Cappe C, Thut G, Romei V, Murray MM (2009) 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

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

Cecere R, Romei V, Bertini C, Làdavas E (2014) Crossmodal enhancement of visual orientation discrimination by looming sounds requires functional activation of primary visual areas: a case study. Neuropsychologia 56:350–358

Cecere R, Rees G, Romei V (2015) Individual differences in alpha frequency drive crossmodal illusory perception. Curr Biol 25(2):231–235

Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. PLoS Comp Biol 5:e1000436

Coull JT, Nobre AC (1998) Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. J Neurosci 18:7426–7435

Cravo AM, Rohenkohl G, Wyart V, Nobre AC (2011) Endogenous modulation of low frequency oscillations by temporal expectations. J Neurophysiol 106:2964–2972

De Meo R, Murray MM, Clarke S, Matusz PJ (2015) Top-down control and early multisensory processes: chicken vs. egg. Front Integr Neurosci 9:17. doi:10.3389/fnint.2015.00017

Dehaene S, Cohen L (2007) Cultural recycling of cortical maps. Neuron 56:384–398

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annual Rev Neurosci 18:193–222

Diaconescu AO, Alain C, McIntosh AR (2011) The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. J Neurophysiol 106(6):2896–2909

Ding Y, Martinez A, Qu Z, Hillyard SA (2014) Earliest stages of visual cortical processing are not modified by attentional load. Hum Brain Map 35:3008–3024

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

Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. Psychol Rev 96(3):433–458

Fairhall SL, Macaluso E (2009) Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. Eur J Neurosci 29:1247–1257

Fernández LM, Visser M, Campos NV, Rivera CÁ, Soto-Faraco S (2015) Top-down attention regulates the neural expression of audiovisual integration. NeuroImage. 119:272–285

Fetsch CR, DeAngelis GC, Angelaki DE (2013) Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. Nat Rev Neurosci 14:429–442

Fiebelkorn IC, Foxe JJ, Butler JS, Mercier MR, Snyder AC, Molholm S (2011) Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. J Neurosci 31(27):9971–9981

Fiebelkorn IC, Snyder AC, Mercier MR, Butler JS, Molhom S, Foxe JJ (2013) cortical cross-frequency coupling predicts perceptual outcomes. Neuroimage 69:126–137

Finisguerra A, Canzoneri E, Serino A, Pozzo T, Bassolino M (2015) Moving sounds within the peripersonal space modulate the motor system. Neuropsychologia 70:421–428

Folk CL, Remington RW, Johnston JC (1992) Involuntary covert orienting is contingent on attentional control settings. J Exp Psychol Hum Percept Perform 18:1030–1044

Fort A, Delpuech C, Pernier J, Giard MH (2002) Dynamics of cortico-subcortical crossmodal operations involved in audio-visual object detection in humans. Cereb Cortex 12:1031–1039

Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci 9:474–480

Frost R, Armstrong BC, Siegelman N, Christiansen MH (2015) Domain generality versus modality specificity: the paradox of statistical learning. Trends Cogn Sci 19(3):117–125

Froyen DJ, Bonte ML, van Atteveldt N, Blomert L (2009) The long road to automation: neurocognitive development of letter–speech sound processing. J Cogn Neurosci 21:567–580

Fujisaki W, Shimojo S, Kashino M, Nishida SY (2004) Recalibration of audiovisual simultaneity. Nat Neurosci 7(7):773–778

Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J Neurosci 25(20):5004–5012

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

Gori M, Sandini G, Burr D (2008) Young children do not integrate visual and haptic form information. Curr Biol 18:694–698

Heron J, Roach NW, Hanson JV, McGraw PV, Whitaker D (2012) Audiovisual time perception is spatially specific. Exp Brain Res 218(3):477–485

Holloway I, van Atteveldt N, Blomert L, Ansari D (2015) Orthographic dependency in the neural correlates of reading: evidence from audiovisual integration in English readers. Cereb Cortex 25(6):1544–1553

Ikumi N, Soto-Faraco S (2014) Selective attention modulates the direction of audio-visual temporal recalibration. PloS One 9:e99311

Iordanescu L, Grabowecky M, Suzuki S (2009) Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. J Vis 9:1

Jones A (2015) Independent effects of bottom-up temporal expectancy and top-down spatial attention. An audiovisual study using rhythmic cueing. Front Integr Neurosci 8:96

Kayser C, Petkov CI, Logothetis NK (2008) Visual modulation of neurons in auditory cortex. Cereb Cortex 18(7):1560–1574

Klapetek A, Ngo MK, Spence C (2012) Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? Atten Percept Psychophys 74:1154–1167

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

Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci 23:571–579

Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT (2004) Semantic congruence is a critical factor in multisensory behavioral performance. Exp Brain Res 158:405–414

Lavie N (2010) Attention, distraction, and cognitive control under load. Cur Dir Psyc Sci 19:143–148

Lewkowicz DJ (2014) Early experience and multisensory perceptual narrowing. Dev Psychobiol 56:292–315

Los SA, Van der Burg E (2013) Sound speeds vision through preparation, not integration. J Exp Psychol Hum Percept Perform 39:1612

Luck SJ, Hillyard SA (1994) Spatial filtering during visual search: evidence from human electrophysiology. J Exp Psychol Hum Percept Perform 20:1000–1014

Lunghi C, Morrone MC, Alais D (2014) Auditory and tactile signals combine to influence vision during binocular rivalry. J Neurosci 34:784–792

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

Martuzzi R et al (2007) Multisensory interactions within human primary cortices revealed by BOLD dynamics. Cereb Cortex 17:1672–1679

Masterberdino S, Santangelo V, Macaluso E (2015) Crossmodal semantic congruence can affect visuo-spatial processing and activity of the fronto-parietal attention networks. Front Integr Neurosci 9:45

Matusz PJ, Eimer M (2011) Multisensory enhancement of attentional capture in visual search. Psychon B Rev 18:904–909

Matusz PJ, Eimer M (2013) Top-down control of audiovisual search by bimodal search templates. Psychophysiology 50:996–1009

Matusz PJ, Traczyk J, Sobkow A, Strelau J (2015a) Individual differences in emotional reactivity moderate the strength of the relationship between attentional and implicit-memory biases towards threat-related stimuli. J Cogn Psyc 27:715–724

Matusz PJ et al (2015b) The role of auditory cortices in the retrieval of single-trial auditory–visual object memories. Eur J Neurosci 41:699–708

Matusz PJ et al (2015c) Multi-modal distraction: insights from children's limited attention. Cognition 136:156–165

Matusz PJ, Retsa C, Murray MM (2016) The context-contingent nature of cross-modal activations of the visual cortex. Neuroimage 125:996–1004

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748

Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci 7:3215–3229

Meredith MA, Allman BL, Keniston LP, Clemo HR (2012) Are bimodal neurons the same throughout the brain? In: Murray MM, Wallace MT (eds) The neural bases of multisensory processes, chapter 4. CRC Press, Boca Raton (FL)

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. Cogn Brain Res 14:115–128

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

Mondloch CJ, Maurer D (2004) Do small white balls squeak? Pitch-object correspondences in young children. Cogn Affect Behav Neurosci 4:133–136

Murray MM et al (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. Neuroimage 21:125–135

Murray MM, Foxe JJ, Wylie GR (2005) The brain uses single-trial multisensory memories to discriminate without awareness. Neuroimage 27:473–478

Murray MM, Wallace MT (eds) (2012) The neural bases of multisensory processes. CRC Press, Boca Raton (FL)

Murray MM, Thelen A, Thut G, Romei V, Martuzzi R, Matusz PJ (2016) The multisensory function of the human primary visual cortex. Neurpsychologia 83C:161–169

Musacchia G, Schroeder CE (2009) Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. Hear Res 258:72–79

Naci L, Taylor KI, Cusack R, Tyler LK (2012) Are the senses enough for sense? Early high-level feedback shapes our comprehension of multisensory objects. Front Integr Neurosci 6:82. doi:10.3389/fnint.2012.00082

Nahorna O, Berthommier F, Schwartz JL (2012) Binding and unbinding the auditory and visual streams in the McGurk effect. J Acoust Soc Am 132:1061–1077

Nardini M, Jones P, Bedford R, Braddick O (2008) Development of cue integration in human navigation. Curr Biol 18:689–693

Nardini M, Bales J, Mareschal D (2015) Integration of audio-visual information for spatial decisions in children and adults. Dev Sci. doi:10.1111/desc.12327

Nardo D, Santangelo V, Macaluso E (2014) Spatial orienting in complex audiovisual environments. Hum Brain Map 35:1597–1614
Neil PA, Chee-Ruiter C, Scheier C, Lewkowicz DJ, Shimojo S (2006) Development of multisensory spatial integration and perception in humans. Dev Sci 9(5):454–464
Niemi P, Näätänen R (1981) Foreperiod and simple reaction time. Psychol Bull 89(1):133–162
Nobre K, Kastner S (eds) (2014) The Oxford handbook of attention. Oxford University Press, Oxford
Orchard-Mills E, Alais D, Van der Burg E (2013a) Crossmodal associations between vision, touch, and audition influence visual search through top-down attention, not bottom-up capture. Atten Percept Psychophys 75:1892–1905
Orchard-Mills E, Van der Burg E, Alais D (2013b) Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies. J Vis 13:6
Palmer TD, Ramsey AK (2012) The function of consciousness in multisensory integration. Cognition 125:353–364
Parise CV, Ernst M (2015) Correlation detection as a general mechanism for multisensory integration. J Vis 15:364
Parise CV, Spence C (2009) 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. PLoS One 4:e5664
Perrodin C, Kayser C, Abel TJ, Logothetis NK, Petkov CI (2015) Who is that? Brain networks and mechanisms for indentifying individuals. doi:10.1016/j.tics.2015.09.002
Powers AR, Hevey MA, Wallace MT (2012) Neural correlates of multisensory perceptual learning. J Neurosci 32:6263–6274
Raij T, Uutela K, Hari R (2000) Audiovisual integration of letters in the human brain. Neuron 28:617–625
Raij T, Ahveninen J et al (2010) Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. Eur J Neurosci 31:1772–1782
Rind FC, Simmons PJ (1999) Seeing what is coming: building collision-sensitive neurons 22(5):215–220
Rohe T, Noppeney U (2015) Cortical hierarchies perform Bayesian causal inference in multisensory perception. PLoS Biol 13(2):e1002073
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) Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. Curr Biol 19:1799–1805
Romei V, Gross J, Thut G (2012) Sounds reset rhythms of visual cortex and corresponding human visual perception. Curr Biol 22(9):807–813
Romei V, Murray MM, Cappe C, Thut G (2013) The contributions of sensory dominance and attentional bias to crossmodal enhancement of visual cortex excitability. J Cogn Neurosci 25:1122–1135
Rowland BA, Stein BE (2007) Multisensory integration produces an initial response enhancement. Front Integr Neurosci 1:4
Sanabria D, Soto-Faraco S, Spence C (2005) Assessing the effect of visual and tactile distractors on the perception of auditory apparent motion. Exp Brain Res 166(3–4):548–558
Santangelo V, Spence C (2007) Multisensory cues capture spatial attention regardless of perceptual load. J Exp Psychol Hum Percept Perform 33(6):1311–1321
Santangelo V, Di Francesco SA, Mastroberardino S, Macaluso E (2015) Parietal cortex integrates contextual and saliency signals during the encoding of natural scenes in working memory. Hum Brain Mapp 36:5003–5017
Sarmiento BR, Shore DI, Milliken B, Sanabria D (2012) Audiovisual interactions depend on context of congruency. Atten Percept Psychophys 74:563–574
Sarmiento B, Matusz PJ, Sanabria D, Murray MM (2016) Contextual factors multiplex to control multisensory processes. Hum Brain Mapp. doi: 10.1002/hbm.23030
Scerif G (2010) Attention trajectories, mechanisms and outcomes: at the interface between developing cognition and environment. Dev Sci 13:805–812
Schiff W, Caviness JA, Gibson JJ (1962) Persistent fear responses in rhesus monkeys to the optical stimulus of "looming". Science 136:982–983
Schroeder CE, Molhom S, Lakatos P, Ritter W, Foxe JJ (2004) Human–simian correspondence in the early cortical processing of multisensory cues. Cogn Proc 5:140–151
Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P (2010) Dynamics of active sensing and perceptual selection. Curr Opin Neurobiol 20:172–176
Senkowski D, Saint-Amour D, Kelly SP, Foxe JJ (2007) Multisensory processing of naturalistic objects in motion: a high-density electrical mapping and source estimation study. Neuroimage 36(3):877–888
Soto-Faraco S, Calabresi M, Navarra J, Werker J, Lewkowicz DJ (2012) The development of audiovisual speech perception. Multisensory development. Oxford University Press, Oxford, pp 207–228
Spence C, Deroy O (2013) How automatic are crossmodal correspondences? Conscious Cogn 22:245–260
Spierer L, Manuel AL, Bueti D, Murray MM (2013) Contributions of pitch and bandwidth to sound-induced enhancement of visual cortex excitability in humans. Cortex 49:2728–2734
Stein BE (2012) The new handbook of multisensory processing. MIT Press, Cambridge
Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. J Cogn Neurosci 19:1964–1973
Stevenson RA, Wallace MT (2013) Multisensory temporal integration: task and stimulus dependencies. Exp Brain Res 277(2):249–261
Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW (2010) Neural processing of asynchronous audiovisual speech perception. Neuroimage 49(4):3308
Stevenson RA et al (2014) Identifying and quantifying multisensory integration: a tutorial review. Brain Topogr 27:707–730
Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and computational mechanisms. Nat Rev Neurosci 15:745–756
Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends Cogn Sci 13:403–409
Sutherland CA, Thut G, Romei V (2014) Hearing brighter: changing in-depth visual perception through looming sounds. Cognition 132:312–323
Talsma D (2015) Predictive coding and multisensory integration: an attentional account of the multisensory mind. Front Integr Neurosci 9:19. doi:10.3389/fnint.2015.00019
Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J Cogn Neurosci 17:1098–1114
Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cereb Cortex 17:679–690
Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. Trends Cogn Sci 14:400–410
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. Cogn Brain Res 14:106–114
ten Oever S, Sack AT, Wheat KL, Bien N, Van Atteveldt N (2013) Audio-visual onset differences are used to determine syllable

identity for ambiguous audio-visual stimulus pairs. Front Psychol 4:331. doi:10.3389/fpsyg.2013.00331
ten Oever S, Schroeder CE, Poeppel D, Van Atteveldt N, Zion Golumbic EM (2014) The influence of temporal regularities and cross-modal temporal cues on auditory detection. Neuropsychologia 63:43–50
Thelen A, Matusz PJ, Murray MM (2014) Multisensory context portends object memory. Curr Biol 24:R734–R735
Thelen A, Talsma D, Murray MM (2015) Single-trial multisensory memories affect later auditory and visual object discrimination. Cognition 138:148–160
Thillay A, Roux S, Gissot V, Carteau-Martin I, Knight RT, Bonnet-Brilhault F, Bidet-Caulet A (2015) Sustained attention and prediction: distinct brain maturation trajectories during adolescence. Front Hum Neurosci 9:519
Tuomainen J, Andersen TS, Tiippana K, Sams M (2005) Audio–visual speech perception is special. Cognition 96:B13–B22
Tyll S, Bonath B, Schoenfeld MA, Heinze HJ, Ohl FW, Noesselt T (2013) Neural basis of multisensory looming signals. NeuroImage 65:13–22
van Atteveldt N, Ansari D (2014) How symbols transform brain function: a review in memory of Leo Blomert. Trends Neurosci Educ 3:44–49
van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. Neuron 43:271–282
van Atteveldt NM, Formisano E, Goebel R, Blomert L (2007) Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. NeuroImage 36:1345–1360
van Atteveldt N, Murray MM, Thut G, Schroeder CE (2014a) Multisensory integration: flexible use of general operations. Neuron 81:1240–1253
van Atteveldt NM, Peterson BS, Schroeder CE (2014b) Contextual control of audiovisual integration in low-level sensory cortices. Human Brain Mapp 35:2394–2411
van der Burg E, Olivers CNL, Bronkhorst A, Theeuwes J (2008) Pip-and-pop: nonspatial auditory signals improve spatial visual search. J Exp Psychol Hum Percept Perform 34:1053–1065
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 Ee R, van Boxtel JJ, Parker AL, Alais D (2009) Multisensory congruency as a mechanism for attentional control over perceptual selection. J Neurosci 29:11641–11649
Van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci USA 102:1181–1186
Vatakis A, Spence C (2007) Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. Percept Psychophys 69:744–756
Vroomen J, Keetels M, de Gelder B, Bertelson P (2004) Recalibration of temporal order perception by exposure to audio-visual asynchrony. Brain Res Cogn Brain Res 22(1):32–35
Walker-Andrews A, Lennon EM (1985) Auditory–visual perception of changing distance by human infants. Child Dev 56:544–548
Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. Psychchol Bull 88:638–667
Werner S, Noppeney U (2010a) Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J Neurosci 30:2662–2675
Werner S, Noppeney U (2010b) Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cereb Cortex 20(8):1829–1842
Yuval-Greenberg S, Deouell LY (2007) What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. J Neurosci 27(5):1090–1096
Zion Golumbic EM, Poeppel D, Schroeder CE (2012) Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain Lang 122:151–161