



Published in final edited form as:

*Neuron*. 2014 March 19; 81(6): 1240–1253. doi:10.1016/j.neuron.2014.02.044.

## Multisensory integration: flexible use of general operations

Nienke van Atteveldt<sup>1,2,3,\*</sup>, Micah M. Murray<sup>4,5</sup>, Gregor Thut<sup>6</sup>, and Charles Schroeder<sup>7,8,\*</sup>

<sup>1</sup>Neuroimaging & Neuromodeling group, Netherlands Institute for Neuroscience, Meibergdreef 47, 1105 BA, Amsterdam, The Netherlands <sup>2</sup>Dept. of Cognitive Neuroscience, Faculty of Psychology & Neuroscience, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands

<sup>3</sup>Dept. of Educational Neuroscience, Faculty of Psychology & Education, VU University Amsterdam, van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands <sup>4</sup>The Laboratory for Investigative Neurophysiology (the LINE), Neuropsychology and Neurorehabilitation Service and Radiodiagnostic Service, University Hospital Center and University of Lausanne, Switzerland

<sup>5</sup>EEG Brain Mapping Core, Centre for Biomedical Imaging (CIBM), Lausanne, Switzerland

<sup>6</sup>Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, G12 8QB, Glasgow, UK <sup>7</sup>Columbia University, Dept. Psychiatry, and the New York State Psychiatric Institute, 1051 Riverside Drive, New York, NY 10032, USA <sup>8</sup>Nathan S. Kline Institute, Cognitive Neuroscience & Schizophrenia Program, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA

### Abstract

Research into the anatomical substrates and “principles” for integrating inputs from separate sensory surfaces has yielded divergent findings. This suggests that multisensory integration is flexible and context-dependent, and underlines the need for dynamically adaptive neuronal integration mechanisms. We propose that flexible multisensory integration can be explained by a combination of canonical, population-level integrative operations, such as oscillatory phase-resetting and divisive normalization. These canonical operations subsume multisensory integration into a fundamental set of principles as to how the brain integrates all sorts of information, and they are being used proactively and adaptively. We illustrate this proposition by unifying recent findings from different research themes such as timing, behavioral goal and experience-related differences in integration.

### Keywords

multisensory integration; oscillations; context; divisive normalization; phase-resetting; canonical operations

---

\*Corresponding authors: Charles Schroeder, Ph.D., Nathan S. Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA, Phone: 845-398-5500, Fax: 845-398-5510, charles.schroeder6@gmail.com; Nienke van Atteveldt, Ph.D., VU University Amsterdam, Faculty of Psychology & Education, ONW, Van der Boechorststraat 1, 1081 BT Amsterdam, Tel: +31 (0)20 5983473, n.m.van.atteveldt@vu.nl.

## 1. Background

Perception is generally a multisensory process. Most situations involve sight, sound, and perhaps touch, taste and smell. Because most of our sensory input is acquired through, or at least modulated by, our motor sampling strategies and routines (Schroeder et al., 2010), perception is also a sensorimotor process. The brain must constantly combine all kinds of information and moreover track and anticipate changes in one or more of these cues. As cues from different sensory modalities initially enter the nervous system in different ways, historically, within-modality and cross-modal integration have been studied separately. These fields have lately converged, leading to the suggestion that neocortical operations are essentially multisensory (Ghazanfar and Schroeder, 2006). Following this notion of integration as an essential part of sensory processing, one might presume that it should not matter whether different inputs onto a neuron in, say primary auditory cortex, come from the same modality or not. Here, we will consider multisensory processing, with a focus on cortical auditory-visual processing, as representative of the brain's integrative operations in general. By extension, we do not consider multisensory integration to be a model, but rather an empirical probe that provides a unique window into the integrative brain and its adaptive nature. Multisensory paradigms provide inputs onto the integrative machinery in the brain that can be segregated to different receptor surfaces and initial input pathways, which facilitates a clean identification of their initial point of convergence; something that is more difficult to do within any of the main sensory modalities.

In the framework of the brain as a “Bayesian” estimator of the environment (Ernst and Banks, 2002; Fetsch et al., 2013; Knill and Pouget, 2004), uncertainty of sensory estimates will be minimized by combining multiple, independent measurements. This is exactly where multisensory integration might play an important role. Multisensory cues often provide complementary estimates of the same event, while within-modality cues tend to be more or less equally reliable. For example, at dusk, audition and vision provide signals of complementary strength, whereas visual shape and texture cues are both degraded. Moreover, cues from different modalities about singular objects or events often predict each other; they typically cross-correlate where one sense lags the other (Parise et al., 2012), whereas within-modality cues have a similar timing. In other words, interpreting multisensory and within-sensory cue-pairings can be considered as different means of reducing uncertainty about an event within the wider class of integrative processes.

In the search for specific underpinnings of multisensory integration, findings have been diverse, for example with regard to the role of binding “principles” such as spatial and temporal correspondence and inverse effectiveness (Stein and Meredith, 1993). The principle of spatial correspondence, a simplified version of which states that inputs are more likely to be integrated when they overlap in space, might be more dependent on task requirements than previously thought (Cappe et al., 2012; Girard et al., 2011; Sperdin et al., 2010). Furthermore, the importance of temporal proximity seems to differ across stimulus types and tasks (Stevenson and Wallace, 2013; Van Atteveldt et al., 2007a). The principle of “inverse effectiveness” is in itself context-dependent, as it predicts that stimulus intensity is a primary determinant of integration effects (Stein and Meredith, 1993). Moreover, the use of the multiple anatomical substrates of integration is also likely to be context-dependent.

For example, different potential sources of multisensory influences on low-level sensory cortices have been put forward, including direct “crossing” projections from sensory cortices of different modalities [e.g. (Falchier et al., 2002; Rockland and Ojima, 2003), direct ascending inputs from so-called “non-specific” thalamic regions (Hackett et al., 2007; Schroeder et al., 2003), as well as more indirect feedback inputs from higher order multisensory cortical regions (Smiley et al., 2007)]. There are several accounts for how these multiple architectures may be functionally complementary (Driver and Noesselt, 2008; Ghazanfar and Schroeder, 2006; van Atteveldt et al., 2013; Werner and Noppeney, 2010). The apparent flexibility in integrative “principles” and neuronal architectures indicates that multisensory integration is not fixed or uniform, but strongly adaptive to contextual factors. This idea of context-dependent integration is not new, especially within the aforementioned view of the brain as a statistically-optimal cue integrator. For example, in the modality-appropriateness framework, context determines which of the senses provides the most appropriate, or reliable, information (Welch and Warren, 1980). In general, visual cues will be most reliable for spatial judgments whereas sounds provide more reliable temporal cues, so either modality can dominate perception in different contexts. In short, multiple factors, such as input properties and behavioral goal, dynamically and flexibly interact to provide the momentary context. This composite context seems to adaptively recruit different neuronal operations and pathways for integration.

Based on recent neurophysiological findings, detailed below, we here take the perspective that multisensory integration can be subsumed under canonical integrative operations in a more unified view. The diversity in research findings in this view reflects the flexible use of these canonical operations. To substantiate this perspective, in Section 2, we will first consider two exemplar canonical integrative operations in detail: divisive normalization (DN) and oscillatory phase-resetting (PR). We then illustrate how a combination of such canonical operations may enable the observed highly adaptive, context-dependent nature of the brain's integrative processing. Rather than providing an exhaustive review of all reported context-effects, we focus here on three main themes within multisensory research: timing, behavioral relevance and effects of experience. Accordingly, in Section 3, we discuss how temporal predictability influences the brain's operation mode, and illustrate the potential complementary role of DN and PR in these different modes. In Section 4, we focus on how different behavioral goals, or task-sets, guide the flexible use of canonical integrative operations. In Section 5, we address how integrative processing is shaped by short-term or longer-term changes in the sensory context, such as that related to experience or training. It should be noted that another main theme- i.e. attention - influences processing within all these themes. Therefore, we frame attention as working in concert with the suggested canonical integrative operations, and discuss its role whenever relevant for explaining how these operations subsume adaptive integration within the different main themes. Finally, we discuss whether certain integrative processes are more context-dependent than others.

Some final notes on what we mean by “context” are worthwhile. In this review, we use context as the immediate “situation” in which the brain operates. Context is shaped by external circumstances, such as properties of sensory events, and internal factors such as behavioral goal, motor plan, and past experiences. In fact, internal and external factors often interactively define the context, for example, when the structure of relevant sensory events

switch the brain in a specific internal “operation mode” such as a “rhythmic mode” (Schroeder and Lakatos, 2009). Longer-lasting determinants of context refer to individual experiences that shape integrative operations, a “developmental context”. Our viewpoint is therefore more basic than, though not mutually exclusive of, the use of context as a more voluntary psychological function, the ability to contextualize information with the purpose of assigning value to events (Maren et al., 2013).

## 2. Multisensory processing as representative of integrative operations in general

The neuronal bases of sensory integration are formed by: 1) the convergence of synaptic inputs from multiple sources onto individual neurons, 2) the operation those neurons perform to produce “integrated” output signals, and 3) interactions with other neurons within and across populations, such as network-level interactions after the initial integration process. Converging inputs can originate from the same or from different sensory modalities. Moreover, sensory inputs also need to be combined with ongoing motor actions, as well as with other top-down signals that relate current inputs with knowledge, memories and predictions. In this review, we advance a general neurophysiological framework that is designed to account for this wide variety of integrative processes that the brain constantly performs.

Prior studies have suggested that within and across modality integration have different behavioral (Forster et al., 2002; Gingras et al., 2009; Girard et al., 2013) and neurophysiological (Alvarado et al., 2007) underpinnings. The question is whether this is due to A) unique intrinsic properties of multisensory neurons, such as their integrative computations; or B) differences in how multiple convergent cues typically interact within versus across the senses, i.e. in which way they provide different estimates about an event, and how this is locally wired. In the latter case, multisensory integration is not special in essence, but the information provided by different modalities may lead to stronger neuronal interactions, as this information tends to be complementary in its ability to reduce uncertainty about events. At the single-neuron level in the superior colliculus (SC), it has been found that it is indeed the *input* that determines the response, and not the neuron type (unisensory or multisensory). Alvarado and colleagues (2007) compared visual-visual integration with visual-auditory integration in multisensory and unisensory neurons in the cat SC. For visual-visual integration, they found the same sub-additive integrative response in multisensory and unisensory neurons. For audiovisual integration, which only occurs in multisensory neurons, the response was different; namely additive or super-additive. Computational models from the same group explain these different responses of multisensory SC neurons by different clustering of synaptic inputs (Alvarado et al., 2008; Rowland et al., 2007). Inputs that cluster together on the same dendritic unit of a neuron, as was the case only for multisensory inputs, will produce a stronger synergistic interaction compared to inputs that do not cluster together. It should be noted that such single-cell interactions may be more determinative in structures like SC than in neocortex. Instead and as we will substantiate below, ensemble processes provide more degrees of freedom for flexibility in differing contexts.

Recently, the divisive normalization model developed for visual processing (Carandini et al., 1997; Reynolds and Heeger, 2009), and described as a “canonical operation” (Carandini and Heeger, 2012), was shown to explain important features of multisensory integration such as inverse effectiveness and the spatial principle (Fetsch et al., 2013; Ohshiro et al., 2011). An important feature of this model is that integrative outputs are normalized by surrounding neurons (Figure 1), and thus, it transcends the level of single neuron responses. The model explains integration effects in both subcortical (SC) and cortical (MST) measurements. Interestingly, it also accounts for adaptive changes in the weighting of different inputs as a function of cue reliability (Morgan et al., 2008), which provides a neural basis for similar effects at the performance level (Ernst and Banks, 2002). In sum, the network-level operation of divisive normalization is able to explain cue integration regardless of the origin of the cues, and as a flexible process depending on cue reliability. An open question in this framework is how *predictive* cue integration is accomplished, i.e., how cues influence the processing of future events. A neural mechanism that is especially suitable to explain such predictive interactions is that of phase-resetting of ongoing oscillatory activity (Kayser et al., 2008; Lakatos et al., 2007; Lakatos et al., 2005). Taking primary auditory cortex as an example, it has been shown that response amplitudes to sounds depend on the phase of ambient oscillations, with “ideal” and “worst” phases in terms of neuronal excitability (Lakatos et al., 2005). A predictive influence can be exerted if one event resets the phase of these ongoing excitability fluctuations and thereby influences processing of upcoming events in the same or a different modality (Figure 1 & 2). The phase-reset mechanism is not specific for multisensory interactions, but rather represents a more general mechanism through which different sensory, motor and attentional cues can modulate ongoing processing (Lakatos et al., 2013; Makeig et al., 2004; Rajkai et al., 2008; Shah et al., 2004) or memory formation (Rizzuto et al., 2003). Therefore, we propose phase-resetting as a second canonical operation enabling flexible integration of multiple sensory, motor and other top-down cues.

Divisive normalization (DN) and oscillatory phase-resetting (PR) by themselves seem two attractive candidates of population-level canonical integrative operations. How can a combination of such operations contribute to the highly adaptive nature of the brain's integrative processing? As schematically depicted in Figure 1, we propose that DN and PR may operate in a complementary fashion rather than in the service of the same goal – they overlap in their outcome but are likely relevant in different brain areas, at different temporal scales, and in different operation modes (“rhythmic” vs. “continuous” modes, see Section 3). It should be noted that there are likely additional canonical operations; DN and PR are interesting candidates based on recent empirical evidence, but are unlikely to explain *all* integrative activity. DN and PR overlap in their outcome in that they both influence a system's sensitivity to weak inputs, either by influencing the ambient excitability (PR) or by combining a pool of responses to amplify output non-linearity for weak inputs (DN). Furthermore, both operations produce a sharpened perceptual tuning, which in both cases can be influenced by attention. In DN, a second stimulus that differs on a certain dimension (e.g., location, timing, cue reliability) suppresses the excitatory response to a first stimulus (Figure 1, green box); this 2<sup>nd</sup> stimulus strongly influences the normalization signal in a broadly tuned population, but only weakly increases the excitatory signal (Ohshiro et al.,

2011). This sharpened tuning likely facilitates the binding of (multi)sensory cues: excitatory signals are reinforced when two stimuli correspond (in time, location, etc.), but suppressed when they are dissimilar. Attention can influence this process by modulating the normalization signal (Reynolds and Heeger, 2009). Resetting the phase of oscillations synchronizes activity between areas, and increases the impact of ascending sensory inputs. For lower frequency oscillations such as delta [1-4 Hz] and theta [5-7 Hz] ranges, PR tunes the sensory systems to specific moments in time (Figure 1 - red box & Figure 2), and possibly to other dimensions such as spectral content (Lakatos et al., 2013). PR occurs both within and across senses, but might be particularly beneficial across senses because of the strong predictive power across modalities related to timing differences [(Schroeder et al., 2008); see Section 3]. Selective attention guides this process by promoting selective entrainment of ongoing activity to the rhythm of the events in the attended stream (Lakatos et al., 2009).

Although DN and PR may overlap in their outcome, we propose that they are complementary in at least three important aspects. The first aspect concerns the type of inputs they operate on, which optimizes these operations for different brain areas (Figure 1A). DN models have been designed to explain interactions among multiple *excitatory* inputs, such as two visual stimuli in V1 (Carandini et al., 1997). This seems crucial, as the suboptimal stimulus should excite the normalization pool in order to cause the suppressive divisive influence. For explaining multisensory interactions, the model may therefore mainly be relevant in areas where neurons exist that receive converging excitatory inputs from different modalities, i.e., higher-level cortical areas, such as macaque MSTd [as in (Ohshiro et al., 2011)] or STP [e.g. (Barraclough et al., 2005; Dahl et al., 2009)] and putatively human STS, or subcortical structures like SC (Stein and Meredith, 1993) – these are what we term “classical” integration areas. In low-level sensory cortices, inputs from a non-preferred modality are often of a modulatory rather than driving nature. Modulatory inputs are shown to induce phase-resetting of ongoing oscillations, e.g. visual or somatosensory inputs in low-level auditory and visual cortices, thus affecting the temporal pattern of firing probability fluctuation in local neurons, rather than driving action potentials per-se [(Lakatos et al., 2007); Figure 2]. In sum, in the context of multisensory integration, DN may be operating mostly in areas that receive converging excitatory inputs (“classical” integration areas), whereas PR may be a more common operation in low-level sensory cortices.

Secondly, DN and PR may operate at different time scales (Figure 1B). The DN operation may be in part mediated by fast-spiking parvalbumin-positive (GABAergic) interneurons [(Reynolds and Heeger, 2009); but see (Carandini and Heeger, 2012)] which cause depolarization-hyperpolarization cycles that correspond to gamma oscillations (Whittington et al., 1995). Recent modeling work suggests that if population responses to different inputs phase-lock to different phases of gamma oscillations, this facilitates the inhibitory division operation [(Montijn et al., 2012); but note that they do not explicitly test different oscillatory frequencies]. In contrast, PR can occur at all time-scales, but many fundamental, task-related modulations occur at time scales corresponding to the lower frequencies [delta/theta; (Schroeder and Lakatos, 2009)]. Different frequency bands are hierarchically coupled, the low-frequency PR produces rhythmic amplitude modulation of higher (e.g., gamma)

frequencies. Cross-frequency PR at lower frequencies is believed to be mediated through modulatory inputs into the most superficial cortical layers, contacting both Layer 1 interneurons (possibly not fast-spiking) and apical tufts of lower layer pyramidal cells, though it may also impact the fast-spiking interneurons. The different time-scales implicate that DN and PR mechanisms may in turn dominate rhythmic vs. continuous modes of neural operation (Figure 1C) that depend on whether or not relevant inputs are predictable in time; this will be further discussed in Section 3.

The complementary goals of DN and PR operations may be summarized as analyzing *content* (DN) versus setting *context* (PR). It is widely accepted that cortical encoding of information (content) entails distributed patterns of action potentials in pyramidal cell ensembles, and albeit perhaps less widely so, that transmission of information across brain areas is enhanced by coordination of neuronal firing through oscillatory coherence (Fries, 2005; Rodriguez et al., 1999). The latter, dynamical modulation of excitability, also performs a “parsing” operation, whose temporal scale depends on the interaction of task parameters (e.g., tempo) with the oscillatory frequencies that can be employed by the brain (Schroeder and Lakatos, 2009), and it corresponds to the neurophysiological context (Buzsaki and Draguhn, 2004; Lakatos et al., 2009). An example is analyzing the detailed frequency structure of a complex auditory signal such as speech. Multisensory cues may enhance the neural representation of speech inputs by direct excitatory convergence and resulting enhancement of neuronal firing in higher-order brain areas such as human STS (Beauchamp, 2005; Van Atteveldt et al., 2010) or macaque STP/MSTd, for which DN has been shown to be a good model (Ohshiro et al., 2011). In contrast, PR operations may contribute to parsing contextual information at lower rates. For example, PR parses lower-frequency fluctuations in speech, reflecting syllables/phrases and prosody; the crucial units to understand speech, and this is also the rate at which visual information, such as articulatory gestures, appears most helpful (Schroeder et al., 2008). While it is unlikely that there is complete segregation of multisensory PR and DN operations across lower and higher levels of the cortical hierarchy, such a bias of these operations seems conceivable; multisensory interactions produced by PR in lower-level sensory areas provide the optimal context to process relevant events, rather than binding cues from different modalities for the purpose of content analysis – which may be accomplished by normalization operations in more “classic” multisensory areas.

### 3. Timing and predictability in encoding: complementary operations in different operation modes

To optimize efficient processing of incoming information, the brain constantly generates predictions about future events (Friston, 2011). This is particularly clear in “active sensing,” when sensory events enter the system as a result of motor activity that the brain initiates (Hatsopoulos and Suminski, 2011; Schroeder et al., 2010). Multisensory cues play an important role in this process of anticipation, as cues in one modality often predict what will happen in other modalities. One reason for this is that different senses have different timing properties (Musacchia and Schroeder, 2009), which can strengthen the predictive value across modalities. Secondly, as the senses provide complementary estimates of the

environment, the brain is able to keep generating predictions even when one type of information is temporarily degraded or unavailable.

It is increasingly clear that the brain is particularly well-equipped to exploit the temporal structure of sensory and motor information. In fact, the active nature of perception and the rhythmic properties of our motor sampling routines predict that most input streams have rhythmic properties (Schroeder et al., 2010). Still, some contexts have a more predictable temporal structure than others, and importantly, task dynamics determine the relevance and usability of temporal structure (Schroeder and Lakatos, 2009). One context in which rhythmic information is very important is social interaction including verbal communication, i.e. for predicting what others will do or say, and when (Hasson et al., 2012; Luo and Poeppel, 2007; Schroeder et al., 2008; Zion Golumbic et al., 2012). Complementary cues from different sensory modalities, or motor cues, may be particularly important as biological rhythmicity is typically not entirely regular, such as in speech (Giraud and Poeppel, 2012). Delays between visual and auditory counterparts of natural events are typically predictable (Figure 3d) and will be used as predictive cues as long as they are reliable (Vroomen and Stekelenburg, 2010). The temporal offsets between, for instance, visual and auditory cues in speech are well-learned (Thorne and Debener, 2013), and this is a basic knowledge that is imposed on incoming information that helps to keep the temporal perception constant despite the fact that audiovisual lag depends on the distance of the source from the subject (Schroeder and Foxe, 2002). There are also circumstances lacking temporal prediction cues, for example if a cat watches a mouse hole and listens to the mouse's scratching noises inside. In this case, multisensory information may help optimize the cat's chances in a different way, e.g., by generating a spatial prediction: the cat knows *where* to expect the mouse, but not *when*. The brain is thought to flexibly switch between the former (“rhythmic”) and latter (“continuous” or “vigilance”) processing modes depending on task demands and the dynamics of the environment.

The different characteristics of the available integrative operations, such as divisive normalization (DN) and oscillatory phase-resetting (PR), suggest a potentially complementary role in rhythmic versus continuous modes (Figure 1c). During continuous mode processing in the absence of predictable input, gamma-rate processes are thought to be used continuously, along with suppression of lower frequency power to avoid periods of weaker excitability. To the extent that the DN operation operates within gamma-cycles it can be used in this mode continuously (at a resolution of  $\sim 25$  ms) to aid multisensory integration and attentional filtering. Taking the example of the cat and the mouse hole, audiovisual spatial tuning may be continuously sharpened by DN processes during this vigilance situation with clear spatial predictions. During rhythmic mode processing, neuronal excitability cycles at low frequencies, and PR operates at this scale to synchronize neuronal high-excitability phases to the anticipated timing of relevant inputs. As gamma-oscillation amplitudes are coupled to the lower-frequency (delta/theta) phase (see Figure 1b), it may be that DN operations also occur during the high-excitability phase of the lower frequency oscillation during rhythmic mode processing.

For illustrating the flexibility in using canonical integrative operations, we will consider how the PR operations are used in various ways at a “cocktail party” (Figure 3) – a situation with



abundant rhythmic information. At such a party, first, we need to select which speaker to focus on and which to ignore. Understanding the speech of the attended speaker is facilitated by anticipatory cues from the visual as well as the motor system, and the interplay of these different cues is orchestrated by selective attention. The first process involves active exploration of the scene (Figure 3a), and the framework of “active sensing” (Schroeder et al., 2010) notes that we sample the environment by systematic patterns of saccades and fixations, and fixations are thought to be able to phase-reset excitability oscillations in visual cortex by efference copy signals (Ito et al., 2011; Melloni et al., 2009; Rajkai et al., 2008). The role of active sensing is clearly context-dependent. For example, it might depend on the dominant modality (that likely differs across individuals, see section 5) as vision is thought to depend more on rhythmic motor routines than audition (Schroeder et al., 2010; Thorne and Debener, 2013); although eye position has also been shown to influence auditory cortex (Werner-Reiss et al., 2003). When the relevant speaker is found, selective attention processes orchestrate the entrainment of ambient oscillatory activity in appropriate sensory areas and frequency ranges to the temporal pattern of the task relevant speech stream (Figure 3b). Visual cues fine-tune the entrainment to the attended speech stream in the auditory cortex (Figure 3c), and this cross-modal phase reset is predictive (Figure 3d); as facial articulatory cues and head movements precede the auditory speech input (Chandrasekaran et al., 2009), they provide predictions that enable the auditory system to anticipate what is coming (Schroeder et al., 2008; van Wassenhove et al., 2005). Recent evidence shows that the brain indeed selectively “tracks” one of multiple speech streams (Zion Golumbic et al., 2013b) and that predictive visual cues enhance the ability of the auditory cortex to use the temporally structured information present in the speech stream. Interestingly, this was the case especially under conditions of selective attention, when the subject had to attend to one of multiple speakers (Zion Golumbic et al., 2013a). In fact, with only auditory information, no enhanced tracking of the attended stream was observed. This suggests that multisensory information enhances the use of temporally predictive information in the input, but mostly so during noisy conditions where selective attention is required. In addition to predicting *when* a speech cue will arrive, the timing between visual and auditory inputs has recently been shown to also aid in predicting *which syllable* will be heard (Ten Oever et al., 2013). The predictive visual influences that facilitate selective listening at a cocktail party are very likely exerted through PR processes (Schroeder et al., 2008; Zion Golumbic et al., 2013a), although this has not been exclusively demonstrated in such a complex real-life situation. Although cross-modal phase-reset has been shown to occur for both transient and extended inputs (Thorne et al., 2011), it seems especially advantageous for ongoing inputs such as a speech stream, such that the temporal input pattern can be matched to the pattern of brain oscillations (Schroeder et al., 2008) and predictive information can build up in strength. Zion-Golumbic and colleagues (2013b) have shown that selective entrainment of both low-frequency and high-gamma oscillations to the attended speech stream increase as a sentence unfolds, indicating the use of accumulated spectro-temporal regularities in both auditory and predictive visual cues.

The evidence reviewed in this section illustrates how the brain dynamically shifts between “rhythmic” and “continuous” modes of operation, and the mechanisms through which the senses interact switch accordingly. If allowed by task and input dynamics, the brain uses

temporally structured information to optimally anticipate incoming information, which may actually often be “enforced” by motor and/or attentional sampling routines during active perception. The brain's anticipatory capacity enables highly efficient processing, and appears to depend strongly on PR operations; PR aligns neuronal excitatory peaks to the input periods anticipated to be most relevant. Multisensory information can be especially helpful in this process as the senses often precede and complement each other and thereby improve predictive power. The impact of multisensory information and the role of motor cues depend on the context, such as the attentional context or dominant modality. Most often, sensory inputs interact with internal cues such as attention and motor efference copies. In the absence of predictable input, lower frequency oscillations are suppressed, resulting in extended periods of high excitability, and thus DN may operate more continuously.

#### **4. How do behavioral goals guide the flexible use of canonical integrative operations?**

Behavioral goals determine which inputs are relevant and which actions are required. Attention can work to select those relevant inputs and actions, although such goal-driven selection is supplemented by pure bottom-up attentional orienting (Talsma et al., 2010). Not surprisingly, many integrative processes are highly adaptive to behavioral goal - in laboratory experiments typically manipulated by task instruction. It has been shown that basic binding “principles”, such as temporal coincidence, are influenced by task demands (Mégevand et al., 2013; Stevenson and Wallace, 2013). An interesting view from sensory substitution research is that cortical functional specialization may actually be more driven by task goals than by the modality of sensory experience (Reich et al., 2012). This is based on repeated findings that task-related specialization in e.g. visual cortex is independent of the input modality (Striem-Amit et al., 2012). The importance of behavioral goal is also inherent to active sensing (Schroeder et al., 2010), since motor actions implementing goal-directed behavior are tightly linked to perception (see Section 3). The profound role of behavioral goal is further shown by findings of different performance and neural effects for integrating identical information with different behavioral goals (Fort et al., 2002b; Girard et al., 2011; Van Atteveldt et al., 2007b; van Atteveldt et al., 2013). For example, in a series of experiments we found that neural integration effects observed under passive conditions (Van Atteveldt et al., 2004) or a unimodal task with irrelevant cross-modal information (Blau et al., 2008) were overruled by an explicit task demand to match auditory and visual inputs (Van Atteveldt et al., 2007b), probably because the task determined the behavioral relevance of the inputs.

How does behavioral relevance influence neuronal operations? In regards to PR, compelling evidence that this process is adaptive to the momentary goal is provided by studies that show task-dependent, adaptive phase-reset when identical inputs are perceived under different task conditions. For example, Bonte and colleagues (Bonte et al., 2009) had participants listen to three vowels spoken by three speakers and instructed them to perform a 1-back task either focused on vowel identity or on speaker identity. The results showed that alpha oscillations temporally realigned across speakers for the vowel task, and across vowels for the speaker task. This demonstrates that phase realignment is transient and highly adaptive to the

momentary goal, and may constitute a mechanism to extract different representations of the same acoustic input, depending on the goal. Whereas alpha-band oscillations may be involved in speech analysis at the vowel level, the same mechanism may apply to other time-scales of analysis, such as theta band for syllables, (Luo and Poeppel, 2007) depending on the input and task. There is indeed evidence that different combinations of oscillatory frequencies can be entrained, depending on the context (Kösem and van Wassenhove, 2012; Schroeder et al., 2008; van Wassenhove, 2013). Perhaps the most intriguing example, albeit still speculative, is that of audiovisual speech [reviewed by (Giraud and Poeppel, 2012; Schroeder et al., 2008)]. Sentences are composed of phrases (lasting ~300-1000 ms), overlapping slightly with a faster segmental unit, the syllable (lasting 150-300 ms), and distinct from even faster elements, such as formant transitions (lasting as little as 25 ms). Interestingly, formants are nested within syllables and syllables within phrases, and there is an *uncanny* resemblance between these speech metrics and the delta (1-4 Hz), theta (5-7 Hz), and gamma (30-50 Hz) frequencies that dominate the ambient oscillatory spectrum in auditory cortex. These frequencies, in particular, exhibit prominent hierarchical cross-frequency couplings (Lakatos et al., 2005; Van Zaen et al., 2013) that strongly parallel the nesting of faster into slower segmental units in speech. Interestingly, different visual cues of the speaker (which all precede the generation of vocalizations by 150-200 ms, (Chandrasekaran et al., 2009)) may reset different oscillatory frequencies, with prosodic cues such as eyebrow raises and head inclinations (occurring at the phrasal rate) resetting delta oscillations, more rapid articulatory gestures of the lips resetting theta oscillations and so on. Obviously, articulatory movements of the lips and prosodic movements of the head can occur separately; they each tend to occur in streams, and thus, if salient to the observer system, can entrain appropriate frequencies separately. In natural conversation, the faster articulatory gestures are generally nested within the slower prosodic gestures, and thus, it is likely that the resetting of higher and lower oscillatory frequencies occurs in a coordinated fashion.

Another aspect of flexibility in PR operations is that it allows task demands to dictate which particular input phase-resets which sensory cortex. The sensory cue arriving first in a given brain region is special in that it allows anticipation of later inputs (perhaps pertaining to the same external event) conveyed via other modalities. This is related to both internal and external timing differences across sensory modalities, and also depends on which modalities are combined (Schroeder and Foxe, 2002). In some situations, when multisensory inputs are brief, discrete events and originate from inert objects that are close (<1 m), visual and auditory information reaches the peripheral sensory surfaces practically simultaneously. As auditory inputs have faster cortical response latencies, they can reset the phase of ongoing oscillations in the visual cortex and thereby enhance visual excitability [Figure 2, (Romei et al., 2012)]. In other situations, these internal timing differences are cancelled by external ones, for example when events occur at a distance and/or originate from moving sources. For example, in many communicatory actions, such as speech but also nonverbal actions like gestures that produce a sound (or a touch), the motor actions produce visible cues before the sounds start. Therefore, in these situations, it is more adaptive that the visual inputs align ongoing auditory oscillations to the upcoming sounds (Schroeder et al., 2008).

Complementary to oscillatory PR processes, other canonical neuronal operations such as DN seem also able to account for adaptive integration processes, for example, related to dynamic changes in cue reliability. Cue reliability changes when a changed behavioral goal influences which inputs are relevant. For example, if a timing-focused goal switches to a spatially oriented one, visual and auditory cues become more and less reliable, respectively. In the DN framework, such changes in cue reliability could dynamically and adaptively change the weighting of different cues in the integrated response (Morgan et al., 2008). Using fMRI, it has been shown at a more macroscopic level that rapid changes in auditory versus visual reliability during speech perception dynamically changed functional connectivity of the respective low-level sensory cortices and superior temporal sulcus (Nath and Beauchamp, 2011). In sum, these examples demonstrate that adaptive use of canonical integrative operations such as PR and DN can provide a neural basis for goal-directed sensory processing and that context factors, such as task goal or the uncertainty-reducing power of one cue over another, determine how internal (motor, attention) and external (sensory) cues interact.

## 5. Experience-related shaping of integration operations

In light of the idea that neuronal oscillations and divisive normalization reflect canonical operations that are adaptively used for integrative processes, the question arises to what extent individual differences in multisensory integration can be accounted for by these operations and vice versa: how an individual's development and experience shape the characteristics of canonical operations. If we return to viewing of the integrative brain from a “Bayesian” perspective, it logically follows that integration is shaped by individual factors, given the role of priors that are shaped by past experiences.

Experience influences multisensory and sensorimotor integration. Increased sensitivity to audiovisual synchrony has been found as a result of perceptual (Powers et al., 2009) or musical (Lee and Noppeney, 2011) training. Powers and colleagues used an audiovisual simultaneity judgment task with feedback, which may have sharpened temporal tuning by DN. Although speculative at this point, it seems plausible that the “temporal principle” of integration as explained by DN (Ohshiro et al., 2011) is fine-tuned through learning, e.g. by sharpening tuning of individual neurons which affects the population-level normalization process. In the case of musical training, temporal sensitivity may increase as a result of more accurate predictions generated by the motor system, and thus more specific PR processes. These studies underscore the flexibility in temporal processing and suggest that experience may fine-tune the accuracy of temporal predictions generated either by motor or by cross-modal cues, possibly by promoting more rapid, accurate selection of the task-relevant sensory or motor rhythms and synchronization of ambient activity to that rhythm. Experience-related effects on multisensory and sensorimotor interactions can be shaped gradually during development (Hillock-Dunn and Wallace, 2012; Lewkowicz, 2012; Lewkowicz and Ghazanfar, 2009), but can also occur very rapidly, as shown by recalibration experiments in temporal, spatial and content (speech) domains (Van der Burg et al., 2013; Vroomen and Baart, 2012). Such recalibration effects are not unique to sensorimotor or multisensory cues, but can also be observed within-modality (Arnold and Yarrow, 2011).

As predicted by “modality-appropriateness” frameworks, sensory dominance in integration depends strongly on task goals. In addition to task dependence, “default” sensory dominance or bias also differs across individuals, perhaps partly due to experience as dominance can be induced by practice (Sandhu and Dyson, 2012). Individual variety in dominance is reflected in EEG-correlates of multisensory integration [(Giard and Peronnet, 1999) but see (Besle et al., 2009)]: stronger multisensory interactions were found in the sensory cortex of the non-dominant modality. Interestingly, such enhanced integration effects on detection of inputs in the non-dominant modality was also found at the behavioral level (Caclin et al., 2011). In analogy, Romei and colleagues (Romei et al., 2013) found that when separating participants according to their attentional preferences (visual or auditory), differences emerged between these groups as to how sounds influenced visual cortex excitability, with audio-to-visual influences being more prominent in participants with low visual/ high auditory preferences. Could individual variation in canonical operations account for the dominance-related individual differences in multisensory integration? Although this is an open question, it is indeed conceivable that cross-modal PR may be related to individual sensory dominance. For instance, cross-modal effects and sensory dominance in the visual modality share the same underlying brain oscillation, i.e. occipital alpha oscillations (8-14 Hz). This oscillation determines both auditory impact on visual cortex excitability by PR [(Romei et al., 2012); Figure 2B] as well as visual dominance, with low visual performers, or low visually excitable participants, showing high alpha-amplitude (Hanslmayr et al., 2005; Romei et al., 2008). Sound-induced alpha-phase reset in the multisensory setting may therefore have disproportional impact in individuals where alpha activity is high and visual performance low. In addition to PR, DN operations can also be individually shaped. Individual differences in modality dominance may be related to the distribution of “modality dominance weights” in the divisive normalization model. In the model, weights for a certain input channel (e.g. visual, vestibular) are fixed for a certain neuron, but vary across neurons in the same pool (Ohshiro et al., 2011). It seems plausible that how these weights are distributed across neurons is related to an individual's sensory dominance, and is shaped by individual factors such as experience or genetics.

Situations in which sensory context is altered further illustrate the importance of adaptive integrative capacity. A case in which context-dependent flexibility is essential for effective sensory processing is that of changes in the sensory modality in which an object is experienced. For example, the initial meeting of someone might involve seeing and hearing her, but later recognition might be limited to seeing her face in a crowd. We and others have examined this [reviewed in (Thelen and Murray, 2013)] from the standpoint of memory processes and the notion of “redintegration” (Hamilton, 1859) where a part is sufficient to reactivate the whole consolidated experience. Single-trial multisensory experiences at one point in time have long-lasting effects on subsequent visual and auditory object recognition. Recognition is enhanced if the initial multisensory experience had been semantically congruent and can be impaired if this multisensory pairing was either semantically incongruent or entailed meaningless information in the task-irrelevant modality, when compared to objects encountered exclusively in a unisensory context (Lehmann and Murray, 2005; Murray et al., 2005; Murray et al., 2004; Thelen et al., 2012; Thelen and Murray, 2013). EEG-correlates of these effects indicate that incoming unisensory information is

rapidly processed by distinct brain networks according to the prior multisensory vs. unisensory context in which an object was initially encountered (Murray et al., 2004; Thelen et al., 2012; Thelen and Murray, 2013). Distinct sub-portions of lateral occipital cortices responded differently at 100ms post-stimulus onset to repeated visual stimuli depending on whether or not these had been initially encountered with or without a semantically congruent sound. Likewise, this was the case even though the presence/absence of sounds was entirely task-irrelevant, and therefore outside current behavioral goals, and the experience was limited to a single-trial exposure. That is, information appears to be adaptively routed to distinct neural populations perhaps as a consequence of prior DN operations that serve to sharpen sensory representations according to if the initial context was unisensory, a semantically congruent multisensory pairing or meaningless multisensory pairing. For example, semantically congruent pairings may result in a weaker normalizing signal and a more robust object representation that is in turn re-accessed with greater reliability even upon subsequent presentation of a unisensory component of this original experience. More generally, we are inclined to interpret these effects as reflecting multisensory enrichment of the adaptive coding context, which increases context-dependent flexibility of perception. However, further experiments, will be necessary to draw more direct links to canonical operations occurring during multisensory processing and their downstream effects on later unisensory processes. Such notwithstanding, the growing interest in multisensory learning (e.g. (Naumer et al., 2009; Shams and Seitz, 2008)) and long-term effects of multisensory interactions more generally (e.g. (Meylan and Murray, 2007; Naue et al., 2011; Shams et al., 2011; Wozny and Shams, 2011; Zangenehpour and Zatorre, 2010)) is not only opening new lines of basic research, but also strategies for education and clinical rehabilitation (e.g. (Johansson, 2012)).

In summary, experience-related individual differences highlight the flexibility in cross-modal temporal processing and suggest that experience may fine-tune the accuracy of temporal predictions generated either by motor or by cross-modal cues, and possibly sharpen temporal integration windows affected by DN operations. A possible link within individuals between sensory dominance and the effectiveness of phase-reset of ambient visual alpha oscillations by sounds, would argue further for the adaptive nature of canonical integrative processes. These observations suggest that canonical integrative operations may be individually shaped and that this process is plastic, for instance to the context of a first experience as in adaptive coding.

## 6. Can integration also be context-independent?

One might argue that some multisensory processes are less context-dependent than others. For example, auditory-visual interactions seem to reliably occur at early post-stimulus latencies (<100ms) and within low-level cortices irrespective of whether the stimuli are presented passively (Vidal et al., 2008), whether stimuli are task-irrelevant (though attended) (Cappe et al., 2010), whether the task required simple detection (Fort et al., 2002a; Martuzzi et al., 2007; Molholm et al., 2002), or whether discrimination was required (Fort et al., 2002b; Giard and Peronnet, 1999; Rajj et al., 2010). This may suggest that some multisensory phenomena are relatively context-free, but this interpretation is complicated by how terms and conditions are defined. Specifically, in passive presentation, there is no way

to determine whether attention is involved unless there is an extremely demanding task that precludes attending to the “passively” presented stimuli. Causal links between short-latency multisensory processes and behavior have been documented in TMS studies wherein sounds lower thresholds for phosphene induction (Romei et al., 2009, 2013; Romei et al., 2007). Such findings could be interpreted to mean that auditory stimuli can have relevant effects on behavior regardless of task and attention context (see also (McDonald et al., 2013)). In regards to the canonical integration operations, the fact that single, task-irrelevant somatosensory stimuli can cause PR in auditory cortex and enhancement of auditory input processing (Lakatos et al., 2007) suggests that it is in the end *saliency* that determines a stimulus's potency in phase-resetting. While attention often confers saliency, stimuli that are inherently salient such as electrical stimuli applied to the periphery (Lakatos et al., 2007), or TMS applied to the brain, clearly can reset oscillations even when not attended. From this perspective, attention is a major, but not exclusive, determinant of saliency, and PR processes may therefore in some cases be disentangled from task goal and attention context. This might benefit processing of novel inputs; i.e., events that occur at unpredicted and/or unattended dimensions.

## 7. Summary and conclusions

The abundant context effects in multisensory integration, as well as the individual variation and inherent coupling with motor and other top-down cues, underline the need for dynamically adaptive neuronal integration mechanisms. We suggest that canonical neuronal operations for cue integration and predictive interactions, such as divisive normalization and phase-reset mechanisms, are very suitable to explain much of the flexibility in multisensory integration. Because they use the same operations as within-modality cue integration and naturally include motor cues, they reinforce the notion of the “essentially integrative” nature of the brain, or at least, the neocortex. As multisensory integration fits very well in this general framework, there is no need to consider multisensory processes as something unique beyond the notion that different senses may decrease cue uncertainty more than within-modality cues, have stronger predictive power, and can be segregated unambiguously in experimental paradigms. Multisensory contexts might likewise be advantageous for learning and memory, which can be taken as specific examples of more general adaptive coding phenomena.

The suppleness of the brain's use of canonical integration operations exemplifies the brain's flexibility and potential for quickly adapting to the statistics of the environment (Altieri et al., 2013) as well as changes in behavioral goals, which undoubtedly confers huge evolutionary advantages. This is evident in human development and also allows embracing more recent and less natural changes in our environment, such as literacy (Van Atteveldt et al., 2009) or the use of sensory-substitution devices (Bach-y-Rita and W Kerzel, 2003), as well as for tracking changes in the sensory modalities in which environmental objects are experienced (e.g. Thelen and Murray, 2013). The high degree of flexibility and abovementioned capacity for adaptation can be readily extended to cases of focal damage, sensory impairment or loss, as well as more diffuse and presumably less-specialized impairments. Finally, individual differences might indicate the need for and efficiency of tailored interventions for deficits such as dyslexia, autism, or schizophrenia, where

integration of sensory and motor cues, and the process of generating predictions across them, appears disturbed (Blau et al., 2009; Stekelenburg et al., 2013; Stevenson et al., 2014).

## Acknowledgments

This work was supported by the Dutch Organization for Scientific Research (NWO, grant 451-07-020 to NvA), the Swiss National Science Foundation (SNSF, grant 320030-149982 to MMM), the National Center of Competence in Research (project “SYNAPSY – The Synaptic Bases of Mental Disease”; project 51AU40\_125759 to MMM), a Wellcome Trust Investigator Award (098434 to GT) and the US National Institutes of Health (DC011490 to CES).

## References

- Altieri N, Stevenson R, Wallace M, Wenger M. Learning to Associate Auditory and Visual Stimuli: Behavioral and Neural Mechanisms. *Brain Topography*. 2013 Epub ahead of print.
- Alvarado J, Rowland B, Stanford T, Stein B. A neural network model of multisensory integration also accounts for unisensory integration in superior colliculus. *Brain Research*. 2008; 1242:13–23. [PubMed: 18486113]
- Alvarado J, Vaughan J, Stanford T, Stein B. Multisensory versus unisensory integration: contrasting modes in the superior colliculus. *Journal of Neurophysiology*. 2007; 97:3193–3205. [PubMed: 17329632]
- Arnold DH, Yarrow K. Temporal recalibration of vision. *Proceedings of the Royal Society B: Biological Sciences*. 2011; 278:535–538.
- Bach-y-Rita P, W Kercel S. Sensory substitution and the human-machine interface. *Trends in Cognitive Sciences*. 2003; 7:541–546. [PubMed: 14643370]
- Barracough NE, Xiao D, Baker CI, Oram MW, Perret DI. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*. 2005; 17:377–391. [PubMed: 15813999]
- Beauchamp M. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*. 2005; 15:1–9. [PubMed: 15721737]
- Besle J, Bertrand O, Giard M. Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hearing Research*. 2009; 258:143–151. [PubMed: 19573583]
- Blau V, Van Atteveldt N, Ekkebus M, Goebel R, Blomert L. Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Current Biology*. 2009; 19:503–508. [PubMed: 19285401]
- Blau V, Van Atteveldt N, Formisano E, Goebel R, Blomert L. Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *European Journal of Neuroscience*. 2008; 28:500–509. [PubMed: 18702722]
- Bonte M, Valente G, Formisano E. Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations. *Journal of Neuroscience*. 2009; 29:1699–1706. [PubMed: 19211877]
- Buzsaki G, Draguhn A. Neuronal Oscillations in Cortical Networks. *Science*. 2004; 304:1926–1929. [PubMed: 15218136]
- Caclin A, Bouchet P, Djoulah F, Pirat E, Pernier J, Giard M. Auditory enhancement of visual perception at threshold depends on visual abilities. *Brain Research*. 2011; 1396
- Cappe C, Morel A, Rouiller EM. Thalamocortical and the dual pattern of corticothalamic projections of the posterior parietal cortex in macaque monkeys. *Neuroscience*. 2007; 146:1371–1387. [PubMed: 17395383]
- Cappe C, Thelen A, Romei V, Thut G, MM M. Looming signals reveal synergistic principles of multisensory integration. *Journal of Neuroscience*. 2012; 32:1171–1182. [PubMed: 22279203]
- Cappe C, Thut G, Romei V, Murray M. Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *Journal of Neuroscience*. 2010; 30:12572–12580. [PubMed: 20861363]



- Carandini M, Heeger D. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*. 2012; 13:51–62.
- Carandini M, Heeger D, Movshon J. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*. 1997; 17:8621–8644. [PubMed: 9334433]
- Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar A. The natural statistics of audiovisual speech. *PLoS Computational Biology*. 2009; 5:e1000436. [PubMed: 19609344]
- Dahl CD, Logothetis NK, Kayser C. Spatial organization of multisensory responses in temporal association cortex. *Journal of Neuroscience*. 2009; 29:11924–11932. [PubMed: 19776278]
- Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron*. 2008; 57:11–23. [PubMed: 18184561]
- Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*. 2002; 415:429–433. [PubMed: 11807554]
- Falchier A, Clavagnier S, Barone P, Kennedy H. Anatomical Evidence of Multimodal Integration in Primate Striate Cortex. *Journal of Neuroscience*. 2002; 22:5749–5759. [PubMed: 12097528]
- Fetsch CR, DeAngelis GC, Angelaki DE. Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*. 2013; 14:429–442.
- Forster B, Cavina-Pratesi C, Aglioti S, Berlucchi G. Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research*. 2002; 143:480–487. [PubMed: 11914794]
- Fort A, Delpuech C, Pernier J, Giard MH. Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex*. 2002a; 12:1031–1039. [PubMed: 12217966]
- Fort A, Delpuech C, Pernier J, Giard MH. Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research*. 2002b; 14:20–30. [PubMed: 12063127]
- Fries P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Science*. 2005; 9:474–480.
- Friston K. Prediction, perception and agency. *International Journal of Psychophysiology*. 2011; 83:248–252. [PubMed: 22178504]
- Ghazanfar AA, Schroeder C. Is neocortex essentially multisensory? *Trends in Cognitive Sciences*. 2006; 10:278–285. [PubMed: 16713325]
- Giard MH, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*. 1999; 11:473–490. [PubMed: 10511637]
- Gingras G, Rowland B, Stein B. The differing impact of multisensory and unisensory integration on behavior. *Journal of Neuroscience*. 2009; 29:4897–4902. [PubMed: 19369558]
- Girard S, Collignon O, Lepore F. Multisensory gain within and across hemispaces in simple and choice reaction time paradigms. *Experimental Brain Research*. 2011; 214:1–8. [PubMed: 21161190]
- Girard S, Pelland M, Lepore F, Collignon O. Impact of the spatial congruence of redundant targets on within-modal and cross-modal integration. *Experimental Brain Research*. 2013; 224:275–285. [PubMed: 23183636]
- Giraud A, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*. 2012; 15:511–517.
- Hackett TA, De La Mothe LA, Ulbert I, Karmos G, Smiley J, Schroeder CE. Multisensory Convergence in Auditory Cortex, II. Thalamocortical Connections of the Caudal Superior Temporal Plane. *Journal of Comparative Neurology*. 2007; 502:924–952. [PubMed: 17444488]
- Hamilton, W. *Lectures on Metaphysics and Logic*. Boston: Gould & Lincoln; 1859.
- Hanslmayr S, Klimesch W, Sauseng P, Gruber W, Doppelmayr M, Freunberger R, Pecherstorfer T. Visual discrimination performance is related to decreased alpha amplitude but increased phase locking. *Neuroscience Letters*. 2005; 375:64–68. [PubMed: 15664124]

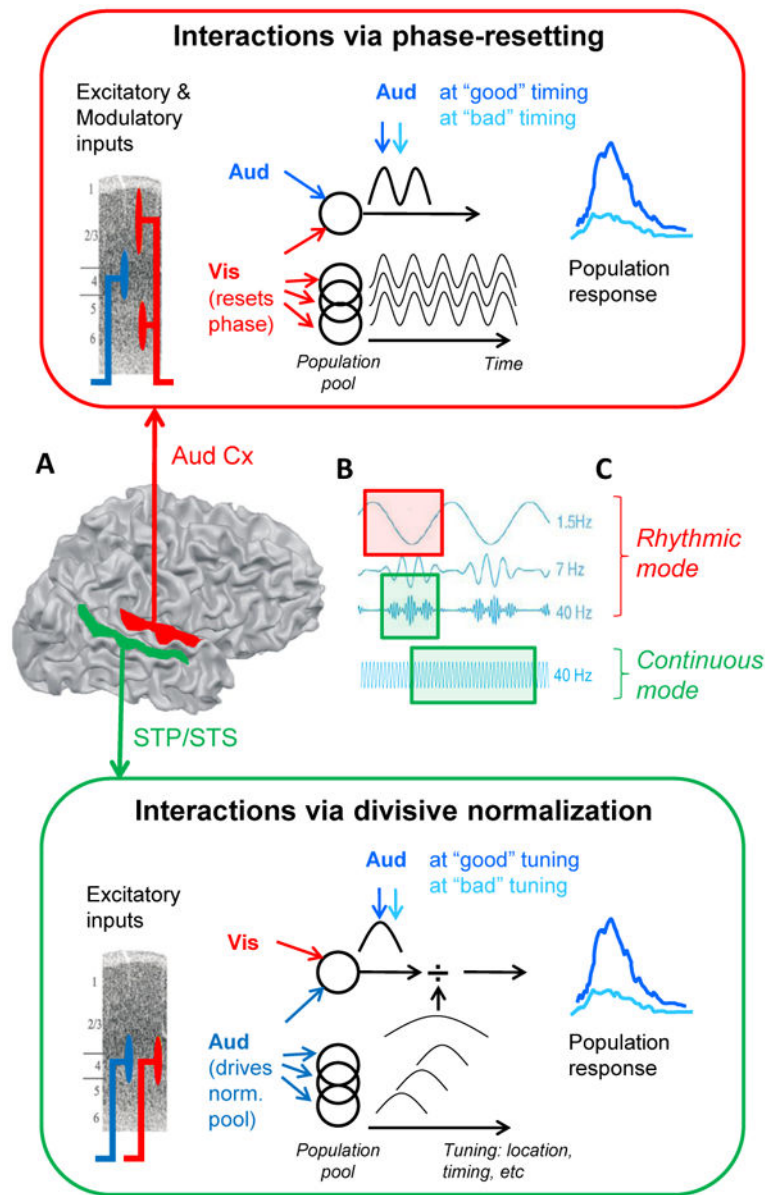
- Hasson U, Ghazanfar A, Galantucci B, Garrod S, Keysers C. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*. 2012; 16:114–121. [PubMed: 22221820]
- Hatsopoulos N, Suminski A. Sensing with the motor cortex. *Neuron*. 2011; 72:477–487. [PubMed: 22078507]
- Hillock-Dunn A, Wallace MT. Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*. 2012; 15:688–696. [PubMed: 22925516]
- Ito J, Maldonado P, Singer W, Grün S. Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes. *Cerebral Cortex*. 2011; 21:2482–2497. [PubMed: 21459839]
- Johansson BB. Multisensory stimulation in stroke rehabilitation. *Frontiers in Human Neuroscience*. 2012; 6:60. [PubMed: 22509159]
- Kayser C, Petkov C, Logothetis N. Visual modulation of neurons in auditory cortex. *Cerebral Cortex*. 2008; 18:1560–1574. [PubMed: 18180245]
- Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*. 2004; 27:712–719. [PubMed: 15541511]
- Kösem A, van Wassenhove V. Temporal structure in audiovisual sensory selection. *PLoS One*. 2012; 7:e40936. [PubMed: 22829899]
- Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*. 2007; 53:279–292. [PubMed: 17224408]
- Lakatos P, Karmos G, Mehta A, Ulbert I, Schroeder C. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*. 2008; 320:110–113. [PubMed: 18388295]
- Lakatos P, Musacchia G, O'Connell M, Falchier A, Javitt D, Schroeder C. The spectrotemporal filter mechanism of auditory selective attention. *Neuron*. 2013; 77:750–761. [PubMed: 23439126]
- Lakatos P, O'Connell M, Barczak A, Mills A, Javitt D, Schroeder C. The leading sense: supramodal control of neurophysiological context by attention. *Neuron*. 2009; 64:419–430. [PubMed: 19914189]
- Lakatos P, Shah A, Knuth K, Ulbert I, Karmos G, Schroeder C. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*. 2005; 94:1904–1911. [PubMed: 15901760]
- Lee H, Noppeney U. Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences USA*. 2011; 108:E1441–1450.
- Lehmann S, Murray M. The role of multisensory memories in unisensory object discrimination. *Brain Res Cogn Brain Res*. 2005; 24:326–334. [PubMed: 15993770]
- Lewkowicz, D. Development of Multisensory Temporal Perception. In: Murray, M.; Wallace, M., editors. *The Neural Bases of Multisensory Processes*. Boca Raton (FL): CRC Press; 2012.
- Lewkowicz D, Ghazanfar A. The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*. 2009; 13:470–478. [PubMed: 19748305]
- Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*. 2007; 54:1001–1010. [PubMed: 17582338]
- Makeig S, Debener S, Onton J, Delorme A. Mining event-related brain dynamics. *Trends in Cognitive Sciences*. 2004; 8:204–210. [PubMed: 15120678]
- Maren S, Phan K, Liberzon I. The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nature Reviews Neuroscience*. 2013; 14:417–428.
- Martuzzi R, Murray M, Michel C, Thiran J, Maeder P, Clarke S, Meuli R. Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cerebral Cortex*. 2007; 17:1672–1679. [PubMed: 16968869]
- McDonald J, Störmer V, Martinez A, Feng W, Hillyard S. Salient sounds activate human visual cortex automatically. *Journal of Neuroscience*. 2013; 33:9194–9201. [PubMed: 23699530]
- Mégevand P, Molholm S, Nayak A, Foxe JJ. Recalibration of the multisensory temporal window of integration results from changing task demands. *PLoS One*. 2013; 8:e71608. [PubMed: 23951203]
- Melloni L, Schwiedrzik C, Rodriguez E, Singer W. (Micro)Saccades, corollary activity and cortical oscillations. *Trends in Cognitive Sciences*. 2009; 13:239–245. [PubMed: 19428286]

- Meylan R, Murray M. Auditory-visual multisensory interactions attenuate subsequent visual responses in humans. *Neuroimage*. 2007; 35:244–254. [PubMed: 17215144]
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research*. 2002; 14:115–128. [PubMed: 12063135]
- Montijn J, Klink P, van Wezel R. Divisive normalization and neuronal oscillations in a single hierarchical framework of selective visual attention. *Frontiers in Neural Circuits*. 2012; 6:22. [PubMed: 22586372]
- Morgan M, DeAngelis G, Angelaki D. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*. 2008; 59:662–673. [PubMed: 18760701]
- Murray M, Foxe J, Wylie G. The brain uses single-trial multisensory memories to discriminate without awareness. *Neuroimage*. 2005; 27:473–478. [PubMed: 15894494]
- Murray M, Michel C, Grave de Peralta R, Ortigue S, Brunet D, Andino S, Schnider A. Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage*. 2004; 21:125–135. [PubMed: 14741649]
- Musacchia G, Schroeder C. Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hearing Research*. 2009; 258:72–79. [PubMed: 19595755]
- Nath A, Beauchamp M. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *Journal of Neuroscience*. 2011; 31:1704–1714. [PubMed: 21289179]
- Naue N, Rach S, Strüber D, Huster R, Zaehle T, Körner U, Herrmann C. Auditory event-related response in visual cortex modulates subsequent visual responses in humans. *Journal of Neuroscience*. 2011; 31:7729–7736. [PubMed: 21613485]
- Naumer MJ, Doehrmann O, Mueller NG, Muckli L, Kaiser J, Hein G. Cortical plasticity of audio-visual object representations. *Cerebral Cortex*. 2009; 19:1641–1653. [PubMed: 19015373]
- Ohshiro T, Angelaki D, DeAngelis G. A normalization model of multisensory integration. *Nature Neuroscience*. 2011; 14:775–782.
- Parise CV, Spence C, Ernst MO. When correlation implies causation in multisensory integration. *Current Biology*. 2012; 22:46–49. [PubMed: 22177899]
- Powers, ARr; Hillock, AR.; Wallace, MT. Perceptual training narrows the temporal window of multisensory binding. *Journal of Neuroscience*. 2009; 29:12265–12274. [PubMed: 19793985]
- Raij T, Ahveninen J, Lin F, Witzel T, Jääskeläinen I, Letham B, Israeli E, Sahyoun C, Vasios C, Stufflebeam S, et al. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *European Journal of Neuroscience*. 2010; 31:1772–1782. [PubMed: 20584181]
- Rajkai C, Lakatos P, Chen C, Pincze Z, Karmos G, Schroeder C. Transient cortical excitation at the onset of visual fixation. *Cerebral Cortex*. 2008; 18:200–209. [PubMed: 17494059]
- Reich L, Maidenbaum S, Amedi A. The brain as a flexible task machine: implications for visual rehabilitation using noninvasive vs. invasive approaches. *Current Opinion in Neurology*. 2012; 25:86–95. [PubMed: 22157107]
- Reynolds JH, Heeger DJ. The normalization model of attention. *Neuron*. 2009; 61:168–185. [PubMed: 19186161]
- Rizzuto D, Madsen J, Bromfield E, Schulze-Bonhage A, Seelig D, Aschenbrenner-Scheibe R, Kahana M. Reset of human neocortical oscillations during a working memory task. *Proceedings of the National Academy of Sciences U S A*. 2003; 100:7931–7936.
- Rockland KS, Ojima H. Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*. 2003; 50:19–26. [PubMed: 14511833]
- Rodriguez E, George N, Lachaux J, Martinerie J, Renault B, Varela F. Perception's shadow: longdistance synchronization of human brain activity. *Nature*. 1999; 397:430–433. [PubMed: 9989408]
- Romei V, Gross J, Thut G. Sounds reset rhythms of visual cortex and corresponding human visual perception. *Current Biology*. 2012; 22:807–813. [PubMed: 22503499]

- Romei V, Murray M, Cappe C, Thut G. Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Current Biology*. 2009; 19:1799–1805. [PubMed: 19836243]
- Romei V, Murray M, Cappe C, Thut G. The Contributions of Sensory Dominance and Attentional Bias to Cross-modal Enhancement of Visual Cortex Excitability. *Journal of Cognitive Neuroscience*. 2013; 25:1122–1135. [PubMed: 23384192]
- Romei V, Murray MM, Merabet LB, Thut G. Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions. *The Journal of Neuroscience*. 2007; 27:11465–11472. [PubMed: 17959789]
- Romei V, Rihs T, Brodbeck V, Thut G. Resting electroencephalogram alpha-power over posterior sites indexes baseline visual cortex excitability. *Neuroreport*. 2008; 19:203–208. [PubMed: 18185109]
- Rowland B, Stanford T, Stein B. A model of the neural mechanisms underlying multisensory integration in the superior colliculus. *Perception*. 2007; 36:1431–1443. [PubMed: 18265826]
- Sandhu R, Dyson B. Re-evaluating visual and auditory dominance through modality switching costs and congruency analyses. *Acta Psychologica (Amst)*. 2012; 140:111–118.
- Schroeder C, Lakatos P. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*. 2009; 32:9–18. [PubMed: 19012975]
- Schroeder CE, Foxe JJ. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research*. 2002; 14:187–198. [PubMed: 12063142]
- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*. 2008; 2:106–113. [PubMed: 18280772]
- Schroeder CE, Smiley JF, Fu KG, McGinnis T, O'Connell MN, Hackett TA. Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *International Journal of Psychophysiology*. 2003; 50:5–17. [PubMed: 14511832]
- Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P. Dynamics of Active Sensing and perceptual selection. *Current Opinion in Neurobiology*. 2010; 20:172–176. [PubMed: 20307966]
- Shah A, Bressler S, Knuth K, Ding M, Mehta A, Ulbert I, Schroeder C. Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cerebral Cortex*. 2004; 14:476–483. [PubMed: 15054063]
- Shams L, Seitz A. Benefits of multisensory learning. *Trends in Cognitive Sciences*. 2008; 12:411–417. [PubMed: 18805039]
- Shams L, Wozny D, Kim R, Seitz A. Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology*. 2011; 2:264. [PubMed: 22028697]
- Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *Journal of Comparative Neurology*. 2007; 502:894–923. [PubMed: 17447261]
- Sperdin H, Cappe C, Murray M. Auditory-somatosensory multisensory interactions in humans: dissociating detection and spatial discrimination. *Neuropsychologia*. 2010; 48:3696–3705. [PubMed: 20833194]
- Stein, BE.; Meredith, MA. *The merging of the senses*. Cambridge, MA: MIT Press; 1993.
- Stekelenburg J, Maes J, Van Gool A, Sitskoorn M, Vroomen J. Deficient multisensory integration in schizophrenia: an event-related potential study. *Schizophrenia Research*. 2013; 147:253–261. [PubMed: 23707640]
- Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynaroski TG, Camarata SM, Wallace MT. Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*. 2014; 34:691–697. [PubMed: 24431427]
- Stevenson RA, Wallace MT. Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*. 2013; 227:249–261. [PubMed: 23604624]
- Striem-Amit E, Cohen L, Dehaene S, Amedi A. Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron*. 2012; 76:640–652. [PubMed: 23141074]
- Talsma D, Senkowski D, Soto-Faraco S, Woldorff M. The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*. 2010; 14:400–410. [PubMed: 20675182]

- Ten Oever S, Sack AT, Wheat KL, Bien N, van Atteveldt N. Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Frontiers in Psychology*. 2013; 4:331. [PubMed: 23805110]
- Thelen A, Cappe C, Murray M. Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *Neuroimage*. 2012; 62:1478–1488. [PubMed: 22609795]
- Thelen A, Murray M. The efficacy of single-trial multisensory memories. *Multisensory Research*. 2013; 26:483–502. [PubMed: 24649531]
- Thorne J, De Vos M, Viola F, Debener S. Cross-modal phase reset predicts auditory task performance in humans. *Journal of Neuroscience*. 2011; 31:3853–3861. [PubMed: 21389240]
- Thorne J, Debener S. Look now and hear what's coming: On the functional role of cross-modal phase reset. *Hearing Research*. 2013 in press.
- Van Atteveldt N, Formisano E, Blomert L, Goebel R. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*. 2007a; 17:962–974. [PubMed: 16751298]
- Van Atteveldt N, Formisano E, Goebel R, Blomert L. Integration of letters and speech sounds in the human brain. *Neuron*. 2004; 43:271–282. [PubMed: 15260962]
- Van Atteveldt N, Roebroek A, Goebel R. Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. *Hearing Research*. 2009; 258:152–164. [PubMed: 19500658]
- Van Atteveldt NM, Blau V, Blomert L, Goebel R. fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. *BMC Neuroscience*. 2010; 11:11. [PubMed: 20122260]
- Van Atteveldt NM, Formisano E, Goebel R, Blomert L. Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *NeuroImage*. 2007b; 36:1345–1360. [PubMed: 17513133]
- van Atteveldt NM, Peterson BS, Schroeder CE. Contextual control of audiovisual integration in low-level sensory cortices. *Human Brain Mapping*. 2013 in press.
- Van der Burg E, Alais D, Cass J. Rapid recalibration to audiovisual asynchrony. *Journal of Neuroscience*. 2013; 33:14633–14637. [PubMed: 24027264]
- van Wassenhove V. Speech through ears and eyes: interfacing the senses with the supramodal brain. *Frontiers in Psychology*. 2013; 4:388. [PubMed: 23874309]
- van Wassenhove V, Grant K, Poeppel D. Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences*. 2005; 102:1181–1186.
- Van Zaen J, Murray MM, Meuli RA, Vesin JM. Adaptive filtering methods for identifying cross-frequency couplings in human EEG. *PLoS One*. 2013; 8:e60513. [PubMed: 23560098]
- Vidal J, Giard M, Roux S, Barthélemy C, Bruneau N. Cross-modal processing of auditory-visual stimuli in a no-task paradigm: a topographic event-related potential study. *Clinical Neurophysiology*. 2008; 119:763–771. [PubMed: 18289935]
- Vroomen, J.; Baart, M. Phonetic Recalibration in Audiovisual Speech. In: Murray, MM.; Wallace, MT., editors. *The Neural Bases of Multisensory Processes*. Boca Raton (FL): CRC Press; 2012.
- Vroomen J, Stekelenburg J. Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience*. 2010; 22:1583–1596. [PubMed: 19583474]
- Welch RB, Warren DH. Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*. 1980; 88:638–667. [PubMed: 7003641]
- Werner-Reiss U, Kelly K, Trause A, Underhill A, Groh J. Eye position affects activity in primary auditory cortex of primates. *Current Biology*. 2003; 13:554–562. [PubMed: 12676085]
- Werner S, Noppeney U. Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *The Journal of Neuroscience*. 2010; 30:2662–2675. [PubMed: 20164350]
- Whittington MA, Traub RD, Jefferys JG. Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature*. 1995; 373:612–615. [PubMed: 7854418]

- Wozny DR, Shams L. Computational characterization of visually induced auditory spatial adaptation. *Frontiers in Integrative Neuroscience*. 2011; 5:75. [PubMed: 22069383]
- Zangenehpour S, Zatorre R. Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia*. 2010; 48:591–600. [PubMed: 19883668]
- Zion Golumbic E, Cogan G, Schroeder C, Poeppel D. Visual input enhances selective speech envelope tracking in auditory cortex at a “cocktail party”. *Journal of Neuroscience*. 2013a; 33:1417–1426. [PubMed: 23345218]
- Zion Golumbic E, Ding N, Bickel S, Lakatos P, Schevon C, McKhann G, Goodman R, Emerson R, Mehta A, Simon J, et al. Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron*. 2013b; 77:980–991. [PubMed: 23473326]
- Zion Golumbic E, Poeppel D, Schroeder C. Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. *Brain and Language*. 2012; 122:151–161. [PubMed: 22285024]



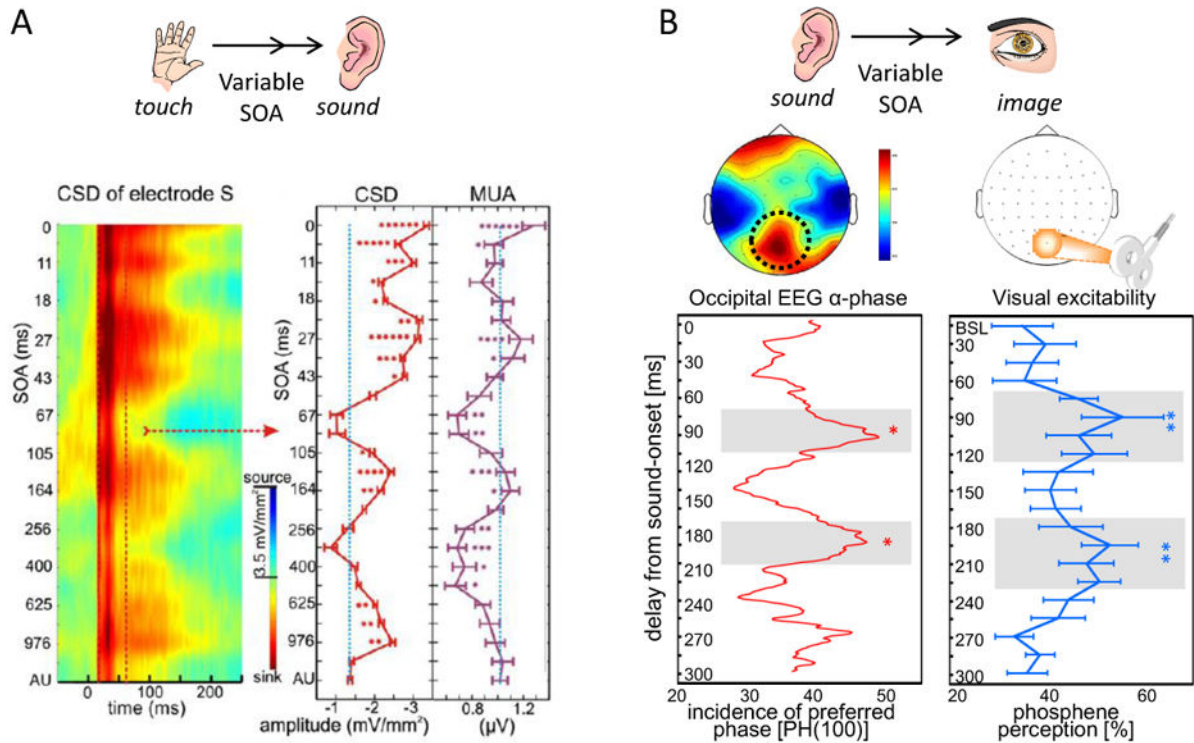
**Figure 1. A schematic representation of the proposed complementary role of canonical integration operations enabling context-dependent integration**  
 Simplified explanation of Phase Resetting (red box) and Divisive Normalization (green box) operations, and how they may complement each other by operating predominantly in different brain areas, time-scales and operation modes.  
**A.** Different brain areas. In low-level sensory cortex, such as primary auditory cortex (A1), cross-modal visual inputs are modulatory (they enter outside cortical layer 4 and do not drive action potentials). By resetting the phase of ambient oscillations in A1, they do change the probability that an appropriately timed excitatory (auditory) input will depolarize neurons above threshold to generate action potentials. It is therefore likely that Phase-Resetting represents a common operation for how multisensory cues interact in low-level sensory cortices. Divisive Normalization models describe interaction of two or more

excitatory inputs. For multisensory integration, this operation seems therefore optimized for brain areas that receive converging excitatory multisensory inputs, such as Superior Temporal Polysensory (STP) area in the macaque monkey (of which the Superior Temporal Sulcus (STS) may be the human homologue).

**B. Different time scales.** PR can occur at all time-scales, but many task-related modulations occur at lower frequencies, such as delta (around 1.5 Hz) and theta (around 7 Hz) (e.g. Schroeder & Lakatos, 2009). The suppressive divisive denominator in the DN operation may in part be mediated by fast-spiking interneurons that produce gamma-range (>30 Hz) oscillations. DN therefore seems appropriate for operating at a fast time-scale.

**C. Different operation modes.** When relevant inputs are predictable in time, the brain assumedly uses a “rhythmic” mode (Schroeder & Lakatos, 2009) where neuronal excitability cycles at low frequencies. PR of these low-frequency oscillations, e.g. by a cross-modal modulatory input, synchronizes high-excitability phases of the oscillations with the anticipated timing of relevant inputs. In the absence of predictable input, the brain is thought to operate in a “continuous mode”. In this mode, gamma-range oscillations are enhanced continuously, along with suppression of lower frequency power to avoid relatively long periods of weaker excitability. As the DN operation likely operates within gamma-cycles it can be used in this mode to continuously facilitate multisensory integration. N.B., in the “rhythmic” mode, gamma amplitude is coupled to the phase of the theta/delta oscillations, so DN may be active during the high-excitability phase of the lower frequency oscillation.

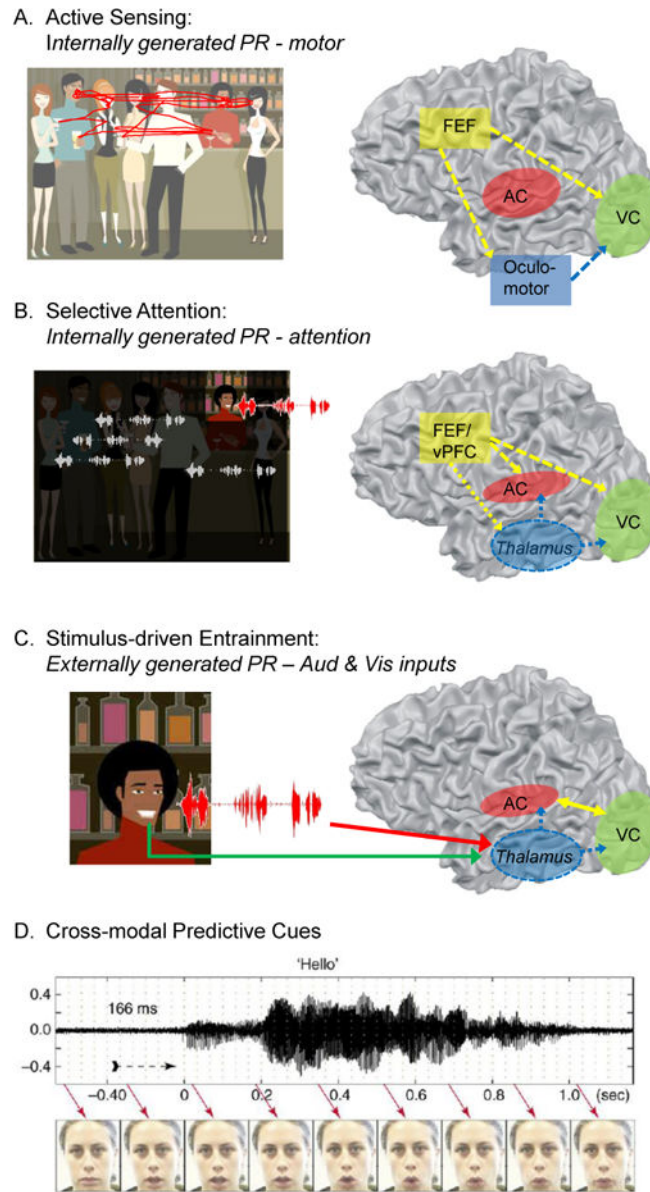




**Figure 2. Evidence in the macaque (A) and human (B) brain for cross-modal phase reset as a mechanism for predictive integration**

**A) Effect of somatosensory-auditory SOA on the bimodal response. (left)** The colormap shows an event related current source density (CSD) response from the site of a large current sink in the supragranular layers of macaque area AI, for different somatosensory-auditory SOAs. CSD is an index of the net synaptic responses (transmembrane currents) that lead to action potentials that lead to action potential generation (indexed by the concomitant multiunit activity, MUA signal) in the local neuronal ensemble. Increasing SOAs are mapped to the y-axis from top to bottom, with 0 on top corresponding to simultaneous auditory-somatosensory stimulation. AU in the bottom represents the auditory alone condition. Red dotted lines denote the 20-80 ms time interval for which we averaged the CSD and MUA in single trials for quantification **(right)** in which we represent mean CSD and MUA amplitude values (x-axis) for the 20-80 ms auditory post-stimulus time interval (error-bars show standard error) with different somatosensory-auditory SOAs (y-axis). Stars denote the number of experiments (out of a total of 6) for which at a given SOA the bimodal response amplitude was significantly different from the auditory. Peaks in the functions occur at ~ SOAs of 27, 45, 114, and 976 msec, which correspond to the periods of oscillations in the gamma (30-50 Hz), beta (14-25 Hz), theta (4-8 Hz) and delta (1-3 Hz) ranges that are phase-reset (and thus aligned over trials) by the initial somatosensory input. As CSD and concomitant MUA increases signify increases in local neuronal excitation, these findings illustrate how the phase reset of ongoing oscillatory activity in A1 predictively prepares local neurons to respond preferentially to auditory inputs with particular timing relationships to the somatosensory (resetting) input. (*Reprinted from Lakatos et al., Neuron 2007*).

**B) Sound-induced (cross-modal) phase locking of alpha-band oscillations in human occipital cortex and visual cortex excitability. (left)** Phase-dynamics in EEG at alpha frequency over posterior recording sites in response to a brief sound (incidence of preferred phase at 100ms post-sound from 0 to 300ms after sound-onset). This EEG alpha-phase dynamics correlated with **(right)** sound-induced cycling of visual cortex excitability over the first 300ms after sound onset as tested through phosphene perception rate in response to single occipital transcranial magnetic stimulation pulses. These findings illustrate co-cycling of perception with underlying perceptually relevant oscillatory activity at identical frequency, here in the alpha-range (around 10Hz) (*Adapted from Romei et al., 2012*). Both A and B support the notion that a sensory input can reset the phase of ongoing oscillations in cortical areas specialized to process another modality, and thereby can facilitate processing at certain periodic intervals and suppress processing at the intervals in-between. With this mechanism, a cross-modal input can reset oscillations to enhance processing specifically at times that relevant input is predicted.



**Figure 3. Different phase-resetting events during a conversation at a cocktail party, and their effects in low-level sensory cortices**

A cocktail party is a good example situation where high flexibility of cue interaction is important for optimal perception and behavior. The rhythmic mode, and hence phase-reset, dominates because of the many rhythmic elements in audiovisual speech. When entering a cocktail party, one first actively explores the scene visually (A). When one speaker is attended (B), the brain's attention system orchestrates the entrainment of ongoing oscillations in low-level sensory cortices to optimally process the relevant speech stream (in red) and visual gestures (person in highlighted square). This guides stimulus-driven entrainment (C), the temporal structure of the acoustic input is being tracked in the auditory cortex (AC), and this process is facilitated by predictive visual cues (D). In parallel, transients in the speech acoustics may also phase-reset oscillatory activity in visual cortex (VC).

**A.** During active visual exploration, eye movements produce internal motor cues that reset low-frequency oscillations in VC to prepare the visual processing system for incoming visual information (Ito et al., 2011; Melloni et al., 2009; Rajkai et al., 2008). The anatomical origins of the motor-related phase-resetting cues are uncertain, but plausible candidates are efference copies from the oculomotor system [pontine reticular formation and/or extraocular muscles, see (Ito et al., 2011)] or a corollary discharge route through the superior colliculus (SC), thalamus and Frontal Eye Fields (FEF), see (Melloni et al., 2009). It is also possible that saccades and the corollary activity are both generated in parallel by attention (Melloni et al., 2009; Rajkai et al., 2008).

**B.** Selective attention orchestrates phase-resetting of oscillations in auditory and visual cortices [e.g. (Lakatos et al., 2008)]. The anatomical origins of the attentional modulatory influence again is not certain, but two plausible candidate mechanisms are cortico-cortical (through ventral prefrontal cortex(vPFC)/FEF) and cortico-thalamic-cortical (reticular nucleus and non-specific matrix) pathways.

**C.** External cross-modal cues can influence processing in low-level sensory cortices by resetting oscillations. Different anatomical pathways are possible for this cross-modal phase-resetting. For example, sensory cortices can influence each other through direct (lateral) anatomical connections [e.g. (Falchier et al., 2002)], or through feedforward projections from nonspecific (Hackett et al., 2007; Lakatos et al., 2007) or higher order (Cappe et al., 2007) thalamic nuclei.

**D.** The cross modal (visual-auditory) phase reset is predictive in that visual gestures in AV speech reliably precede the related vocalizations.

Cocktail party image: iStock. Cross-modal timing figure in D reprinted from Schroeder *et al.*, 2008.