Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin

Functionally relevant low frequency oscillations in multisensory illusion and multisensory response speed facilitation

Dissertation
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von
Georgios Michail, M.Sc.

Berlin, 2021

Erstgutachter: Prof. Dr. Michael Niedeggen
Zweitgutachter: Prof. Dr. Tim Rohe

Tag der Disputation: 02.09.2021

Acknowledgements

First, I would like to express my heartfelt appreciation and gratitude to Prof. Julian Keil for giving me to opportunity to do this project and for his unrelenting support and encouragement at every stage of the thesis. His immediate support at all times and in all sorts of issues and his readiness to provide top-notch scientific advice and engage in energetic insightful discussions has been of paramount importance for completing the thesis and enjoying research. Julian, thank you for being such a tremendous and caring mentor.

Next, I would like to thank Prof. Michael Niedeggen, for enabling this project, for providing critical feedback and monitoring and for encouraging me to present and discuss my research in his research colloquium.

I would also like to thank Prof. Daniel Senkowski for his continuous support throughout the thesis and his rigorous and valuable insights into my work. My sincere thanks go also to my colleagues, James Moran and Mathis Kaiser, for the great advice and assistance, the pleasant working atmosphere and the interesting non-academic discussions. I also want to thank Alex Masurovsky, Amal Sarhan, and Teresa Ramme, for their assistance in acquiring the data for this work. Many thanks go also to my wonderful friends for keeping up my spirits during the difficult stages of the thesis.

I would also like to express my gratitude to my siblings and my parents, Fotini and Savva, for all they have done for me.

Lastly, a special thank you goes to Maria for being an endless source of energy and unwavering support.

Table of Contents

Abbreviations

ACC	anterior cingulate cortex
ANOVA	analysis of variance
BOLD	blood oxygenation level dependent
ECoG	electrocorticography
EEG	electroencephalography
ERP	event-related potential
fMRI	functional magnetic resonance imaging
MEG	magnetoencephalography
SIFI	sound-induced flash illusion
STG	superior temporal gyrus
STS	superior temporal sulcus
TFR	time frequency representation

Abstract

The integration of information across multiple senses is a flexible process shaped by stimulus-driven and contextual influences. Understanding how these influences interact to shape crossmodal processing and the perceptual or behavioral outcome is a major goal of multisensory research. Over the last decades, neuroscientific work promoted a hierarchical framework of multisensory processing, including a dynamic interplay between primary sensory and association cortical areas unfolding at multiple stages. In parallel, behavioral research characterized the extent to which various multisensory phenomena are affected by contextual influences. Despite the accumulating evidence informing these frameworks, there are still important knowledge gaps. To address critical aspects of multisensory perception, that are, to date, poorly understood I used two multisensory phenomena: (i) an established multisensory illusion paradigm, the sound-induced flash illusion (SIFI) in which the integration of a flash together with two rapid beeps can induce the illusory perception of two flashes, and (ii) the crossmodal response speed facilitation, as manifested in the speeding of simple visual responses by concurrent task-irrelevant auditory information.

In the first study, we show that susceptibility to the SIFI is altered when cognitive resources are depleted through a secondary working memory task. This finding suggests that multisensory integration producing SIFI, previously considered a stimulus-driven process, is subject to cognitive resource limitations. The second study, using EEG and a similar design, replicated this finding and extended it by demonstrating a pronounced effect of working memory load on the oscillatory power related to the SIFI. Specifically, SIFI under high load was associated with low frequency oscillations in the theta and beta range unfolding at multiple stages of crossmodal processing. This finding suggests that SIFI, previously linked to gamma oscillations, is an adaptive process that depends on the availability of cognitive resources. Critically, the observed pattern of oscillatory responses is remarkably similar with the literature on an audiovisual speech illusion (McGurk effect), suggesting that low frequency oscillations might reflect general integrative mechanisms. The last study used EEG and ECoG recordings to explore the oscillatory signatures of crossmodal response speed facilitation, for which there is little and inconclusive evidence. We found that crossmodal response speed facilitation is associated with reduced beta power in association areas occurring at early processing stages. Taken together, we provide strong evidence supporting the adaptive nature of multisensory integration in the SIFI and the functional relevance of low frequency oscillations at multiple stages of crossmodal processing.

Zusammenfassung

Die Integration von Informationen über mehrere Sinne hinweg ist ein flexibler Prozess, der von reizgesteuerten und kontextuellen Einflüssen geprägt ist. Zu verstehen, wie diese Einflüsse interagieren, um die crossmodale Verarbeitung und das Wahrnehmungs- oder Verhaltensergebnis zu formen, ist ein wichtiges Ziel der multisensorischen Forschung. In den letzten Jahrzehnten hat die neurowissenschaftliche Arbeit ein hierarchisches Modell der multisensorischen Verarbeitung postuliert, dass ein dynamisches Zusammenspiel zwischen primär-sensorischen und assoziativen kortikalen Arealen beinhaltet, das sich in mehreren Stufen entfaltet. Parallel dazu charakterisierte die Verhaltensforschung das Ausmaß, in dem verschiedene multisensorische Phänomene durch kontextuelle Einflüsse beeinflusst werden. Trotz der Fülle an Erkenntnissen, die diese Modelle untermauern, gibt es immer noch wichtige Wissenslücken. Um kritische Aspekte der multisensorischen Wahrnehmung zu beleuchten, die bis heute nur unzureichend verstanden sind, habe ich zwei multisensorische Phänomene verwendet: (i) ein etabliertes multisensorisches Illusionsparadigma, die sound-induced flash illusion (SIFI), bei der die Integration eines Blitzes zusammen mit zwei schnellen Pieptönen die illusorische Wahrnehmung von zwei Blitzen hervorrufen kann, und (ii) die crossmodale Reaktionsgeschwindigkeitserleichterung, die sich in der Beschleunigung einfacher visueller Reaktionen durch gleichzeitige aufgabenirrelevante auditive Informationen manifestiert.

In der ersten Studie zeigen wir, dass die Anfälligkeit für die SIFI verändert wird, wenn die kognitiven Ressourcen durch eine sekundäre Arbeitsgedächtnisaufgabe erschöpft sind. Dieser Befund deutet darauf hin, dass die multisensorische Integration, welche die SIFI hervorruft und bisher als reiz-gesteuerter Prozess angesehen wurde, Einschränkungen der kognitiven Ressourcen unterliegt. Die zweite Studie, die EEG und ein ähnliches Design verwendete, replizierte diesen Befund und erweiterte ihn, indem sie einen ausgeprägten Effekt der Arbeitsgedächtnisbelastung auf die oszillatorische Power in Verbindung mit der SIFI zeigte. Insbesondere war die SIFI unter hoher Belastung mit niederfrequenten Oszillationen im Theta- und Beta-Bereich assoziiert, die sich in mehreren Phasen der crossmodalen Verarbeitung entfalten. Dieses Ergebnis legt nahe, dass die SIFI, welche zuvor mit Gamma-Oszillationen in Verbindung gebracht wurde, ein adaptiver Prozess ist, der von der Verfügbarkeit kognitiver Ressourcen abhängt. Entscheidend ist, dass das beobachtete Muster der oszillatorischen Antworten eine bemerkenswerte Ähnlichkeit mit der Literatur über eine audiovisuelle Sprachillusion (McGurk-Effekt) aufweist, was darauf hindeutet, dass niederfrequente Oszillationen allgemeine integrative Mechanismen widerspiegeln könnten. In der letzten Studie wurden EEG-

und ECoG-Aufnahmen verwendet, um die oszillatorischen Signaturen der crossmodalen Reaktionsgeschwindigkeitserleichterung zu untersuchen, für die es nur wenige und nicht schlüssige Beweise gibt. Wir fanden heraus, dass die crossmodale Reaktionsgeschwindigkeitserleichterung mit einer reduzierten Beta-Power in Assoziationsarealen verbunden ist, die in frühen Verarbeitungsphasen auftreten. Zusammengenommen liefern wir starke Belege für die adaptive Natur der multisensorischen Integration in der SIFI und die funktionelle Relevanz von niederfrequenten Oszillationen in verschiedenen Stadien der crossmodalen Verarbeitung.

List of original research articles

The dissertation is based on the following research articles:

Michail G, Keil J (2018) High cognitive load enhances the susceptibility to non-speech audiovisual illusions. *Scientific Reports* 8:11530.

Michail G, Senkowski D, Niedeggen M, Keil J (2021) Memory Load Alters Perception-Related Neural Oscillations during Multisensory Integration. *The Journal of Neuroscience*, 41:1505–1515.

Michail G, Senkowski D, Holtkamp M, Wächter B, Keil J (*submitted*) Early beta oscillations in multisensory association areas underlie crossmodal performance enhancement. *The Journal of Neuroscience.*

1 Introduction

In everyday life, we constantly receive a stream of sensory information. The processing and filtering of the sensory input as well as its integration with behavioral goals and prior knowledge is essential for the construction of accurate representations of the environment and the guidance of behavior. The question of how we bind objects from different senses to achieve the unity of perceptual experience has long been the subject of philosophical pursuits, reaching back to Plato and Aristotle. In his treatises ‘Sense and Sensibilia’ (*Περὶ αἰσθήσεως καὶ αἰσθητῶν*) and ‘On the Soul’ (*Περὶ Ψυχῆς*), Aristotle postulates the existence of a higher level perceptual capacity of the soul (what he called the ‘common sense’) which is complex enough to coordinate the five senses and accomplish the binding of heterogenous (i.e., crossmodal) objects into unified percepts (Gregoric, 2007). Leaving aside the phenomenological reflections of Aristotle on perception, the first sporadic experimental reports on crossmodal influences came in late 19th, early 20th century (e.g., Todd, 1912), while the systematic investigation of multisensory perception by means of behavioral psychological methods started at the second half of the 20th century (Alais et al., 2010).

The study of multisensory perception started with the investigation of the consequences of crossmodal influences on behavioural performance, typically the facilitation of response speed and accuracy in multimodal compared to unimodal stimulation (Todd, 1912; Hershenson, 1962; Raab, 1962; Heller, 1982; Marks, 1987). For instance, Herschenson (1962) showed that simple light and sound stimuli are detected faster when presented together than when presented alone. The discovery of multisensory illusions, such as the biasing of sound localization by visual information in the ventriloquist effect (Choe et al., 1975) added another dimension by showing that crossmodal interactions do not only modulate performance but can fundamentally alter the perceptual experience. Critically, the important advances in neuroimaging methods in the last decades allowed the systematic investigation of the neurobiological and neurophysiological substrates of multisensory interactions, adding another layer of complexity to our understanding of perception (Stein et al., 2020). What became gradually but steadily apparent in the course of the last decades of experimental research on multisensory perception is the complexity and multifaceted nature of intersensory interactions (van Atteveldt et al., 2014). The scientific quest to delineate the nuanced interactions between multimodal signals in the multitude of different possible constellations but at the same time the strive to identify general principles and all-encompassing theoretical and quantitative models has been the inspiring and driving force behind the current thesis.

In the following sections, I first provide a short overview of the basic principles underlying multisensory interactions and the most prominent models that are used to provide quantitative characterisation of multisensory interactions at the behavioral and neural level. Then, I outline the hierarchical processing of multisensory information in terms of when and where in the brain crossmodal interactions take place. In the context of the presented hierarchical framework I, subsequently, describe our current understanding as well some open questions regarding (a) the role of oscillatory activity during multisensory processing and (b) the influence of context-related parameters, often termed 'top-down', on multisensory processing. In the last section of the Introduction, I present the objectives of the current thesis.

1.1 Spatiotemporal hierarchy in multisensory processing

Over the last decades, empirical research on multisensory perception employed a vast range of experimental approaches to identify the basic principles and the underlying mechanisms. A common experimental approach is to investigate how performance in one modality, assessed in terms of accuracy, detection, or response speed, is influenced by redundant or complementary information in another modality. A prominent example of this approach is the redundant target effect, in which multisensory stimulation results in faster responses compared with unisensory stimulation (Miller, 1982). Another common experimental approach consists of investigating the bias in the performance or perception in one modality by incongruent information provided in another modality. This approach led to the discovery of several multisensory illusory phenomena, in which crossmodal temporal, spatial or semantical asymmetries produce an illusory percept, qualitatively different than the veridical one. For instance, in the ventriloquist effect, the localization of a sound is biased by spatially incongruent visual stimuli (Choe et al., 1975). Another prominent illusion is the McGurk effect, in which the presentation of an auditory phoneme with incongruent visual lip movements produces the illusory perception of a different auditory phoneme (McGurk and MacDonald, 1976; Alsius et al., 2018). In the sound-induced flash illusion (SIFI), a flash presented together with two rapid beeps results in the illusory perception of a second flash (Shams et al., 2000; Hirst et al., 2020; Keil, 2020). Multisensory illusions have not only been reported in the audiovisual domain. In the rubber hand illusion, when experimenters stroke an attended rubber hand synchronously with the covered and not seen real hand, participants feel the rubber hand as their own (Botvinick and Cohen, 1998). The following sections include an overview of the principles and mechanisms governing multisensory perception. This overview is based mainly on evidence obtained from studies examining crossmodal influences during multisensory performance enhancement and the multisensory illusions described above.

1.1.1 Measures and principles of crossmodal interactions

Behavioral and neural measures of multisensory integration

At the behavioral level, a typical way to characterize multisensory interactions is to assess whether performance, typically accuracy, response speed or stimulus detection, in the multisensory condition is better than the best performance among the unimodal conditions (Stevenson et al., 2014). Several studies investigating the multisensory facilitation of response time (RT), have applied statistical approaches, such as the race model inequality (Miller, 1982), that allow to control whether the speeding of multisensory responses is not just due to statistical facilitation but is the result of multisensory integration. More recently, several studies used drift diffusion models (Ratcliff and Smith, 2004) which have the advantage to incorporate information about both response speed and accuracy and are therefore suitable to model more realistic multisensory scenarios involving a speed-accuracy trade-off (Drugowitsch et al., 2012).

Another important goal of multisensory research was to develop quantitative measures of multisensory interactions that would enable the comparison of findings between different paradigms and studies and the development of quantitative models. One of the first measures that was employed as a metric of multisensory interactions in neural responses was additivity. Based largely on the work of Meredith and Stein on the neuronal responses in superior colliculus (Meredith and Stein, 1983; Stein and Meredith, 1993) the additive model proposes that neural responses to multisensory stimuli reflect multisensory integration or crossmodal enhancement when they are larger than the linear sum of unimodal responses. Thus, a nonlinear enhancement of neural responses, termed ‘superadditivity’, is considered an index of multisensory integration. Since its establishment, the additive model has been employed to demonstrate and quantify crossmodal enhancement in neuronal responses in subcortical structures (e.g., in basal ganglia Nagy et al., 2006) but also at the cortical level in event-related potentials (Giard and Peronnet, 1999; Molholm et al., 2002; Talsma and Woldorff, 2005), oscillatory power (Senkowski et al., 2005, 2006; Mercier et al., 2015), and fMRI BOLD signal (Calvert et al., 2000; Calvert, 2001). The use of superadditivity, even though widespread, has received some criticism over methodological limitations (Teder-Sälejärvi et al., 2002; Stevenson et al., 2014; Colonius and Diederich, 2020) and concerns that the excessive focus on superadditivity might undermine the complexity and plurality of the ways in which crossmodal interactions shape multisensory behavior (Stanford and Stein, 2007). An alternative way to measure multisensory integration in neural responses is by using bistable multisensory paradigms, in which the unimodal stimuli are sometimes integrated and sometimes not (Keil and Senkowski, 2018). For instance, in the SIFI (Shams et al., 2000), one visual flash presented together with two auditory beeps is sometimes perceived as one flash (no illusion, no integration) and sometimes as two (illusion, integration). According to this approach, differences in neural responses between illusory and non-illusory trials reflect neural correlates of multisensory integration. Besides its employment in SIFI studies (Keil et al., 2014; Balz et al., 2016; Michail et al., 2021a), this approach has been used to characterize neural responses reflecting multisensory integration in various multisensory illusions including the visuotactile version of SIFI (Lange et al., 2011), the ventriloquist effect (Kaiser et al., 2021), the McGurk effect (Keil et al., 2012), and the rubber hand illusion (Rao and Kayser, 2017).

Principles and models underlying multisensory processing

Another central aim of research in multisensory perception has been to elucidate the principles and rules that control how we combine signals from different modalities. Based on the influential work of Stein and Meredith on single-cell neuronal responses in cat's superior colliculus (Meredith and Stein, 1983; Stein and Meredith, 1993), three basic rules of multisensory integration were established, that were subsequently used in numerous neural and behavioral multisensory studies. First, the 'temporal rule' states that multisensory integration is more likely when the unimodal stimuli are aligned in time. Second, the 'spatial rule' states that spatial congruence between unimodal signals facilitates multisensory integration. And third, the rule of 'inverse effectiveness' states that multisensory integration is more likely when the response to the unimodal stimuli is weak. Despite offering a clear and simple framework of multisensory integration, aspects of which were confirmed in human studies (e.g., Senkowski et al., 2011), these principles fail to explain multisensory interactions during more complicated tasks in which context-related parameters come into play (Colonius and Diederich, 2020). For instance, a study showed that the visually-driven enhancement of language comprehension is larger at medium and not at the highest level of auditory noise, as it would be predicted by the principle of inverse effectiveness (Ma et al., 2009). To characterize such contextual influences, theoretical frameworks derived from early behavioral studies proposed a number of guiding principles that define the behavioral outcome of crossmodal influences (Welch and Warren, 1980). For instance, the ‘directed-attention’ framework, derived from early evidence of crossmodal shifts in stimulus localization (e.g., Canon, 1971), suggested that selective attention to one modality biases multisensory perception towards the attended modality. Moreover, under the ‘modality precision’ hypothesis, the relative reliability or acuity of the unimodal components of the stimulus is a decisive parameter that defines the dominant modality. Finally, the ‘modality appropriateness’ hypothesis highlights the importance of the varying precision of the different modalities in specific stimulus features such as the higher spatial resolution of vision compared with audition. Welch and Warren (1980) argued that these frameworks are insufficient and proposed their own model in which intersensory biases are the result of the propensity of our perceptual system to process discrepant crossmodal signals based on the assumption that they arise from the same external event. Critically, their model incorporates the role of several features of the stimulus and context in shaping the perceptual outcome of crossmodal influences. Further research in more recent years has provided an increasingly broader understanding on the dependency of multisensory perception on stimulus properties, contextual factors as well as on the spatial, temporal, and semantic relationships between the combined sensory modalities (see reviews: Macaluso and Driver, 2005; Alais et al., 2010; Stevenson and Wallace, 2013; Macaluso et al., 2016).

In an effort to formalize aspects of multisensory interactions, several quantitative models have been proposed. One prominent model suggests that our perceptual system combines signals in a statistically optimal fashion using the maximum-likelihood estimation (Blake et al., 1993; Ernst and Banks, 2002). According to this model, when two signals from different modalities are combined, the integration of these signals is weighted by their relative reliabilities. Naturally, this model incorporates the varying modality-specific precision in specific stimulus features, as proposed by the ‘modality appropriateness’ hypothesis. Thus, by assigning large weight to reliable signals – that is signals with low variability – and small weight to unreliable signals, the maximum-likelihood estimation integrates crossmodal signals in a statistically optimal fashion to obtain the least variable and thus, most reliable multisensory percept. Previous psychophysical studies provide evidence that, in line with the maximum-likelihood model, humans integrate crossmodal signals in a statistically optimal way during visuotactile size estimation (Ernst and Banks, 2002), audiovisual localization (Alais and Burr, 2004) and audiovisual temporal discrimination (Shams et al., 2005b). Moreover, previous research in primates has shown that, also, single neurons in temporal cortex perform reliability-weighted integration of visual and vestibular signals (Fetsch et al., 2012). In that case, the output activity of a single neuron is normalized by the activity of the surrounding neurons following the predictions of an integrative operation termed divisive normalization (Ohshiro et al., 2011).

Despite experimental support from several psychophysical and neural studies, the maximum-likelihood model fails to explain findings when the assumption of crossmodal signals arising from the same external event (i.e., the assumption of a common cause) is violated (Körding et al., 2007). To address this gap, the Bayesian causal inference framework proposes that crossmodal signals are integrated or segregated based on the subjective perception of a common or an independent cause. Besides the sensory reliabilities, the optimal Bayesian observer uses also prior expectations to infer whether crossmodal signals have a common cause and should be integrated (and how they should be integrated) or whether they have independent causes and should be segregated (Körding et al., 2007). This model has been supported by several psychophysical (Wozny et al., 2008; Odegaard et al., 2015; Rohe and Noppeney, 2015a), neuroimaging (Rohe and Noppeney, 2015b) and electrophysiological studies (Cao et al., 2019; Rohe et al., 2019). Despite the considerable progress in the Bayesian psychophysical models of multisensory perception, there is currently little evidence on the neural implementation of the integration of prior information as well as on the temporal evolution of the different influences.

1.1.2 The ‘when and where’ of multisensory processing

According to classical hierarchical models of multisensory processing, the unisensory information of a multisensory object is first processed sequentially in sensory-specific primary and secondary cortical areas before being combined in multisensory convergence or association areas (Mesulam, 1998). As described in the previous section, a prevalent method to characterize multisensory convergence is to identify areas in which the neural response to multisensory stimuli is superadditive, meaning it is larger than the linear sum of the unisensory responses. Previous neuroscientific evidence has established a number of cortical areas with a multisensory convergence profile, including the superior temporal sulcus (STS) and posterior superior temporal gyrus (STG; Calvert et al., 2000; Beauchamp et al., 2004; Barraclough et al., 2005) as well as areas in the parietal cortex (Hyvärinen, 1981; Calvert et al., 2000; Bremmer et al., 2001; Schlack et al., 2005; Molholm et al., 2006).

However, growing evidence of crossmodal influences occurring in lower order unisensory areas during early processing stages challenge this traditional sequential view of multisensory processing, pointing towards a more complex spatiotemporal hierarchy. Several EEG studies in humans have provided evidence of early crossmodal interactions within the first 100 ms after stimulus onset (Giard and Peronnet, 1999; Fort et al., 2002; Molholm et al., 2002; Senkowski et al., 2005; Cappe et al., 2010). Extending these findings, intracranial recordings in primates and humans with a higher spatial resolution than EEG demonstrated early crossmodal influences in primary sensory areas (Ghazanfar et al., 2005; Lakatos et al., 2007; Kayser and Logothetis, 2009; Mercier et al., 2013, 2015). The notion that activity in unisensory cortical areas is sensitive to crossmodal influences is further supported by several fMRI studies showing multisensory influences in the unisensory visual (Macaluso et al., 2000, 2002; Kayser et al., 2007) and unisensory auditory cortex (Calvert et al., 1997; Laurienti et al., 2002; Eckert et al., 2008).

The amounting evidence of ‘paradoxical’ crossmodal influences in areas and latencies traditionally associated with early hierarchical processing led to an ongoing debate regarding the spatiotemporal signatures of multisensory integration (Calvert and Thesen, 2004; Bizley et al., 2016a). The early integration account proposes that crossmodal sensory information are combined at short latencies in sensory-specific cortices. Conversely, the late integration account posits that the integration of crossmodal information involves predominantly higher order association areas at late processing stages. An alternative account proposes that crossmodal interactions can occur at multiple stages of processing in a parallel way and the relative importance of the interactions at each stage depends on the nature of the task. This idea is supported by evidence showing that parietal association areas integrate information according to task relevance and sensory reliability, whereas integration in sensory areas is sensitive to the level of spatial congruence between the unimodal signals (Rohe and Noppeney, 2016). This finding suggests that unisensory and association cortical areas perform distinct computations during crossmodal processing that can differentially influence the perceptual or behavioral outcome. Further support for this idea comes from evidence showing that early multisensory interactions are more often observed in tasks employing simple audiovisual stimuli (Giard and Peronnet, 1999; Fort et al., 2002; Molholm et al., 2002; Senkowski et al., 2005; Mercier et al., 2013) and might reflect early perceptual binding (Bizley et al., 2016b), whereas later crossmodal effects are more often reported in complex tasks such as visual motion or visual object recognition (Kayser et al., 2017; Franzen et al., 2020) and might relate to decision-making processes. Taken together, the current evidence points towards a parallel processing framework in which crossmodal interactions occur at multiple stages of the processing hierarchy and the relative contribution of each stage to the perceptual outcome is defined by stimulus and task-related parameters. Consistent with this notion – on a more mechanistic level – is the proposal that during crossmodal processing our perceptual system adaptively engages certain integrative operations, such as oscillatory phase resetting and divisive normalization, based on the demands posed by the stimuli and the environment (van Atteveldt et al., 2014).

An adjacent debate concerns the cortical pathways mediating the interplay between different modalities, especially at early processing stages. Anatomical studies using retrograde and anterograde neuronal tracing to map anatomical connections in primates and rodents suggest that crossmodal influences could in principle be mediated in three different ways (Cappe et al., 2009). First, crossmodal influences could be conveyed through thalamocortical connections to primary sensory areas from thalamic nuclei specific for other modalities (Budinger et al., 2006; Hackett et al., 2007; Henschke et al., 2015). Second, one modality can directly influence processing associated with another modality through corticocortical projections between unisensory cortical areas (Henschke et al., 2015; Majka et al., 2019). Finally, crossmodal influences could be mediated through connections between multisensory association and unisensory areas (Rockland and Ojima, 2003; Smiley et al., 2007). It has been argued that crossmodal influences occurring at short latencies are mediated by thalamocortical or direct corticocortical feedforward connections that allow fast information transfer while later crossmodal effects reflect predominantly feedback influences from association areas (Cappe et al., 2009; Bizley et al., 2016b). In contrast, other researchers argue that given the extensive reciprocal connections between multisensory and unisensory areas (e.g., Felleman and Van Essen, 1991; Young, 1993), including all types of anatomical connections (feedforward, feedback, lateral), crossmodal information transfer can use several alternative routes, an idea consistent with the notion of a parallel and dynamic processing framework (Foxe and Schroeder, 2005; Kayser and Logothetis, 2007).

1.2 The role of neural oscillations in multisensory processing

Any account of hierarchical multisensory processing would be incomplete without incorporating neural oscillations. Neural oscillations reflect the periodic fluctuation of excitability in neuronal ensembles (Wang, 2010). The brain’s “ability to package information in time” (Buzsáki, 2006) and implement local information processing and efficient interregional communication is largely based on the temporal organization properties of neural oscillations (Buzsáki and Watson, 2012). Consequently, neural oscillations are instrumental in supporting perceptual and cognitive abilities.

Similarly, extensive empirical and theoretical work has established that neural oscillations play an essential role during multisensory perception (Keil and Senkowski, 2018).

As a periodic signal, an oscillation is characterized by its amplitude (*power*), its speed, measured in the number of oscillations per unit time (*frequency*) and the angle of the oscillatory cycle (*phase*). Neural oscillations extracted from mesoscopic and macroscopic brain signals, measured with EEG, MEG or ECoG, are classified in terms of their frequency and typically include the following frequency bands: delta (0.5-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-29 Hz), and gamma (30-150 Hz). Neurophysiological evidence suggests that many aspects of multisensory processing are associated with phase synchronization and power modulations in the various frequency bands (Keil and Senkowski, 2018).

Clearly, mapping cognitive and perceptual processes to specific frequency bands is far from straightforward (Wang, 2010; Herrmann et al., 2016). First, the functional role of oscillations in cognitive and perceptual processes is often contingent on the time frame and the brain area in which they occur. Second, the contribution of oscillations depends largely on the feature of oscillatory activity which is relevant in a certain process (phase, power, frequency). Third, there is ample evidence that certain aspects of cognitive and perceptual processing involve a synergy between different frequency bands (e.g., in cross-frequency coupling, see Jensen and Colgin, 2007; Canolty and Knight, 2010). Therefore, the different cognitive and perceptual processes involve an interplay between different oscillatory mechanisms in the various frequency bands. While acknowledging this complexity, some aspects of multisensory perception have been consistently linked to specific oscillatory mechanisms. In the following paragraphs, I review studies of oscillatory mechanisms in multisensory perception, with a focus on paradigms linked to multisensory illusions and multisensory performance enhancement. This is however by no means an exhaustive review of the literature on multisensory perception.

1.2.1 Crossmodal phase resetting

The phase synchronization of the oscillatory activity within or between cortical areas is considered to play an important role during crossmodal processing by optimizing sensory processing and allowing efficient information transfer (Keil and Senkowski, 2018; Bauer et al., 2020). Oscillatory phase synchronization can be achieved through a range of different mechanisms. In *crossmodal phase resetting*, input from one modality can reset the phase of oscillations in an area processing information from another modality. The resulting alignment of excitability states between modalities is suggested to facilitate crossmodal processing (Bauer et al., 2020). Electrophysiological studies in macaques have demonstrated that visual and somatosensory input could modulate auditory information processing in primary auditory cortex through phase resetting of low frequency oscillations in the delta, theta, and alpha bands (Lakatos et al., 2007; Kayser et al., 2008). In accordance with these findings, ECoG studies in humans using simple audiovisual stimuli showed that phase resetting of low frequency oscillations is involved in the auditory-driven modulation of visual stimulus processing in the visual cortex (delta, theta and alpha band, Mercier et al., 2013), and vice versa in the modulation of auditory stimulus processing in auditory cortex by visual input (delta band, Mercier et al., 2015). Similarly, another ECoG study found that visual speech enhanced auditory speech representations in auditory cortex through phase resetting in the delta band (Mégevand et al., 2020).

1.2.2 Functional connectivity

Interregional communication is assumed to occur also though *phase coherent oscillations*, that is when fluctuations of excitability between two or more areas are synchronized (Bastos and Schoffelen, 2016). Measuring corticocortical phase consistency during the motion bounce illusion (Sekuler et al., 1997), an EEG study demonstrated the emergence of large-scale cortical synchronization in the beta and the gamma band, which was predictive of the individual susceptibility to integrate auditory and visual signals and report a bounce percept (Hipp et al., 2011). Similarly, two MEG studies demonstrated that audiovisual integration in the McGurk effect and the SIFI is influenced by alpha and beta cortical synchronization prior to the stimulation (Keil et al., 2012, 2014). In addition, recent evidence suggests that visuotactile processing might involve long-range synchronization in both the low frequencies (Wang et al., 2019) and the gamma range (Misselhorn et al., 2019). In sum, evidence from phase modulations during multisensory processing indicates that while crossmodal phase resetting is reflected mostly in the low frequencies, large-scale cortical synchronization influencing crossmodal processing can occur in both low and high frequency ranges.

1.2.3 Local cortical activity in multisensory illusion perception

In addition to phase synchronization, research in the multisensory field has shown that oscillatory power modulations in the different frequency bands are implicated in a wide range of crossmodal effects (Keil and Senkowski, 2018). Brain gamma oscillations are characterized by biophysical properties enabling the processing and integration of information at the local cortical level (Tallon-Baudry and Bertrand, 1999; von Stein and Sarnthein, 2000; Buzsáki and Wang, 2012). Consistently, previous multisensory research showed that gamma power is associated with the binding of simple audiovisual stimuli (Senkowski et al., 2005, 2007). In a similar vein, several studies have demonstrated that the integration of a simple flash with two beeps leading to an illusory percept in the SIFI is associated with increased gamma power (Bhattacharya et al., 2002; Mishra et al., 2007; Balz et al., 2016). Gamma power was associated not only with the binding of simple multisensory stimuli but also in multisensory integration during more complex phenomena such as speech and body ownership. For instance, a study found that the integration of audiovisual speech stimuli in the McGurk effect was associated with increased gamma power (Kaiser et al., 2005). Similarly, gamma power changes were also linked to visuotactile integration producing illusory proprioceptive shifts in the rubber hand illusion (Kanayama et al., 2007, 2009). Yet, several other studies have demonstrated that crossmodal processing underlying the McGurk effect and the rubber hand illusion engages activity in the low frequency range, below 30 Hz. Two recent studies have shown that the perception of the McGurk effect is associated with beta power suppression (Roa Romero et al., 2015; Kumar et al., 2016) and theta power modulations (Keil et al., 2012; Roa Romero et al., 2016; Morís Fernández et al., 2018). Regarding the rubber hand illusion, several recent EEG studies demonstrated the involvement of low frequency power modulations in the alpha and beta range during crossmodal influences related with the illusory rubber hand ownership (Evans and Blanke, 2013; Faivre et al., 2017; Rao and Kayser, 2017). Taken together, these findings suggest that in multisensory illusion paradigms, crossmodal binding of simple stimuli is reflected in gamma power, whereas integration in more complicated paradigms, utilizing speech or body ownership, relies on the additional engagement of activity in the low frequencies, presumably linked to top-down control (Keil and Senkowski, 2018). This idea is consistent with a growing consensus that, from an information processing standpoint, bottom-up feedforward processing is reflected in gamma oscillations, while activity in low frequencies facilitates top-down feedback information processing (Fries, 2015).

1.2.4 Local cortical activity in crossmodal performance enhancement

A separate line of empirical research has assessed contributions of oscillatory power in paradigms in which information from one modality enhanced performance in another modality. In an EEG study examining visual motion perception, participants had to identify coherently moving dots presented together with congruent or incongruent moving sound or with a stationary sound (Gleiss and Kayser, 2014a). The accuracy of visual motion detection was enhanced by congruent auditory information and the facilitation in performance was predicted by delta (1-4 Hz) and alpha (8-12 Hz) power, presumably reflecting attention and priming-related mechanisms. In another EEG study conducted by the same authors, the enhancement of visual contrast detection by acoustic noise was related to power modulations in the alpha and beta band (Gleiss and Kayser, 2014b). A small number of other studies examined the role of oscillations in the multisensory facilitation of response speed. An EEG study, in which participants had to make speeded responses following detection of simple visual, auditory, or audiovisual stimuli found a link between shorter RTs and beta power in frontal and occipital electrodes (Senkowski et al., 2006). Moreover, an ECoG study, employing a similar paradigm, found that faster responses where related with increased phase coupling in the delta band between auditory and motor cortex (Mercier et al., 2015). In both studies, however, the correlation of RTs with oscillatory responses was observed in both the multisensory and the unisensory conditions. Therefore, the reported oscillatory effects are not specific for multisensory processing and could reflect motor-related processing. It thus remains an open question whether low frequency oscillations have any functional role at early sensory stages of processing during crossmodal RT facilitation.

With regard to the hierarchical model of multisensory processing, the neurophysiological evidence reviewed in this section suggests that crossmodal influences engage a wide range of oscillatory mechanisms at multiple processing stages and that several factors such as the stimulus complexity or the modality pairings define the adaptive engagement and dynamic interplay between these mechanisms (van Atteveldt et al., 2014).

1.3 Contextual influences on multisensory behavioral and neural responses

The extent to which multisensory processing is sensitive to contextual influences associated with attention or expectations has been a matter of debate (Macaluso et al., 2016). Early behavioral studies on the ventriloquist effect argued that multisensory integration is driven by bottom-up stimulus-driven processing and is not sensitive to selective attention (Bertelson et al., 2000; Vroomen et al., 2001). Moreover, based on evidence from ERP studies showing early multisensory interactions within 100 ms after stimulus onset (Giard and Peronnet, 1999; Fort et al., 2002; Molholm et al., 2002; Cappe et al., 2010; Van der Burg et al., 2011) it has been suggested that early crossmodal influences are bottom-up in nature and largely independent from top-down control (De Meo et al., 2015; ten Oever et al., 2016). Yet, there is growing evidence challenging this view. For instance, evidence of high attentional load diminishing the perception of the McGurk effect (Alsius et al., 2005, 2007) suggests that audiovisual speech integration is sensitive to attentional limitations and does not occur automatically as previously argued (Soto-Faraco et al., 2004). Furthermore, previous ERP studies have shown that early multisensory interactions can be modulated by spatial attention (Talsma and Woldorff, 2005) and selective attention (Talsma et al., 2007). Further evidence implicating top-down control during crossmodal processing comes from fMRI studies showing that multisensory integration effects in STG are sensitive to task demands (van Atteveldt et al., 2007) and that unisensory neural and behavioral weights shaping audiovisual integration in the ventriloquist effect can be modulated by modality-specific attention (Rohe and Noppeney, 2018). Moreover, behavioral and neuroimaging studies from the same group show that spatial attention and expectations can influence multisensory perception by engaging neural systems in frontoparietal and sensory cortical areas (Zuanazzi and Noppeney, 2018, 2019).

Previous research demonstrated that contextual influences on neural responses associated with multisensory processing might be reflected in modulations of neural oscillations, primarily in the low frequency range (Keil and Senkowski, 2018). For instance, previous EEG studies linked power modulations and functional connectivity in the alpha and beta band with influences of modality-specific attention and temporal expectations during a visuotactile detection task (Pomper et al., 2015; Keil et al., 2016). Furthermore, reduced alpha and beta power was associated with modality-specific attention and spatial attention in a visuotactile oddball task (Bauer et al., 2012) and with increased task demands during visuotactile congruence evaluation (Göschl et al., 2015). The engagement of low frequency activity during contextual influences on multisensory processing is consistent with theoretical frameworks proposing that activity in low frequency oscillations is related to the top-down feedback stream of information (Arnal and Giraud, 2012; Fries, 2015) and the optimization of sensory processing through the routing of information to task-relevant regions (Jensen and Mazaheri, 2010).

Interestingly, a recent behavioral study found that while top-down attentional control influenced audiovisual integration of ambiguous stimuli in the motion bounce illusion, it had no effect on a simple audiovisual simultaneity judgement task (Donohue et al., 2015). This finding suggests that the engagement of top-down attentional control in multisensory processing might depend on the level of perceptual uncertainty or stimulus complexity. This notion is consistent with theoretical proposals of multisensory perception relying on a dynamic and adaptive interplay between bottom-up and top-down processing, with the latter being favored when stimulus or context parameters intensify crossmodal competition or perceptual uncertainty (Talsma et al., 2010; ten Oever et al., 2016). Figure 1 provides an illustration of the different scenarios depending on the level of crossmodal competition and perceptual uncertainty.

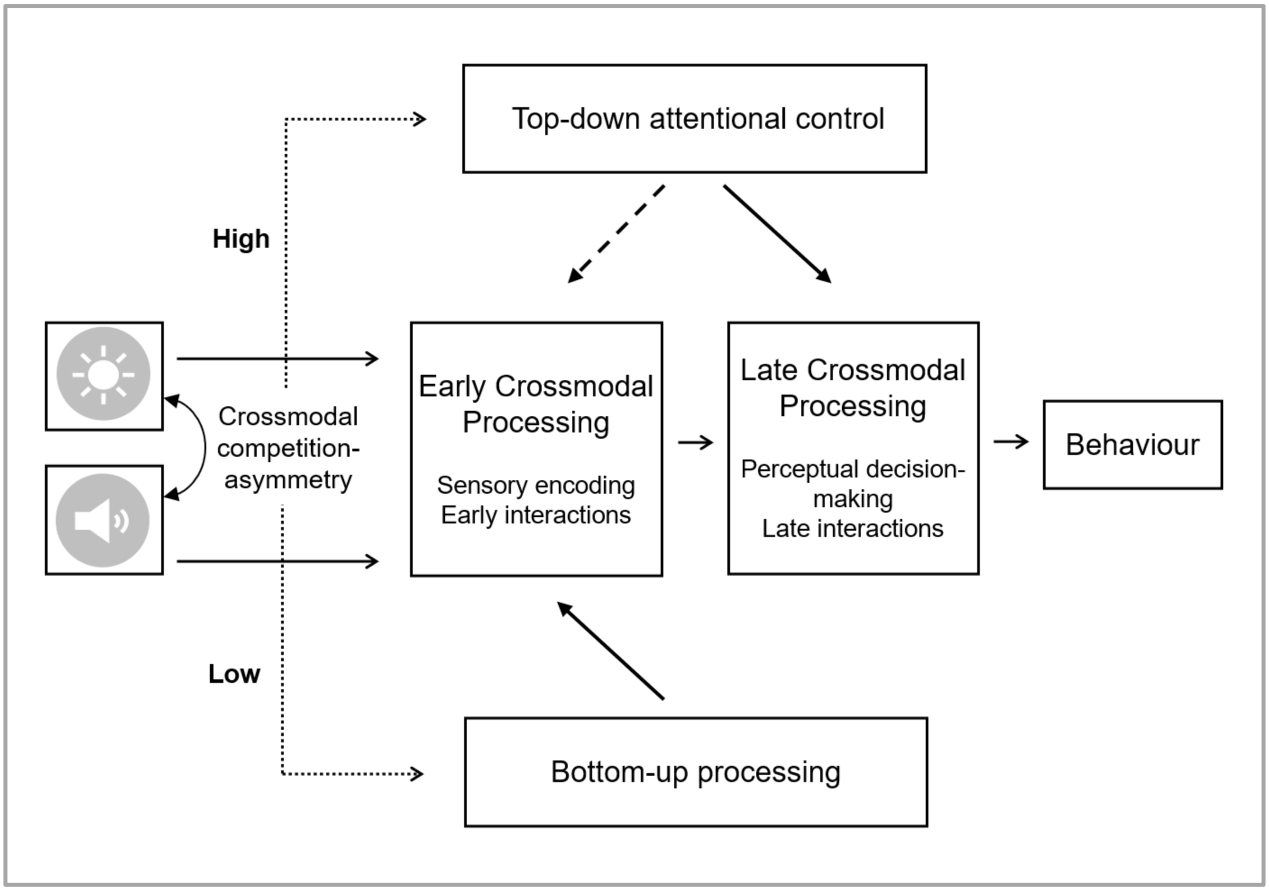

Figure 1. A conceptual model of the interplay between bottom-up and top-down processing in multisensory perception. Dotted lines indicate possible scenarios and dashed lines indicate inconclusive evidence. According to theoretical proposals (Talsma et al., 2010), under increased crossmodal competition and enhanced perceptual uncertainty, multisensory processing might rely on top-down attentional control. Current evidence supports that top-down processing can modulate late crossmodal processing, whereas its influence on early processing stages is a matter of debate. Conversely, multisensory perception under low crossmodal competition relies on bottom-up processing at early stages of processing.

1.4 Aims of thesis

As discussed above, multisensory perception displays a considerable degree of flexibility and adaptability based on the properties of the stimuli and the environment in a given task or situation. In general, the reviewed evidence indicates that in multisensory settings involving stimuli of low complexity and with small crossmodal incongruence, multisensory perception relies mainly on bottom-up processing and is rather automatic. In contrast, multisensory processing in tasks involving more complex stimuli and large crossmodal incongruence as well as environmental conditions enhancing perceptual uncertainty seems to engage top-down attentional control. Critically, this understanding on the duality of our perceptual system is derived from largely independent lines of investigation with very few studies assessing top-down influences in multisensory paradigms involving simple stimuli, which are presumably integrated in an automatic fashion.

Therefore, one critical question, addressed in the current thesis (Study I and II), concerned the extent to which top-down attentional control is engaged when environmental conditions challenge the presumably stimulus-driven nature of multisensory phenomena relying predominantly on bottom-up processing. To address this question, I investigated the sensitivity of behavioral and oscillatory responses linked to the multisensory integration of simple audiovisual stimuli, as manifested in the SIFI, to the amount of available of cognitive resources. Multisensory integration in the SIFI is linked to early crossmodal interactions (Hirst et al., 2020), and gamma band oscillations (Mishra et al., 2007; Balz et al., 2016) indicating a reliance on bottom-up processing. By employing a secondary orthogonal working memory task performed in parallel to the SIFI, I examined whether the assumed automaticity of crossmodal binding in the SIFI is resistant to the depletion of available cognitive resources. According to the proposed model (Figure 1), which is based on previous theoretical proposals (Talsma et al., 2010), the depletion of resources through a working memory task, should enhance the crossmodal competition and perceptual uncertainty, and thus necessitate the recruitment of top-down control mechanisms. If this assumption is valid, then limiting the resources available for multisensory processing in the SIFI will modify the susceptibility to the illusion and crucially, it will lead to the engagement of top-down control mechanisms, modulating the activity of low frequency oscillations.

An overarching question in the current thesis concerns the spatiotemporal dynamics of the oscillatory responses in multisensory perception. As mentioned in the previous sections, a major debate in multisensory research revolves around the question whether early crossmodal influences implicate not only feedforward processing in primary sensory areas, but also feedback processing from sensory association areas. A major goal of the current thesis was to inform this debate by investigating the spatiotemporal profile of neural oscillations involved in two multisensory phenomena: (i) the SIFI performed under varying levels of working memory load (Study II), and (ii) the crossmodal response speed facilitation, as manifested in the speeding of simple visual responses by concurrent task-irrelevant auditory information (Study III). The examination of oscillatory responses in the SIFI under high cognitive load could address critical assumptions of the proposed model regarding the role of top-down control (Figure 1). If the influence of cognitive load on SIFI entails low frequency oscillations at early stages and in association areas this would strengthen the notion that top-down attentional control can influence early crossmodal processing. In addition, the investigation of oscillatory responses in multisensory performance enhancement could provide further insights regarding the hierarchy of crossmodal processing. If, for instance, response speed facilitation relies on early crossmodal influences in primary sensory areas involving high frequency oscillations this would argue for a larger dependence on feedforward processing. Conversely, if early crossmodal processing relates to low frequency oscillations in association areas, this would indicate an involvement of feedback processing at early stages consistent with the extensive recurrent cortical connections and in line with a parallel and dynamic model of crossmodal processing.

Overall, the current thesis assessed whether and under which conditions oscillations in different frequency bands are involved at sensory and non-sensory stages of crossmodal processing during different multisensory phenomena. Accordingly, the findings of the current thesis could provide insights regarding common oscillatory mechanisms engaged in different multisensory settings and thus, extend frameworks on the role of oscillations in the spatiotemporal hierarchy governing multisensory perception.

2 Summary of the dissertation studies

2.1 Study I: High working memory load enhances perception of the SIFI

Despite advances in our understanding about the parameters shaping multisensory perception (Macaluso et al., 2016; ten Oever et al., 2016), little is known about the sensitivity of multisensory integration to varying availability of cognitive resources. Critically, only few studies assessed this question by employing secondary orthogonal tasks to increase demands and therefore limit resources for multisensory processing. Moreover, these studies provided contradicting evidence showing that audiovisual speech integration in the McGurk effect is sensitive to attentional load (Alsius et al., 2005, 2007), whereas visuotactile spatial integration is not affected by working memory load (Zimmer and Macaluso, 2007). Therefore, the first study aimed to investigate, at the behavioral level, whether the integration of audiovisual signals inducing the SIFI is influenced by varying levels of working memory load.

Participants in this experiment performed a dual-task combining the SIFI paradigm with an orthogonal *n*-back task (0-, 1- and 2-back) which was used to manipulate the available cognitive resources prior to the SIFI task. Each trial started with the *n*-back task, in which a letter was presented, and participants were asked to indicate if it matched the letter present *n* trials before. In 0-back trials the target was letter ‘X’. After the *n*-back task, participants were presented with a combination of auditory and visual stimuli. Besides the combination producing the SIFI (1 flash and 2 beeps), the design consisted of several other flash-beep combinations that were used to prevent the formation of expectations and to control the perception in auditory-only (one beep, two beeps), visual-only (one flash, two flashes) and congruent audiovisual stimulation (one flash-beep pair, two flash-beep pairs). To control for potential response biases, the experimental setup incorporated another stimulus combination, which, similar to the SIFI-inducing stimulus combination, consisted of one flash and two beeps, but with the second beep being presented 180 ms after the first one (Mishra et al., 2007). The inclusion of this condition aimed to ensure that any effect of working memory load on SIFI reflected perceptual changes.

Multisensory integration in the SIFI trials (one flash, two beeps) was assessed in terms of the individual susceptibility to perceive the crossmodally integrated percept of two flashes (i.e., the illusion rate). A one-way repeated measures ANOVA was used to examine whether working memory load had any effect on illusion rates or response times. We found that perception of the SIFI was higher with increasing working memory load (Figure 2). Further analysis showed that this effect was not confounded by response bias. Moreover, working memory load had no effect on response times. These findings indicate that crossmodal integration of audiovisual stimuli producing the SIFI perception is subject to the availability of cognitive resources. More specifically, our data show that when cognitive resources are depleted by high working memory load, participants are more likely to integrate simple audiovisual signals and report an illusory percept.

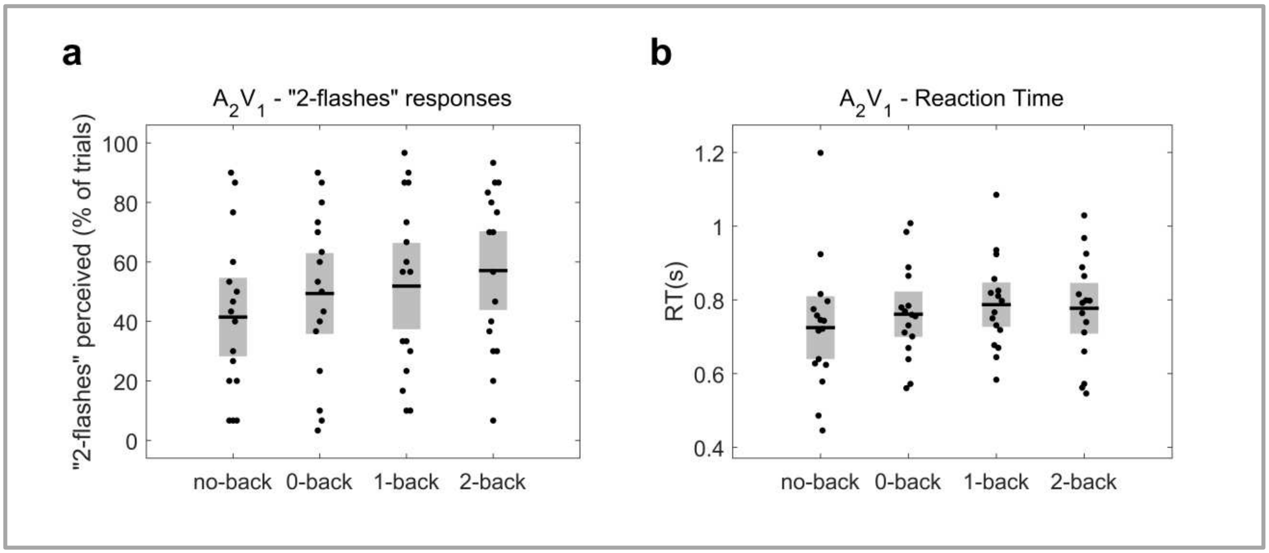

Figure 2. High working memory load induces increased SIFI illusion rates. (a) Illusion rate (i.e., the percentage "2-flashes" responses in the critical "one flash-two beeps" trials) for the different working memory levels. A one-way ANOVA revealed that working memory load had a significant effect on the illusion rates. Post-hoc comparisons showed that illusion perception in 2-back was higher compared to no-back and 0-back. (b) Working memory load had no effect on the RTs in the "one flash-two beeps" trials. Horizontal black lines denote the mean and grey bars the standard error of the mean. This figure corresponds to Figure 2 in the published manuscript by Michail et al., 2018.

These findings are consistent with studies on the McGurk effect demonstrating that audiovisual speech integration is sensitive to increased task demands. Interestingly, while Alsius and colleagues (2005, 2007) report a decrease in the McGurk effect under increased demands, our data reveal the opposite effect for the SIFI, meaning an enhancement of illusion perception under high cognitive load. This discrepancy can be explained by differences in the characteristics of two paradigms. During speech, which relies predominantly on sound, the depletion of resources might result in a larger sensory weight for auditory information, which in turn causes the reduced perception of the McGurk effect. Conversely, as the SIFI relies on temporal discrimination, for which audition is more precise, the exhaustion of resources could increase the sensory weight attributed to the sound, thus producing higher illusion susceptibility. Overall, the first study showed that integration of simple non-speech audiovisual signals is modulated by working memory load and thus, is not a hardwired automatic process unaffected by contextual influences.

2.2 Study II: Low frequency oscillations at multiple stages reflect the influence of working memory load on audiovisual integration in SIFI

The first study showed that the susceptibility to the SIFI is influenced by working memory load. This finding indicates that SIFI is modulated by the available cognitive resources and leads to the question regarding the neural responses associated with this effect. Therefore, in the second study we used EEG recordings and a similar experimental design as the first study and sought first, to replicate the findings of the first study and second, to explore the oscillatory power modulations associated with the effect of working memory load on multisensory integration during the SIFI. As reviewed above, previous evidence indicates that integration of audiovisual signals in SIFI is related to gamma band oscillations, presumably reflecting bottom-up processing (e.g., Mishra et al., 2007). Here, we probed whether, consistent with theoretical proposals (Talsma et al., 2010), a reduction of available resources, induced by high working memory load, would result in the recruitment of top-down control reflected in power modulation of low frequency oscillations.

The experimental design was similar to the first study with the difference of including only two levels of working memory load (0-back/low load, 2-back/high load). The analysis was performed on the trials producing the SIFI (one flash, two beeps) and included a 2 x 2 ANOVA on the spectral power after the SIFI stimulation with factors Load (0-back/low load, 2-back/high load) and Perception (no illusion, illusion). If working memory load has any impact on power modulations associated with the SIFI, this should manifest in the interaction between Load and Perception. In addition, the analysis pipeline included source localization of the interaction effects in order to gain more insight about the origins of these effects.

The behavioral analysis replicated the findings of the first study, showing increased illusion rates under high compared with low working memory load. In addition, we found that participants who were less accurate in the *n*-back task displayed a larger load-induced increase in the illusion rate. Critically, memory load also affected oscillatory power reflecting integrative processing in SIFI, as revealed by the interaction between memory load and the illusion perception (Figure 3). The interaction consisted of three distinct effects in: (i) early beta band power (18-26 Hz, ~70 ms) localized in left auditory and motor cortical regions, (ii) mid-frontal theta band power (7-9 Hz, ~120ms), localized in the mid-anterior cingulate cortex and (iii) in later beta band power (13-22 Hz, ~350ms), localized in regions of the prefrontal and auditory cortex, including STG. Based on the spectral and spatiotemporal characteristics of the interaction and the insights obtained by the post-hoc analysis of power we argue that audiovisual integration in SIFI under high load involves three stages. First, early beta power suppression in an auditory-motor network mirrors the detection of crossmodal mismatch, which is enhanced by high working memory load. Subsequently, increased theta power in cingulate cortex reflects the need for enhanced top-down control in the face of high perceptual uncertainty. Finally, the late beta power suppression in frontal and auditory cortex, including STG, is presumably linked to the engagement of top-down attentional control to facilitate late integrative processing in association cortex. Remarkably, this pattern of oscillatory responses is highly consistent with oscillatory activities associated with the integration of audiovisual speech stimuli in the McGurk effect (Roa Romero et al., 2015, 2016; Morís Fernández et al., 2018).

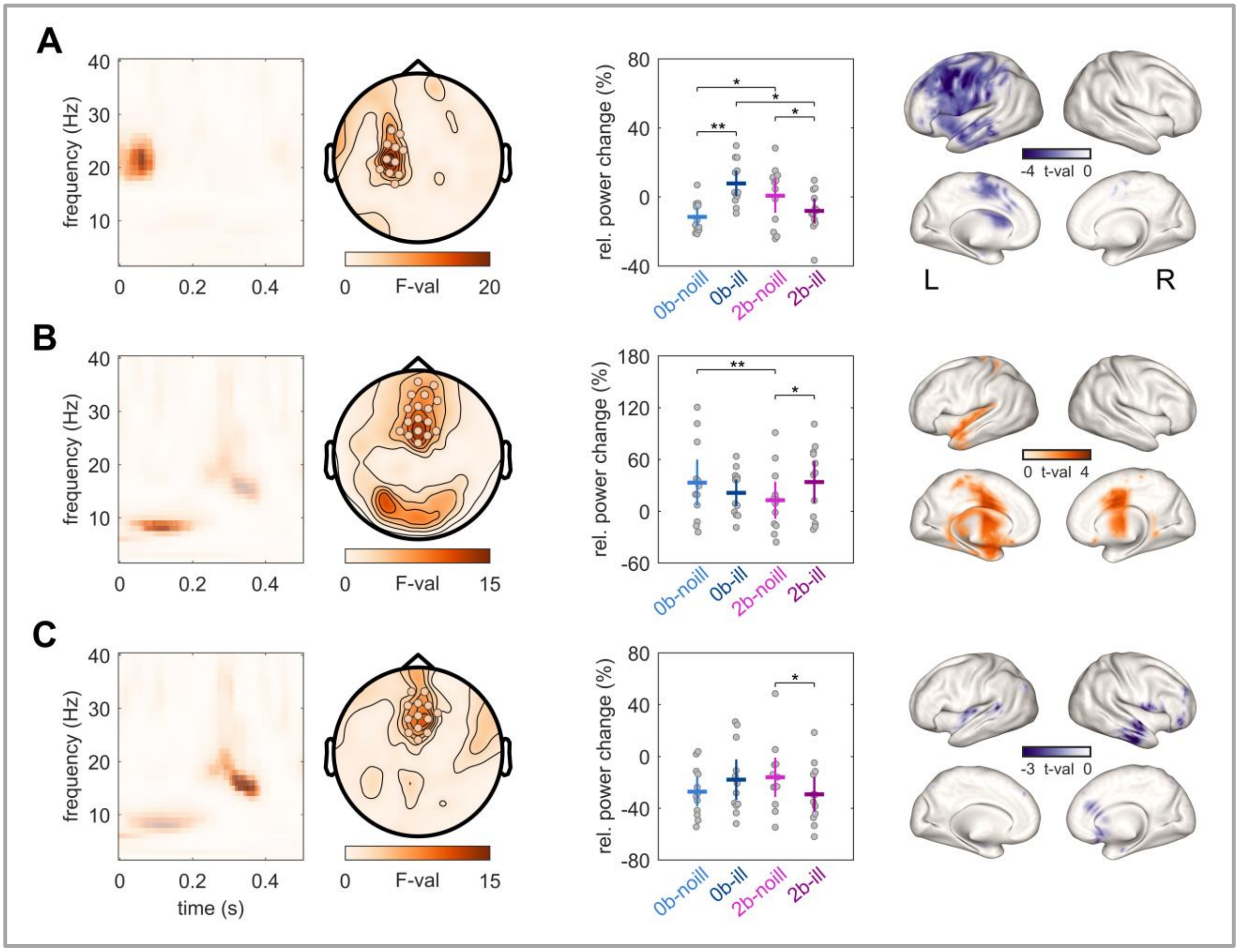

Figure 3. Interaction between working memory load and illusion perception. Each row represents a cluster of interaction between Load (0-back/low, 2-back/high) and Perception (no illusion, illusion). Three interaction effects were observed. **A** The first interaction effect was found in early beta power (18–26 Hz; 0 to 90 ms) over left fronto-central electrodes. Source analysis identified a corresponding illusion-dependent activity difference in the left motor and auditory cortex. **B** The second interaction cluster was observed in frontal theta power (7–9 Hz; 30 to 200 ms). Source analysis identified higher illusion-dependent theta activity in 2-back vs. 0-back, in middle and anterior cingulate cortex. **C** The third interaction was found in frontal beta power (13–22 Hz; 250 to 380 ms). Source analysis showed significant 2-back vs. 0-back differences in the right PFC, ACC, and bilateral temporal regions. *Left panels:* TFRs of significant interactions averaged across the cluster's electrodes and a topography plot showing the spatial distribution and the contributing electrodes (dots). *Middle panel*s: Post–hoc comparisons of the average power in the cluster between the four conditions.

Horizontal lines denote the mean and verticals the standard error of the mean. *Right panels:* Source contrast for power modulation differences associated with illusion perception ($\Delta_{\text{illusion-noillusion}}$) between 2-back and 0-back. *$p < 0.05$, **$p < 0.01$. This figure corresponds to Figure 5 in the published manuscript by Michail et al., 2021.

The results of the second study have three important implications. First, the finding of working memory load modulating behavioral and neural responses associated with SIFI extends the results of the first study, in supporting that integration of simple audiovisual stimuli is not immune to cognitive load. Second, the findings of the second study suggest that integration of audiovisual stimuli in SIFI under reduced cognitive resources relies on the involvement of low frequency oscillations in the theta and beta band, unfolding at multiple processing stages. Third, the striking similarity between the observed pattern of oscillatory responses and the spectral signatures of the McGurk effect indicates that low frequency oscillations are not only recruited to bind audiovisual incongruent speech stimuli but might constitute general integrative mechanisms recruited when stimuli or context enhance crossmodal competition or perceptual uncertainty.

2.3 Study III: Crossmodal facilitation of response times relates to early beta power in association cortex

Importantly, the second study showed that crossmodal processing in SIFI under high cognitive load is associated with low frequency oscillations at both early and late stages of processing. Specifically, SIFI under high load engaged beta oscillations at early stages of crossmodal processing. In the third study we investigated whether early crossmodal influences and low (vs. high) frequency oscillations are also relevant in another paradigm, the multisensory performance enhancement. To this end, the third study used EEG and ECoG data, obtained independently, to investigate the neural correlates of crossmodal response speed facilitation. While the crossmodal facilitation of response times (RTs) is well-documented, the underlying neural correlates are poorly understood. Most of the early ERP studies examining the crossmodal RT facilitation have not associated neural responses related to multisensory interactions with individual performance gains (e.g., Fort et al., 2002; Molholm et al., 2002). Moreover, the oscillatory responses associated with RT facilitation in the few studies testing such relationships, most likely reflect motor-related processing which is not specific to multisensory perception (Senkowski et al., 2006; Mercier et al., 2015). To address this gap in the literature, the last study of the thesis examined the oscillatory correlates of crossmodal RT facilitation, as manifested in the speeding of simple visual responses by concurrent task-irrelevant auditory information.

The EEG data analysis for the third study used the dataset obtained in the second study and focused on the visual-only (one flash) and audiovisual (one flash, one beep) trials. Note, that we used only trials from the 0-back condition assuming no effect of the 0-back task on the processing of the audiovisual stimuli. The ECoG data were obtained through intracranial electrodes implanted for presurgical monitoring in four patients with drug-resistant focal epilepsy. The patients performed the SIFI task, as described in the first study, but without the secondary working memory task. For the purposes of the third study, only the visual-only (one flash) and audiovisual (one flash, one beep) trials were analyzed.

In the first step of the analysis, individual RT gains after audiovisual compared with visual-only stimulation were correlated with the corresponding difference in oscillatory power. This analysis was applied on the EEG data and followed a data-driven approach based on the clustering of correlation values in time, frequency, and sensor space. Based on the results of the correlation analysis in sensor space, source localization was applied to provide further information about the exact cortical sources. Next, we leveraged the high spatial resolution of the ECoG data and compared oscillatory power between audiovisual and visual-only trials. This step aimed to provide further information, at the individual level, regarding the spatial profile of oscillatory power modulations associated with the crossmodal RT facilitation.

The correlation analysis of the EEG data revealed that larger individual benefits in response speed in the audiovisual compared with the visual trials were associated with reduced early beta band power (13-25 Hz, 80-200 ms) in lower parietal and extrastriate visual cortical areas (Figure 4). Interestingly, while theta power was increased in audiovisual compared with visual trials, this power modulation was not related to individual performance gains. Extending this finding, the analysis of the ECoG data revealed a suppression of beta power in STG in audiovisual compared with visual trials, starting approximately 150 ms after stimulus onset. Notably, this effect was consistent across all participants in the ECoG experiment (n=4). Taken together, these data indicate that crossmodal RT facilitation is related to beta power suppression in association areas occurring at relatively early stages of sensory processing.

These results suggest that the facilitation of visual responses by auditory information relies on a sharpening of early sensory processing linked to beta power in association areas. The facilitation of visual responses by beta power might reflect an attention-dependent routing of information at early stages of processing. This notion is in agreement with the role of low frequency oscillations in optimizing information processing in task relevant areas (Jensen and Mazaheri, 2010) as well with several ERP studies showing that crossmodal influences can occur at early processing stages (Giard and Peronnet, 1999; Fort et al., 2002; Van der Burg et al., 2011). Previous evidence on the oscillatory correlates of crossmodal RT facilitation is sparse and inconclusive regarding the involvement of crossmodal influences at early, sensory stages of processing (Senkowski et al., 2006; Mercier et al., 2015). As such, the third study addressed a critical gap in the literature by showing the involvement of beta power at early sensory stages during the crossmodal facilitation of response speed.

In the framework of the current thesis, the findings of the third study highlight that the functional importance of low frequency oscillations, especially in the beta band, is not only restricted to the top-down facilitation of multisensory integration but extends to the crossmodal processing underlying multisensory performance enhancement.

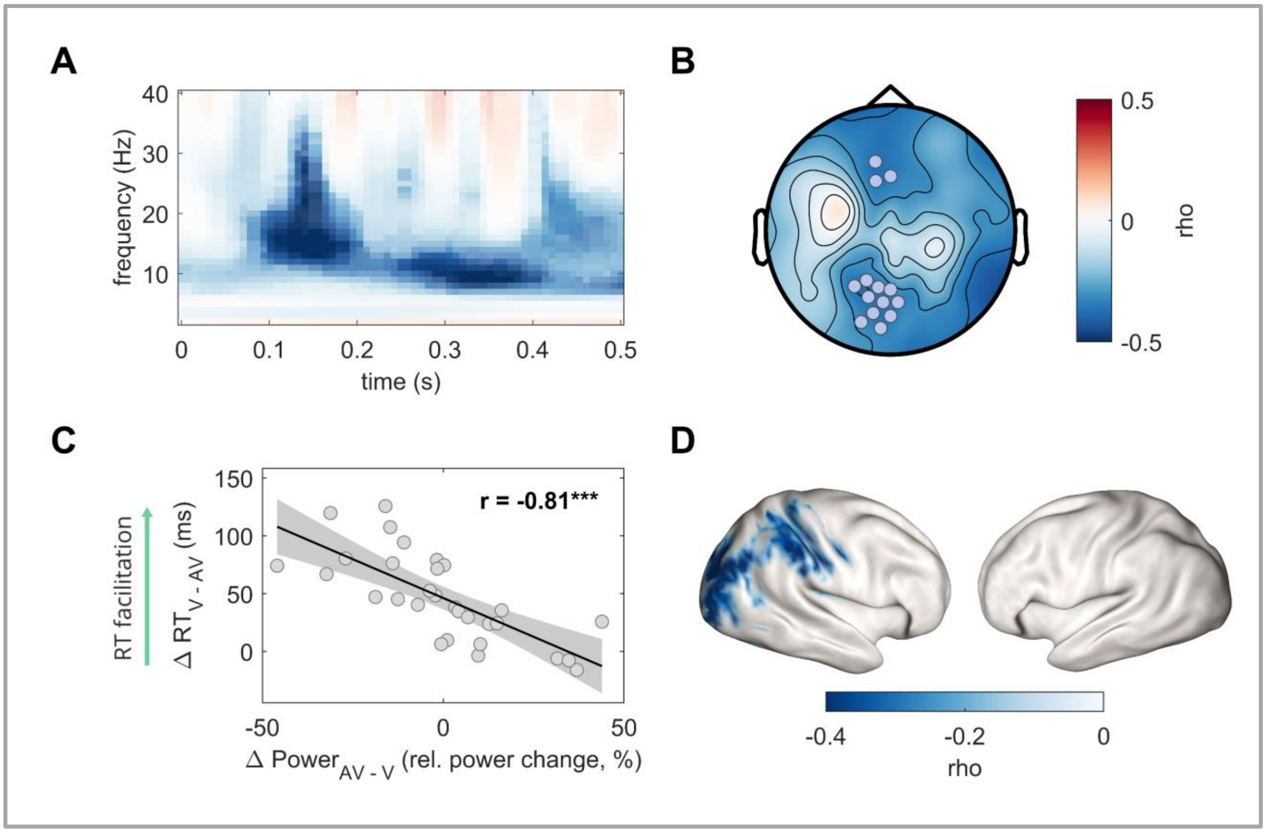

Figure 4. Correlation analysis between 'audiovisual (AV) – visual (V)' power difference and the crossmodal response speed facilitation. The correlation analysis revealed that crossmodal facilitation of RTs was associated with lower early beta power (80-200 ms), localized in right extrastriate visual and inferior parietal areas. **A.** TFR of the correlation between the AV-V power difference and the V-AV RT change, averaged across channels with the highest contribution to the cluster. Lower values indicate that crossmodal RT facilitation correlates with smaller AV-V power difference. **B.** Topographic map showing the distribution of the correlation between AV-V power change and the crossmodal RT facilitation. Channels with the highest contribution to the cluster are highlighted with dots. **C.** Scatterplot depicting the correlation between the individual A-V power difference in the cluster and the crossmodal RT facilitation. The lower the power in the cluster for the AV vs. V condition, the larger the crossmodal RT facilitation. Black lines represent the best-fitting linear regression and shaded areas the 95% confidence interval. **D.** Correlation in source space between the early beta band power difference (AV-V, 80-200 ms, 13-25 Hz) and the crossmodal response speed facilitation. Lower AV-V beta power in extrastriate visual and inferior parietal areas correlated with crossmodal response speed facilitation.; *** $p < 0.001$

3 General Discussion

The current thesis employed the SIFI paradigm performed under various levels of working memory load and the crossmodal response speed facilitation in order to address certain research objectives. First, I aimed to examine whether the supposedly automatic integration of stimuli causing the SIFI is sensitive to limitations of cognitive resources. Regarding this goal, the first two studies showed that susceptibility to the SIFI is modulated under conditions of depleted cognitive resources, suggesting that integration of simple audiovisual stimuli is sensitive to cognitive load. An adjacent objective, addressed by the second study, was to elucidate the oscillatory signatures of the effect of memory load on integrative processing during the SIFI. In particular, we were interested to examine whether low frequency oscillations reflecting top-down processing are engaged in integrative crossmodal processing when resources are depleted. Accordingly, the second study, using EEG, revealed a pronounced effect of load on crossmodal processing associated with SIFI, unfolding at multiple stages, and involving low frequency oscillations in the theta and beta range. A third objective of the thesis was to examine the oscillatory correlates of crossmodal response speed facilitation which are, to date, poorly understood. To that end, the third study used EEG and ECoG data and demonstrated that beta suppression in association cortex at early stages of processing is related to individual RT benefits in visual responses driven by concurrent task-irrelevant auditory information. Figure 5 provides a summary of the oscillatory effects observed in the current thesis with respect to the hierarchical model of multisensory processing.

Taken together, the significance of the research in the current thesis lies on three major contributions. First, by demonstrating and replicating the sensitivity of audiovisual integration producing the SIFI to manipulations of working memory load, the current thesis indicates that the binding of simple audiovisual stimuli, previously considered stimulus-driven and automatic, is subject to cognitive resources limitations. Second, the finding that the SIFI, previously linked to bottom-up gamma oscillations, relies on low frequency oscillations when cognitive resources are depleted, constitutes an important addition to hierarchical models of multisensory processing. Finally, the finding that crossmodal facilitation of response speed relates to beta power in association cortex affecting early sensory processing addresses a critical gap in the literature on multisensory performance enhancement and suggests a diverse role of beta oscillations in multisensory perception. The implications of these contributions are discussed in the following paragraphs.

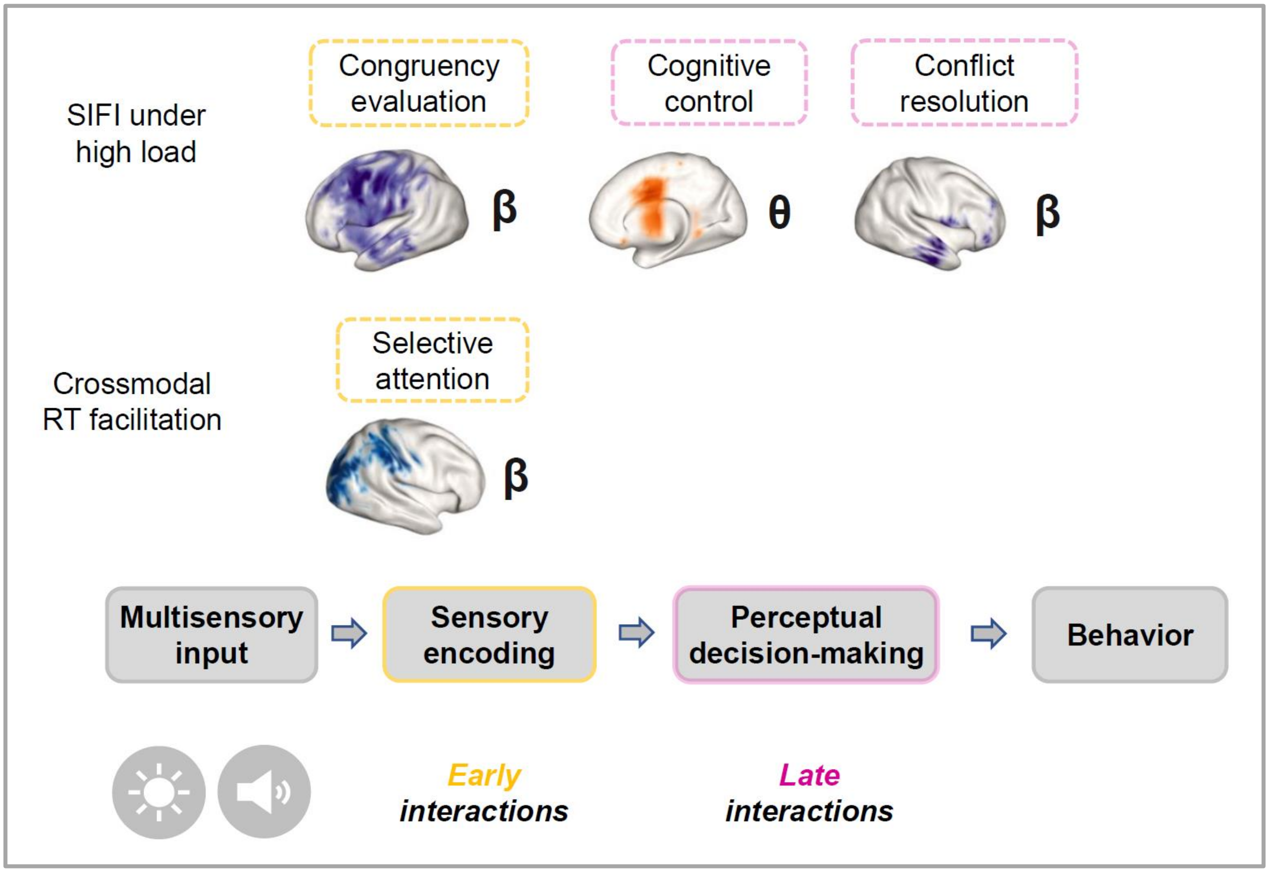

Figure 5. Overview of the oscillatory effects in the present thesis. The identified oscillatory power modulations are positioned on the left-right axis, representing the hierarchy of crossmodal processing, relative to the latency of each effect. The oscillatory responses in the first row correspond to the findings of the second study (Figure 3), while the responses in the second row correspond to the findings of the third study (Figure 4). The brain maps show the source localization of the corresponding effects observed at the scalp level.

3.1 Cognitive resources influence multisensory integration

The extent to which contextual influences affect multisensory perception has been a matter of intense debate. Contrary to earlier reports suggesting that the binding of crossmodal signals is purely stimulus-driven and impervious to manipulations of attentional control, growing evidence advances a dynamic framework incorporating contextual influences in a wide range of multisensory settings (Macaluso et al., 2016). Similarly, models of multisensory integration such as the maximum-likelihood estimation, are based solely on the stimulus properties and largely overlook the influences of context. In contrast, the contribution of contextual factors is incorporated in the Bayesian inference model, as the estimation of sensory reliability is influenced by task properties and prior knowledge. An important challenge emerging from this dynamic framework, is to obtain a nuanced understanding of the role of specific contextual influences within and across different multisensory phenomena. While numerous studies focused on manipulations of spatial and selective attention only very few examined the influence of the availability of cognitive resources on multisensory perception. It is exactly this knowledge gap that the current thesis tried to address. We used the SIFI paradigm as a proxy of multisensory integration and probed its resistance to limitations in cognitive resources achieved through a secondary orthogonal working memory task. Regarding that objective, we found and replicated that the susceptibility to the SIFI is enhanced under high working memory load.

An important implication of this finding relates to the assumptions associated with the integration of audiovisual stimuli in the SIFI. Previous human EEG and MEG studies using evoked responses showed that SIFI is characterized by early crossmodal influences (Shams et al., 2005a; Mishra et al., 2007). Moreover, there is ample evidence that the binding of audiovisual signals in SIFI is related to gamma band oscillations, presumably reflecting bottom-up processing (Bhattacharya et al., 2002; Mishra et al., 2007; Balz et al., 2016). A reasonable assumption, based on this evidence, is that perceptual binding of stimuli in the SIFI relies largely on stimulus-driven bottom-up processing (Bizley et al., 2016b; Hirst et al., 2020). Yet, this assumption is challenged by our finding of high cognitive load altering the perception of SIFI as well as the oscillatory power modulations associated with it. This finding provides strong evidence that the SIFI is not the result of a hardwired automatic process but that it is subject to constraints on the available cognitive resources. Moreover, it corroborates the idea that multisensory perception is a flexible and adaptive process (van Atteveldt et al., 2014).

Another interesting aspect of the memory-load effect on the SIFI observed in the present thesis, relates to the juxtaposition of the present findings with related findings in the McGurk effect. Alsius and colleagues (2005, 2007) found that when attentional resources are diverted to a secondary visual, auditory, or tactile detection task, the perception of the McGurk effect is reduced. Conversely, we found that depleting cognitive resources through a working memory task results in the enhanced perception of the SIFI. This discrepancy can be attributed to the nature of the two audiovisual illusions. Natural speech relies predominantly on sound. Therefore, the depletion of resources in the McGurk effect might have enhanced the weight of auditory information and hence, weakened the susceptibility to the illusion. On the other hand, temporal discrimination – on which the SIFI is based – is more accurate in the auditory compared to the visual modality. Therefore, the enhanced susceptibility to SIFI under load might be related to increased reliance on auditory information under conditions of limited cognitive resources. An alternative explanation of this discrepancy relates to differences in the nature of the secondary task employed in these two sets of studies. In the studies by Alsius and colleagues, the secondary tasks induce high perceptual load whereas in our experimental design, the orthogonal *n*-back task induces increased cognitive load. According to the “attentional load theory” (Lavie, 2005), perceptual and cognitive load have an opposite effect on the processing of task-irrelevant information. The theory postulates that perceptual load suppresses task-irrelevant information processing, whereas the cognitive load enhances it. In this view, the enhanced susceptibility to the SIFI under high cognitive load can be interpreted as being caused by the enhanced processing of task-irrelevant auditory information. Antithetically, the reduced perception of the McGurk effect under high perceptual load might result from the suppression of the processing of task-irrelevant visual information. An interesting prediction, following from this logic, is that the combination of the McGurk effect with a secondary task inducing high cognitive instead of perceptual load, should result in the opposite effect, that is an enhancement of the likelihood to perceive the illusion.

Overall, the altered susceptibility to the SIFI by high cognitive load observed in the present thesis argues that, even seemingly stimulus-driven multisensory phenomena, such as the SIFI, depend on the availability of cognitive resources. Moreover, the divergence of this finding with respective findings on the McGurk effect highlights the adaptability of integrative mechanisms, and the importance of taking the specific task and stimulus characteristics into consideration when planning experimental designs and when interpreting empirical research on multisensory perception. Future investigations could address the potential influence of cognitive load on the McGurk effect, as discussed above, but also on other multisensory illusions relying on other forms of crossmodal asymmetries, such as the ventriloquist effect (audiovisual spatial incongruence) or the rubber hand illusion (visuotactile incongruence).

3.2 The diverse functional role of beta oscillations in multisensory perception

Our understanding on the functional role of beta oscillations in cognition and perception has been significantly broadened in the last years. Classically, beta oscillations have been linked to sensorimotor functions (Pfurtscheller, 1981; Neuper and Pfurtscheller, 2001). However, recent studies associate beta oscillations with a much more diverse functional repertoire, including the maintenance of endogenous information during working memory and decision making (Spitzer and Haegens, 2017), memory encoding and retrieval (Hanslmayr et al., 2016), the coding of prior expectations (Betti et al., 2021) and the transfer of top-down predictions in the feedback stream of information (Arnal and Giraud, 2012; Bressler and Richter, 2015; Fries, 2015).

With respect to multisensory perception, the role of beta oscillations is contingent on the characteristics of the task and the employed stimuli. As demonstrated in multisensory illusion studies, reviewed in the Introduction (see section: *The role of neural oscillations in multisensory processing*), while gamma oscillations are considered a hallmark of crossmodal binding, multisensory integration in more complex tasks such as speech and body ownership, engages low frequency activity, including beta oscillations. Accordingly, multisensory integration in the SIFI, which employs simple audiovisual stimuli, has been associated with gamma band oscillations (Bhattacharya et al., 2002; Mishra et al., 2007; Balz et al., 2016). Our finding that the SIFI under high cognitive load is related to low frequency activity indicates that integrative processing in the SIFI is adaptive and depends on the availability of cognitive resources. Two different effects involved modulations of beta power, one at early (~70 ms poststimulus) and one at late stages of processing (~350 ms). We propose that the early beta power suppression in auditory and motor cortex reflects the evaluation of crossmodal conflict and the detection of mismatch. This notion is consistent with evidence implicating motor beta oscillations in sensory conflict processing (Huang et al., 2014) and the extensive auditory-motor connections that could enable the observed effect (Zatorre et al., 2007; Rauschecker and Scott, 2009; Nelson et al., 2013; Cheung et al., 2016; Zhang et al., 2016). On the other hand, the late beta power decrease, localized in prefrontal and temporal cortex – including STG – is likely related to top-down modulation of late integrative processing in temporal association areas. This idea is in line with the proposed role of beta oscillations in mediating top-down transfer of information (Arnal and Giraud, 2012; Fries, 2015), as well as with the well-established function of STG as a multisensory convergence area (Calvert et al., 2000; Beauchamp et al., 2004). The observed pattern of early and late beta power modulations indicates that when cognitive resources are limited, multisensory integration in the SIFI relies on spatiotemporally distinctive beta activities subserving conflict detection and late top-down integrative processing.

Regarding the crossmodal facilitation of response speed, while a previous study suggested a link with beta power, the evidence is tentative and most likely reflects motor processing, non-specific to multisensory stimulation (Senkowski et al., 2006). In the third study, we demonstrate that auditory-driven individual gains in visual response speed correlate with reduced beta power in multisensory association and secondary visual areas. Importantly the beta effect occurred within 80 ms after stimulus onset, arguing for a modulation of visual processing at early sensory stages. We propose that this early beta effect, reflects an attentional gain modulation of visual processing originating in multisensory association and extrastriate visual cortex. This notion is in agreement with empirical and modeling work demonstrating the role of beta oscillations in gating sensory processing through selective attention (Fries et al., 2001; Buschman and Miller, 2007; Lee et al., 2013; Alzueta et al., 2020; Limanowski et al., 2020). This finding indicates that crossmodal facilitation of response speed might be linked to beta power enhancing the efficiency of early sensory processing through the attention-dependent routing of information (Figure 6b).

Overall, the present findings suggest a prominent role for beta oscillations in the dynamic and adaptive processing underlying multisensory perception. This role includes the facilitation of multisensory integration of simple stimuli when cognitive resources are scarce, and the multisensory enhancement of response speed. Critically, the demonstration of functionally relevant beta power modulations at early processing stages during these multisensory phenomena argues for an important role of feedback processing on early crossmodal interactions (Driver and Noesselt, 2008).

3.3 An integrative framework for speech and non-speech audiovisual illusions

In the present thesis (Study II), we provide evidence that multisensory integration in the SIFI under high working memory load is associated with oscillatory power modulations including: (i) an early beta power suppression in auditory and motor cortex followed by (ii) a theta power increase in mid-anterior cingulate cortex (mid-ACC), and (iii) a late beta power decrease, in prefrontal and auditory cortex, including the STG (Figure 3 and 5). This pattern of responses is remarkably similar with oscillatory power modulations previously associated with the integration of audiovisual speech stimuli in the McGurk effect. For instance, an EEG study showed that audiovisual speech integration during the McGurk effect is linked to an early beta suppression in left fronto-central electrodes at 50-100 ms poststimulus and a late beta suppression in frontal electrodes at 500-600 ms (Roa Romero et al., 2015). Another study by the same group showed that in healthy individuals the McGurk effect is also linked to a mid-frontal theta power increase occurring at 230-370 ms (Roa Romero et al., 2016). Subsequent electrophysiological and neuroimaging evidence from another group suggests that the mid-frontal theta power during the McGurk effects might reflect conflict processing (Morís Fernández et al., 2017, 2018). The striking similarity in terms of the spectral and spatiotemporal characteristics of the oscillatory responses associated with the

McGurk effect and the SIFI under high cognitive load points to a substantial overlap in the mechanisms underlying these two phenomena.

Prompted by the significant convergence in the oscillatory signatures of these two phenomena, I propose a framework that attempts to integrate this convergent evidence and provides a mechanistic model for the crossmodal processing of audiovisual incongruent signals (Figure 6a). A critical assumption of this framework is that the processing of audiovisual incongruence engages low frequency oscillations reflecting top-down control, only when the integration of conflicting audiovisual stimuli requires more processing resources either due to the nature of the crossmodal asymmetry (e.g., semantic asymmetry in the McGurk effect) or because of a scarcity in the available resources (e.g., SIFI under high cognitive load). Therefore, this model does not account for integrative processing in the SIFI when no additional task is performed, as in this case crossmodal binding is considered a bottom-up process relying on gamma oscillations (Bhattacharya et al., 2002; Mishra et al., 2007; Balz et al., 2016).

The proposed framework suggests that multisensory integration of incongruent stimuli consists of three stages. In the early stage, the crossmodal incongruence is evaluated and detected and this step is subserved by beta power in an auditory-motor network. This proposal is based on the common left frontocentral topography of the early beta suppression in both phenomena (for McGurk see Roa Romero et al., 2016/Figure 4B), which in the case of the SIFI under high load, was localized in left auditory and motor areas. This notion is consistent with the reciprocal functional and anatomical connections between auditory and motor areas (Zatorre et al., 2007; Rauschecker and Scott, 2009; Nelson et al., 2013; Cheung et al., 2016; Zhang et al., 2016) and previous evidence implicating beta oscillations in sensory conflict and prediction error processing (Arnal et al., 2011; Arnal and Giraud, 2012; Huang et al., 2014). Following the beta suppression, increased theta power in mid-anterior cingulate cortex signals the need for cognitive control in the face of enhanced crossmodal conflict and perceptual uncertainty. This notion is consistent with the well-established role of mid-frontal theta oscillations as a signal for cognitive control (Cavanagh and Frank, 2014) as well as with evidence implicating theta oscillations in top-down guided attention (Helfrich et al., 2019). This proposal is also in accordance with studies demonstrating the involvement of theta power in the processing of intersensory conflict for both speech (Lange et al., 2013) and non-speech stimuli (Cohen and Donner, 2013). Following the theta power signal for the need of cognitive control, a beta power suppression reflects the top-down frontal modulation of late crossmodal processing in association areas in temporal cortex. This notion accords with evidence and proposals relating beta oscillations with top-down influences from higher order to lower order sensory areas (Buschman and Miller, 2007; Bastos et al., 2015; Fries, 2015; Richter et al., 2017). The assumed spatial profile of this late beta effect is based on its localization in the prefrontal and bilateral temporal areas, including STG, extending the frontal topography of the effect at sensor space identified in both the current thesis and the study by Roa Romero et al. (2016; Figure 5B). The involvement of beta power in multisensory association areas at late processing stages is further supported by a previous EEG study demonstrating a late beta suppression in inferior parietal cortex during the processing of incongruent speech stimuli (Lange et al., 2013).

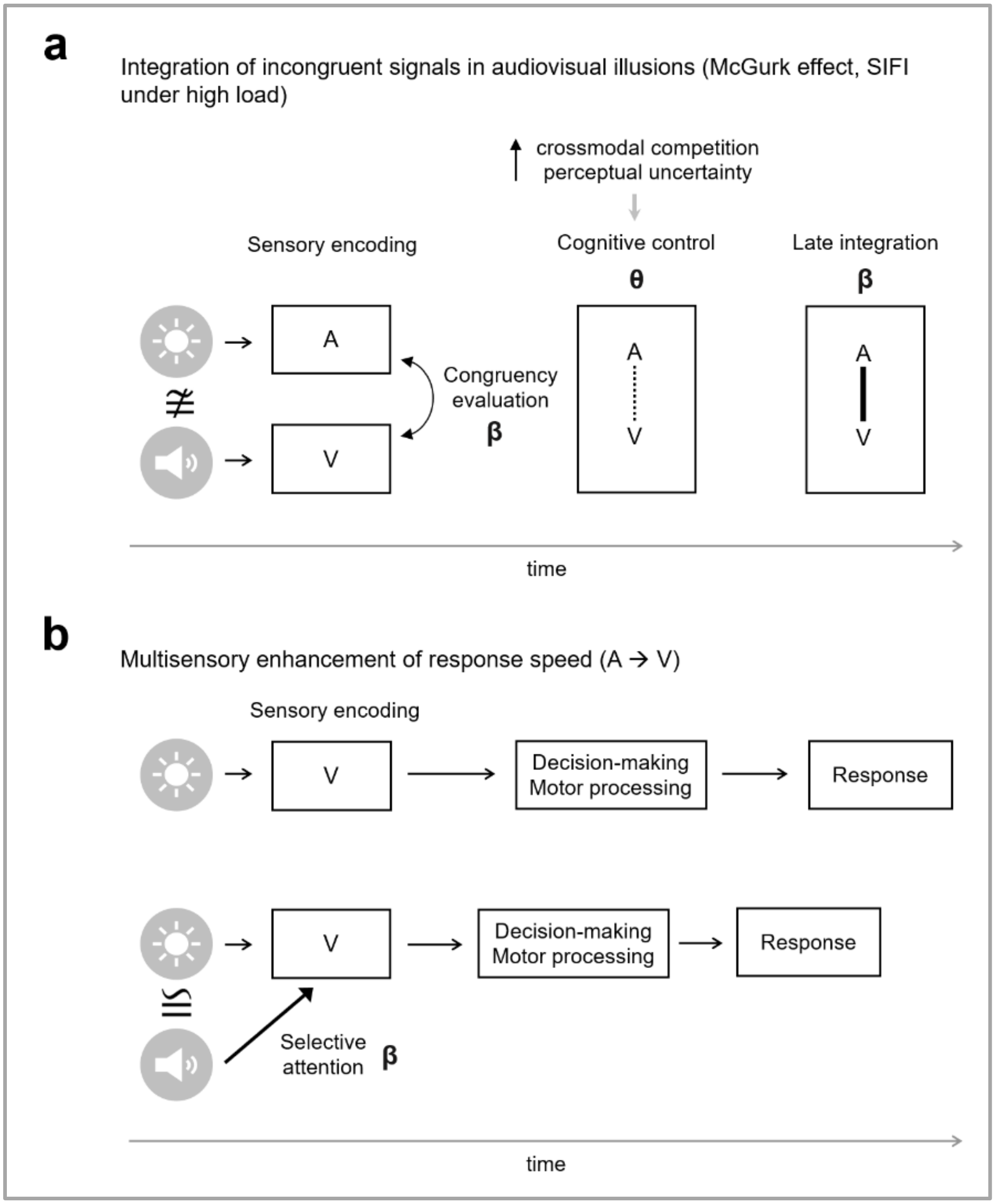

Figure 6. Conceptual frameworks incorporating the findings of the current thesis. **a.** Framework for the integration of audiovisual stimuli in the McGurk effect and the SIFI under high cognitive load. The framework is based on the pattern of oscillatory responses associated with SIFI under high load in the current thesis, which is similar to the oscillatory power modulations previously linked to the McGurk effect. **b.** A mechanistic explanation for the finding of the third study,

associating crossmodal RT facilitation with early beta power. Based on the latency and topography of the observed effect, we propose that auditory-driven early beta power reflects a sharpening of early visual processing through selective attention. This, in turn, accelerates all the subsequent processing steps and finally, the response times. A and V indicate auditory and visual processing. Greek letters β and θ are used to indicate beta and theta oscillations.

Overall, this framework provides a common mechanistic basis for the integration of audiovisual stimuli in the McGurk effect and the SIFI under limited cognitive resources. Importantly, the spatiotemporally distinctive contributions of low frequency oscillations, associated with both of these phenomena, might reflect general integrative mechanisms involved in various multisensory settings. Further investigations should address the generalization of this framework by testing whether low frequency oscillations play a similar role in other multisensory tasks, and especially in illusory phenomena produced by crossmodal asymmetries.

3.4 Methodological limitations

One restriction of the present thesis, in the context of the EEG studies (Study II and III), is that the role of neural oscillations in multisensory perception was assessed only in terms of power modulations. As reviewed in the introduction, there is evidence that multisensory perception in the studied paradigms relies on the communication between different areas (SIFI: Keil et al., 2014; crossmodal RT facilitation: Mercier et al., 2015). Therefore, assessing crossmodal phase-resetting or large-scale cortical synchronization could provide complementary evidence to the observed power modulations that would allow the deduction of a more comprehensive understanding of the underlying neural mechanisms, especially with regard to interregional communication. Such approaches could for instance elucidate the network of areas involved in the top-down control influences observed during the SIFI under high cognitive load as well as the direction of these modulations (Study II).

Another limitation of the thesis relates to the use of ECoG recordings in the third study. Despite the important advantages of ECoG recordings such as the high spatial resolution and the high signal-to-noise ratio, there are some apparent limitations (Parvizi and Kastner, 2018). First, given the rarity of such operations and the potential implications precluding participation, it was difficult to obtain a large sample size. Despite our analysis yielding highly consistent results across all individuals, the small sample size prevented us from performing analyses at the group level which would be comparable with the analyses of the EEG data. Moreover, the positioning of the implanted electrodes is purely dictated by the clinical needs of each patient, often resulting in heterogenous cortical coverage. In our study, electrode grids covered mainly the temporal cortex, however with a large interindividual variability. In addition, in three individuals the electrode grid covered the left temporal cortex and only in one individual the right temporal cortex. Hence, we could not compare differences between hemispheres and test whether the ECoG data corroborated the right hemisphere localization of the correlation between beta power and RTs, observed in the analysis of the EEG data. Therefore, future investigations should recruit larger participant cohorts and with largely homogenous cortical coverage. In parallel, our understanding on the role of different regions in multisensory perception could benefit from ECoG investigations in other cortical areas that are critical for crossmodal processing (e.g., primary sensory areas and association areas in parietal and frontal cortex).

3.5 Outlook

The findings of the current thesis regarding the role of power modulations in the SIFI under high cognitive load open up several directions for future investigations. As mentioned in the previous section, future studies may employ measures of phase synchrony to assess the role of large-scale cortical synchronization in the SIFI under conditions of increased cognitive demands. Another interesting direction of future studies would be to employ non-invasive brain stimulation techniques such as the transcranial magnetic stimulation (TMS) or transcranial alternating current stimulation (tACS) to selectively modulate activity in cortical areas and frequency bands of interest (in the case of tACS) and obtain causal evidence regarding their role in crossmodal processing. For instance, by modulating activity in the left motor cortex while participants perform the SIFI under high load, we could assess how crucial is the involvement of motor cortex in early crossmodal processing that was observed in the current thesis. These methods have been successfully used in previous studies to demonstrate causal links regarding the role of angular gyrus (Kamke et al., 2012) and the individual alpha frequency in SIFI (Cecere et al., 2015).

Future studies may also attempt to elucidate the role of prestimulus oscillatory activity in the studied multisensory paradigms. Growing evidence suggests that ongoing brain activity, reflected in the spontaneous oscillatory activity, influences the perception of the upcoming stimulus. In particular, the phase and power of alpha oscillations, thought to reflect the current state of cortical excitability, have been consistently implicated in the modulation of visual perception (Busch et al., 2009; Mathewson et al., 2009; Iemi and Busch, 2018; Samaha et al., 2020; Michail et al., 2021b). With regard to the multisensory performance enhancement, future studies may analyze prestimulus alpha activity in visual cortex to investigate whether the power or the phase of alpha oscillations affect the magnitude of the facilitation of the visual responses driven by concurrent information in another modality. Regarding the multisensory illusion perception, previous studies have demonstrated that prestimulus alpha oscillations can influence the upcoming perception in several multisensory illusions (Lange et al., 2014; Keil and Senkowski, 2018). Moreover, previous studies have linked the SIFI to various prestimulus oscillatory activities, including the individual alpha frequency in visual cortex (Cecere et al., 2015; Keil and Senkowski, 2017), visual gamma power (Kaiser et al., 2019), as well as beta power in temporal cortex and alpha and beta synchrony between auditory and multisensory association areas (Keil et al., 2014). Future investigations could assess whether the influence that memory load exerts on SIFI is related to prestimulus power or phase in the low or high frequency oscillations.

More generally, future investigations could examine to what extent the current findings generalize to other paradigms. For instance, is the early beta power as relevant for the crossmodal enhancement of stimulus detection as it is for the facilitation of response speed? Or is the effect of working memory load on the McGurk effect or the rubber hand illusion analogous to its influence on the SIFI? These investigations could shed light on commonalities as well as differences between crossmodal processing in different settings and enable an understanding of the factors underlying the large heterogeneity of results among studies on multisensory perception.

3.6 Conclusion

The present thesis examined the behavioral and oscillatory responses associated with the SIFI performed under various cognitive demands, and the crossmodal facilitation of response speed. The significance of the thesis's findings can be summarized in three main contributions. First, I showed that integration of simple non-speech audiovisual stimuli is sensitive to the available cognitive resources and therefore is not as stimulus-driven as previously assumed. Second, the depletion of cognitive resources in the SIFI resulted in the engagement of low frequency oscillations following a multistage spatiotemporal pattern, remarkably similar with the oscillatory signatures of audiovisual speech integration in the McGurk effect. This finding constitutes a characteristic illustration of the adaptability of multisensory integration, in that the SIFI, while usually linked to bottom-up gamma oscillations, relies on top-down low frequency oscillations when resources are scarce. At the same time, this finding demonstrates the flexibility of integrative mechanisms as revealed by the shared oscillatory signatures underlying the McGurk effect and SIFI under load. The last contribution of the thesis regards the finding of early beta power being associated with the crossmodal facilitation of response speed. By correlating oscillatory responses with individual performance gains, we demonstrate that beta power starting within 80 ms after stimulus onset reflects the enhancement of early sensory processing which in turn accelerates response times. This finding covers a critical gap in the literature regarding the oscillatory correlates of multisensory enhancement and suggests a functional role of beta oscillations in early sensory stages of crossmodal processing. Regarding the debate on the hierarchical model of multisensory perception (Driver and Noesselt, 2008; Bizley et al., 2016a), the present findings highlight the major role of feedback processing at early stages, reflected in the beta band power.

Bibliography

Alais D, Burr D (2004) The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. Current Biology 14:257–262.

Alais D, Newell FN, Mamassian P (2010) Multisensory processing in review: from physiology to behaviour. Seeing Perceiving 23:3–38.

Alsius A, Navarra J, Campbell R, Soto-Faraco S (2005) Audiovisual Integration of Speech Falters under High Attention Demands. Current Biology 15:839–843.

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

Alsius A, Paré M, Munhall KG (2018) Forty Years After Hearing Lips and Seeing Voices: the McGurk Effect Revisited. Multisensory Research 31:111–144.

Alzueta E, Melcón M, Jensen O, Capilla A (2020) The ‘Narcissus Effect’: Top-down alpha-beta band modulation of face-related brain areas during self-face processing. NeuroImage 213:116754.

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

Arnal LH, Wyart V, Giraud AL (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat Neurosci 14:797–801.

Balz J, Keil J, Roa Romero Y, Mekle R, Schubert F, Aydin S, Ittermann B, Gallinat J, Senkowski D (2016) GABA concentration in superior temporal sulcus predicts gamma power and perception in the sound-induced flash illusion. NeuroImage 125:724–730.

Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of Visual and Auditory Information by Superior Temporal Sulcus Neurons Responsive to the Sight of Actions. Journal of Cognitive Neuroscience 17:377–391.

Bastos AM, Schoffelen JM (2016) A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. Front Syst Neurosci 9:175.

Bastos AM, Vezoli J, Bosman CA, Schoffelen JM, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. Neuron 85:390–401.

Bauer AKR, Debener S, Nobre AC (2020) Synchronisation of Neural Oscillations and Cross-modal Influences. Trends Cogn Sci 24:481–495.

Bauer M, Kennett S, Driver J (2012) Attentional selection of location and modality in vision and touch modulates low-frequency activity in associated sensory cortices. Journal of Neurophysiology 107:2342–2351.

Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of Auditory and Visual Information about Objects in Superior Temporal Sulcus. Neuron 41:809–823.

Bertelson P, Vroomen J, De Gelder B, Driver J (2000) The ventriloquist effect does not depend on the direction of deliberate visual attention. Perception & Psychophysics 62:321–332.

Betti V, Della Penna S, de Pasquale F, Corbetta M (2021) Spontaneous Beta Band Rhythms in the Predictive Coding of Natural Stimuli. Neuroscientist 27:184–201.

Bhattacharya J, Shams L, Shimojo S (2002) Sound-induced illusory flash perception: role of gamma band responses. NeuroReport 13:1727–1730.

Bizley JK, Jones GP, Town SM (2016a) Where are multisensory signals combined for perceptual decision-making? Current Opinion in Neurobiology 40:31–37.

Bizley JK, Maddox RK, Lee AKC (2016b) Defining Auditory-Visual Objects: Behavioral Tests and Physiological Mechanisms. Trends in Neurosciences 39:74–85.

Blake A, Bülthoff HH, Sheinberg D (1993) Shape from texture: Ideal observers and human psychophysics. Vision Research 33:1723–1737.

Botvinick M, Cohen J (1998) Rubber hands ‘feel’ touch that eyes see. Nature 391:756–756.

Bremmer F, Schlack A, Shah NJ, Zafiris O, Kubischik M, Hoffmann K-P, Zilles K, Fink GR (2001) Polymodal Motion Processing in Posterior Parietal and Premotor Cortex: A Human fMRI Study Strongly Implies Equivalencies between Humans and Monkeys. Neuron 29:287–296.

Bressler SL, Richter CG (2015) Interareal oscillatory synchronization in top-down neocortical processing. Current Opinion in Neurobiology 31:62–66.

Budinger E, Heil P, Hess A, Scheich H (2006) Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. Neuroscience 143:1065–1083.

Busch NA, Dubois J, VanRullen R (2009) The Phase of Ongoing EEG Oscillations Predicts Visual Perception. J Neurosci 29:7869–7876.

Buschman TJ, Miller EK (2007) Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. Science 315:1860–1862.

Buzsáki G (2006) Rhythms of the brain. New York, NY, US: Oxford University Press.

Buzsáki G, Wang XJ (2012) Mechanisms of Gamma Oscillations. Annu Rev Neurosci 35:203–225.

Buzsáki G, Watson BO (2012) Brain rhythms and neural syntax: implications for efficient coding of cognitive content and neuropsychiatric disease. Dialogues Clin Neurosci 14:345–367.

Calvert GA (2001) Crossmodal Processing in the Human Brain: Insights from Functional Neuroimaging Studies. Cerebral Cortex 11:1110–1123.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS (1997) Activation of Auditory Cortex During Silent Lipreading. Science 276:593–596.

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Current Biology 10:649–657.

Calvert GA, Thesen T (2004) Multisensory integration: methodological approaches and emerging principles in the human brain. Journal of Physiology-Paris 98:191–205.

Canolty RT, Knight RT (2010) The functional role of cross-frequency coupling. Trends in Cognitive Sciences 14:506–515.

Canon LK (1971) Directed attention and maladaptive “adaptation” to displacement of the visual field. Journal of Experimental Psychology 88:403–408.

Cao Y, Summerfield C, Park H, Giordano BL, Kayser C (2019) Causal Inference in the Multisensory Brain. Neuron 102:1076-1087.e8.

Cappe C, Rouiller EM, Barone P (2009) Multisensory anatomical pathways. Hearing Research 258:28–36.

Cappe C, Thut G, Romei V, Murray MM (2010) Auditory-Visual Multisensory Interactions in Humans: Timing, Topography, Directionality, and Sources. Journal of Neuroscience 30:12572–12580.

Cavanagh JF, Frank MJ (2014) Frontal theta as a mechanism for cognitive control. Trends in Cognitive Sciences 18:414–421.

Cecere R, Rees G, Romei V (2015) Individual Differences in Alpha Frequency Drive Crossmodal Illusory Perception. Current Biology 25:231–235.

Cheung C, Hamilton LS, Johnson K, Chang EF (2016) The auditory representation of speech sounds in human motor cortex. eLife 5:e12577.

Choe CS, Welch RB, Gilford RM, Juola JF (1975) The “ventriloquist effect”: Visual dominance or response bias? Perception & Psychophysics 18:55–60.

Cohen MX, Donner TH (2013) Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. Journal of Neurophysiology 110:2752–2763.

Colonius H, Diederich A (2020) Formal models and quantitative measures of multisensory integration: a selective overview. European Journal of Neuroscience 51:1161–1178.

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.

Donohue SE, Green JJ, Woldorff MG (2015) The effects of attention on the temporal integration of multisensory stimuli. Front Integr Neurosci 9:32.

Driver J, Noesselt T (2008) Multisensory Interplay Reveals Crossmodal Influences on ‘Sensory-Specific’ Brain Regions, Neural Responses, and Judgments. Neuron 57:11–23.

Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A (2012) The Cost of Accumulating Evidence in Perceptual Decision Making. J Neurosci 32:3612–3628.

Eckert MA, Kamdar NV, Chang CE, Beckmann CF, Greicius MD, Menon V (2008) A cross-modal system linking primary auditory and visual cortices: evidence from intrinsic fMRI connectivity analysis. Hum Brain Mapp 29:848–857.

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433.

Evans N, Blanke O (2013) Shared electrophysiology mechanisms of body ownership and motor imagery. NeuroImage 64:216–228.

Faivre N, Dönz J, Scandola M, Dhanis H, Ruiz JB, Bernasconi F, Salomon R, Blanke O (2017) Self-Grounded Vision: Hand Ownership Modulates Visual Location through Cortical β and γ Oscillations. J Neurosci 37:11–22.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47.

Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE (2012) Neural correlates of reliability-based cue weighting during multisensory integration. Nature Neuroscience 15:146–154.

Fort A, Delpuech C, Pernier J, Giard MH (2002) Dynamics of Cortico-subcortical Cross-modal Operations Involved in Audio-visual Object Detection in Humans. Cerebral Cortex 12:1031–1039.

Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. NeuroReport 16:419–423.

Franzen L, Delis I, De Sousa G, Kayser C, Philiastides MG (2020) Auditory information enhances post-sensory visual evidence during rapid multisensory decision-making. Nature Communications 11:5440.

Fries P (2015) Rhythms for Cognition: Communication through Coherence. Neuron 88:220–235.

Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention. Science 291:1560–1563.

Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory Integration of Dynamic Faces and Voices in Rhesus Monkey Auditory Cortex. J Neurosci 25:5004–5012.

Giard MH, Peronnet F (1999) Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. Journal of Cognitive Neuroscience 11:473–490.

Gleiss S, Kayser C (2014a) Oscillatory mechanisms underlying the enhancement of visual motion perception by multisensory congruency. Neuropsychologia 53:84–93.

Gleiss S, Kayser C (2014b) Acoustic Noise Improves Visual Perception and Modulates Occipital Oscillatory States. Journal of Cognitive Neuroscience 26:699–711.

Göschl F, Friese U, Daume J, König P, Engel AK (2015) Oscillatory signatures of crossmodal congruence effects: An EEG investigation employing a visuotactile pattern matching paradigm. NeuroImage 116:177–186.

Gregoric P (2007) Aristotle on the Common Sense. Oxford University Press.

Hackett TA, Mothe LADL, Ulbert I, Karmos G, Smiley J, Schroeder CE (2007) Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. Journal of Comparative Neurology 502:924–952.

Hanslmayr S, Staresina BP, Bowman H (2016) Oscillations and Episodic Memory: Addressing the Synchronization/Desynchronization Conundrum. Trends in Neurosciences 39:16–25.

Helfrich RF, Breska A, Knight RT (2019) Neural entrainment and network resonance in support of top-down guided attention. Current Opinion in Psychology 29:82–89.

Heller MA (1982) Visual and tactual texture perception: Intersensory cooperation. Perception & Psychophysics 31:339–344.

Henschke JU, Noesselt T, Scheich H, Budinger E (2015) Possible anatomical pathways for short-latency multisensory integration processes in primary sensory cortices. Brain Structure and Function 220:955–977.

Herrmann CS, Strüber D, Helfrich RF, Engel AK (2016) EEG oscillations: From correlation to causality. International Journal of Psychophysiology 103:12–21.

Hershenson M (1962) Reaction time as a measure of intersensory facilitation. Journal of Experimental Psychology 63:289–293.

Hipp JF, Engel AK, Siegel M (2011) Oscillatory Synchronization in Large-Scale Cortical Networks Predicts Perception. Neuron 69:387–396.

Hirst RJ, McGovern DP, Setti A, Shams L, Newell FN (2020) What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. Neuroscience & Biobehavioral Reviews 118:759–774.

Huang S, Rossi S, Hämäläinen M, Ahveninen J (2014) Auditory Conflict Resolution Correlates with Medial–Lateral Frontal Theta/Alpha Phase Synchrony. PLOS ONE 9:e110989.

Hyvärinen J (1981) Regional distribution of functions in parietal association area 7 of the monkey. Brain Research 206:287–303.

Iemi L, Busch NA (2018) Moment-to-Moment Fluctuations in Neuronal Excitability Bias Subjective Perception Rather than Strategic Decision-Making. eNeuro 5.

Jensen O, Colgin LL (2007) Cross-frequency coupling between neuronal oscillations. Trends in Cognitive Sciences 11:267–269.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci 4:186.

Kaiser J, Hertrich I, Ackermann H, Mathiak K, Lutzenberger W (2005) Hearing Lips: Gamma-band Activity During Audiovisual Speech Perception. Cereb Cortex 15:646–653.

Kaiser M, Senkowski D, Busch NA, Balz J, Keil J (2019) Single trial prestimulus oscillations predict perception of the sound-induced flash illusion. Sci Rep 9:1–8.

Kaiser M, Senkowski D, Keil J (2021) Mediofrontal theta-band oscillations reflect top-down influence in the ventriloquist illusion. Human Brain Mapping 42:452–466.

Kamke MR, Vieth HE, Cottrell D, Mattingley JB (2012) Parietal disruption alters audiovisual binding in the sound-induced flash illusion. NeuroImage 62:1334–1341.

Kanayama N, Sato A, Ohira H (2007) Crossmodal effect with rubber hand illusion and gamma-band activity. Psychophysiology 44:392–402.

Kanayama N, Sato A, Ohira H (2009) The role of gamma band oscillations and synchrony on rubber hand illusion and crossmodal integration. Brain and Cognition 69:19–29.

Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal information? Brain Struct Funct 212:121–132.

Kayser C, Logothetis NK (2009) Directed Interactions Between Auditory and Superior Temporal Cortices and their Role in Sensory Integration. Front Integr Neurosci 3:7.

Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional Imaging Reveals Visual Modulation of Specific Fields in Auditory Cortex. J Neurosci 27:1824–1835.

Kayser C, Petkov CI, Logothetis NK (2008) Visual Modulation of Neurons in Auditory Cortex. Cerebral Cortex 18:1560–1574.

Kayser SJ, Philiastides MG, Kayser C (2017) Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. NeuroImage 148:31–41.

Keil J (2020) Double flash illusions: current findings and future directions. Front Neurosci 14:298.

Keil J, Müller N, Hartmann T, Weisz N (2014) Prestimulus Beta Power and Phase Synchrony Influence the Sound-Induced Flash Illusion. Cereb Cortex 24:1278–1288.

Keil J, Müller N, Ihssen N, Weisz N (2012) On the Variability of the McGurk Effect: Audiovisual Integration Depends on Prestimulus Brain States. Cereb Cortex 22:221–231.

Keil J, Pomper U, Senkowski D (2016) Distinct patterns of local oscillatory activity and functional connectivity underlie intersensory attention and temporal prediction. Cortex 74:277–288.

Keil J, Senkowski D (2017) Individual Alpha Frequency Relates to the Sound-Induced Flash Illusion. Multisensory Research 30:565–578.

Keil J, Senkowski D (2018) Neural Oscillations Orchestrate Multisensory Processing. The Neuroscientist 24:609–626.

Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal Inference in Multisensory Perception. PLOS ONE 2:e943.

Kumar GV, Halder T, Jaiswal AK, Mukherjee A, Roy D, Banerjee A (2016) Large Scale Functional Brain Networks Underlying Temporal Integration of Audio-Visual Speech Perception: An EEG Study. Front Psychol 7:1558.

Lakatos P, Chen C-M, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. Neuron 53:279–292.

Lange J, Christian N, Schnitzler A (2013) Audio–visual congruency alters power and coherence of oscillatory activity within and between cortical areas. NeuroImage 79:111–120.

Lange J, Keil J, Schnitzler A, van Dijk H, Weisz N (2014) The role of alpha oscillations for illusory perception. Behavioural Brain Research 271:294–301.

Lange J, Oostenveld R, Fries P (2011) Perception of the touch-induced visual double-flash illusion correlates with changes of rhythmic neuronal activity in human visual and somatosensory areas. NeuroImage 54:1395–1405.

Laurienti PJ, Burdette JH, Wallace MT, Yen Y-F, Field AS, Stein BE (2002) Deactivation of Sensory-Specific Cortex by Cross-Modal Stimuli. Journal of Cognitive Neuroscience 14:420–429.

Lavie N (2005) Distracted and confused?: Selective attention under load. Trends in Cognitive Sciences 9:75–82.

Lee JH, Whittington MA, Kopell NJ (2013) Top-Down Beta Rhythms Support Selective Attention via Interlaminar Interaction: A Model. PLOS Computational Biology 9:e1003164.

Limanowski J, Litvak V, Friston K (2020) Cortical beta oscillations reflect the contextual gating of visual action feedback. NeuroImage 222:117267.

Ma WJ, Zhou X, Ross LA, Foxe JJ, Parra LC (2009) Lip-Reading Aids Word Recognition Most in Moderate Noise: A Bayesian Explanation Using High-Dimensional Feature Space. PLOS ONE 4:e4638.

Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. Trends in Neurosciences 28:264–271.

Macaluso E, Frith CD, Driver J (2000) Modulation of Human Visual Cortex by Crossmodal Spatial Attention. Science 289:1206–1208.

Macaluso E, Frith CD, Driver J (2002) Crossmodal Spatial Influences of Touch on Extrastriate Visual Areas Take Current Gaze Direction into Account. Neuron 34:647–658.

Macaluso E, Noppeney U, Talsma D, Vercillo T, Hartcher-O’Brien J, Adam R (2016) The Curious Incident of Attention in Multisensory Integration: Bottom-up *vs.* Top-down. Multisensory Research 29:557–583.

Majka P, Rosa MGP, Bai S, Chan JM, Huo B-X, Jermakow N, Lin MK, Takahashi YS, Wolkowicz IH, Worthy KH, Rajan R, Reser DH, Wójcik DK, Okano H, Mitra PP (2019) Unidirectional monosynaptic connections from auditory areas to the primary visual cortex in the marmoset monkey. Brain Struct Funct 224:111–131.

Marks LE (1987) On cross-modal similarity: Auditory–visual interactions in speeded discrimination. Journal of Experimental Psychology: Human Perception and Performance 13:384–394.

Mathewson KE, Gratton G, Fabiani M, Beck DM, Ro T (2009) To See or Not to See: Prestimulus α Phase Predicts Visual Awareness. J Neurosci 29:2725–2732.

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

Mégevand P, Mercier MR, Groppe DM, Golumbic EZ, Mesgarani N, Beauchamp MS, Schroeder CE, Mehta AD (2020) Crossmodal Phase Reset and Evoked Responses Provide Complementary Mechanisms for the Influence of Visual Speech in Auditory Cortex. J Neurosci 40:8530–8542.

Mercier MR, Foxe JJ, Fiebelkorn IC, Butler JS, Schwartz TH, Molholm S (2013) Auditory-driven phase reset in visual cortex: Human electrocorticography reveals mechanisms of early multisensory integration. Neuroimage 79:19–29.

Mercier MR, Molholm S, Fiebelkorn IC, Butler JS, Schwartz TH, Foxe JJ (2015) Neuro-Oscillatory Phase Alignment Drives Speeded Multisensory Response Times: An Electro-Corticographic Investigation. The Journal of Neuroscience 35:8546–8557.

Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. Science 221:389–391.

Mesulam MM (1998) From sensation to cognition. Brain 121:1013–1052.

Michail G, Senkowski D, Niedeggen M, Keil J (2021a) Memory Load Alters Perception-Related Neural Oscillations during Multisensory Integration. J Neurosci 41:1505–1515.

Michail G, Toran Jenner L, Keil J (2021b) Prestimulus alpha power but not phase influences visual discrimination of long-duration visual stimuli. Eur J Neurosci.

Miller J (1982) Divided attention: Evidence for coactivation with redundant signals. Cognitive Psychology 14:247–279.

Mishra J, Martinez A, Sejnowski TJ, Hillyard SA (2007) Early Cross-Modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion. J Neurosci 27:4120–4131.

Misselhorn J, Schwab BC, Schneider TR, Engel AK (2019) Synchronization of Sensory Gamma Oscillations Promotes Multisensory Communication. eNeuro 6.

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. Cognitive Brain Research 14:115–128.

Molholm S, Sehatpour P, Mehta AD, Shpaner M, Gomez-Ramirez M, Ortigue S, Dyke JP, Schwartz TH, Foxe JJ (2006) Audio-Visual Multisensory Integration in Superior Parietal Lobule Revealed by Human Intracranial Recordings. Journal of Neurophysiology 96:721–729.

Morís Fernández L, Macaluso E, Soto-Faraco S (2017) Audiovisual integration as conflict resolution: The conflict of the McGurk illusion. Human Brain Mapping 38:5691–5705.

Morís Fernández L, Torralba M, Soto-Faraco S (2018) Theta oscillations reflect conflict processing in the perception of the McGurk illusion. European Journal of Neuroscience 48:2630–2641.

Nagy A, Eördegh G, Paróczy Z, Márkus Z, Benedek G (2006) Multisensory integration in the basal ganglia. European Journal of Neuroscience 24:917–924.

Nelson A, Schneider DM, Takatoh J, Sakurai K, Wang F, Mooney R (2013) A Circuit for Motor Cortical Modulation of Auditory Cortical Activity. Journal of Neuroscience 33:14342–14353.

Neuper C, Pfurtscheller G (2001) Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates. International Journal of Psychophysiology 43:41–58.

Odegaard B, Wozny DR, Shams L (2015) Biases in Visual, Auditory, and Audiovisual Perception of Space. PLOS Computational Biology 11:e1004649.

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

Parvizi J, Kastner S (2018) Promises and limitations of human intracranial electroencephalography. Nat Neurosci 21:474–483.

Pfurtscheller G (1981) Central beta rhythm during sensorimotor activities in man. Electroencephalography and Clinical Neurophysiology 51:253–264.

Pomper U, Keil J, Foxe JJ, Senkowski D (2015) Intersensory selective attention and temporal orienting operate in parallel and are instantiated in spatially distinct sensory and motor cortices. Hum Brain Mapp 36:3246–3259.

Raab DH (1962) Division of Psychology: Statistical Facilitation of Simple Reaction Times*. Transactions of the New York Academy of Sciences 24:574–590.

Rao IS, Kayser C (2017) Neurophysiological Correlates of the Rubber Hand Illusion in Late Evoked and Alpha/Beta Band Activity. Front Hum Neurosci 11:377.

Ratcliff R, Smith PL (2004) A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. Psychol Rev 111:333–367.

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724.

Richter CG, Thompson WH, Bosman CA, Fries P (2017) Top-Down Beta Enhances Bottom-Up Gamma. J Neurosci 37:6698–6711.

Roa Romero Y, Keil J, Balz J, Gallinat J, Senkowski D (2016) Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia. Journal of Neurophysiology 116:1396–1407.

Roa Romero Y, Senkowski D, Keil J (2015) Early and late beta-band power reflect audiovisual perception in the McGurk illusion. Journal of Neurophysiology 113:2342–2350.

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. International Journal of Psychophysiology 50:19–26.

Rohe T, Ehlis A-C, Noppeney U (2019) The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. Nat Commun 10:1–17.

Rohe T, Noppeney U (2015a) Sensory reliability shapes perceptual inference via two mechanisms. Journal of Vision 15:22–22.

Rohe T, Noppeney U (2015b) Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. PLOS Biology 13:e1002073.

Rohe T, Noppeney U (2016) Distinct Computational Principles Govern Multisensory Integration in Primary Sensory and Association Cortices. Current Biology 26:509–514.

Rohe T, Noppeney U (2018) Reliability-Weighted Integration of Audiovisual Signals Can Be Modulated by Top-down Attention. eneuro 5:ENEURO.0315-17.2018.

Samaha J, Iemi L, Haegens S, Busch NA (2020) Spontaneous Brain Oscillations and Perceptual Decision-Making. Trends Cogn Sci 24:639–653.

Schlack A, Sterbing-D'Angelo SJ, Hartung K, Hoffmann K-P, Bremmer F (2005) Multisensory Space Representations in the Macaque Ventral Intraparietal Area. J Neurosci 25:4616–4625.

Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. Nature 385:308.

Senkowski D, Molholm S, Gomez-Ramirez M, Foxe JJ (2006) Oscillatory Beta Activity Predicts Response Speed during a Multisensory Audiovisual Reaction Time Task: A High-Density Electrical Mapping Study. Cereb Cortex 16:1556–1565.

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

Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG (2007) Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia 45:561–571.

Senkowski D, Talsma D, Herrmann CS, Woldorff MG (2005) Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. Exp Brain Res 166:411–426.

Shams L, Iwaki S, Chawla A, Bhattacharya J (2005a) Early modulation of visual cortex by sound: an MEG study. Neuroscience Letters 378:76–81.

Shams L, Kamitani Y, Shimojo S (2000) Illusions: What you see is what you hear. Nature 408:788.

Shams L, Ma WJ, Beierholm U (2005b) Sound-induced flash illusion as an optimal percept. NeuroReport 16:1923–1927.

Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE (2007) Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. Journal of Comparative Neurology 502:894–923.

Soto-Faraco S, Navarra J, Alsius A (2004) Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. Cognition 92:B13–B23.

Spitzer B, Haegens S (2017) Beyond the Status Quo: A Role for Beta Oscillations in Endogenous Content (Re)Activation. eNeuro 4:ENEURO.0170-17.2017.

Stanford TR, Stein BE (2007) Superadditivity in multisensory integration: putting the computation in context. NeuroReport 18:787–792.

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

Stein BE, Stanford TR, Rowland BA (2020) Multisensory Integration and the Society for Neuroscience: Then and Now. J Neurosci 40:3–11.

Stevenson RA, Ghose D, Fister JK, Sarko DK, Altieri NA, Nidiffer AR, Kurela LR, Siemann JK, James TW, Wallace MT (2014) Identifying and Quantifying Multisensory Integration: A Tutorial Review. Brain Topogr 27:707–730.

Stevenson RA, Wallace MT (2013) Multisensory temporal integration: Task and stimulus dependencies. Exp Brain Res 227:249–261.

Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. Trends in Cognitive Sciences 3:151–162.

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 in Cognitive Sciences 14:400–410.

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.

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. Cognitive Brain Research 14:106–114.

ten Oever S, Romei V, van Atteveldt N, Soto-Faraco S, Murray MM, Matusz PJ (2016) The COGs (context, object, and goals) in multisensory processing. Exp Brain Res 234:1307–1323.

Todd JW (1912) Reaction to multiple stimuli. The Science Press.

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

van Atteveldt N, Murray MM, Thut G, Schroeder CE (2014) Multisensory Integration: Flexible Use of General Operations. Neuron 81:1240–1253.

von Stein A, Sarnthein J (2000) Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. International Journal of Psychophysiology 38:301–313.

Vroomen J, Bertelson P, Gelder BD (2001) The ventriloquist effect does not depend on the direction of automatic visual attention. Perception & Psychophysics 63:651–659.

Wang P, Göschl F, Friese U, König P, Engel AK (2019) Long-range functional coupling predicts performance: Oscillatory EEG networks in multisensory processing. NeuroImage 196:114–125.

Wang XJ (2010) Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. Physiological Reviews 90:1195–1268.

Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. Psychological Bulletin 88:638–667.

Wozny DR, Beierholm UR, Shams L (2008) Human trimodal perception follows optimal statistical inference. Journal of Vision 8:24–24.

Young MP (1993) The organization of neural systems in the primate cerebral cortex. Proceedings of the Royal Society of London Series B: Biological Sciences 252:13–18.

Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: auditory–motor interactions in music perception and production. Nature Reviews Neuroscience 8:547–558.

Zhang S, Xu M, Chang WC, Ma C, Hoang Do JP, Jeong D, Lei T, Fan JL, Dan Y (2016) Organization of long-range inputs and outputs of frontal cortex for top-down control. Nature Neuroscience 19:1733–1742.

Zimmer U, Macaluso E (2007) Processing of multisensory spatial congruency can be dissociated from working memory and visuo-spatial attention. Eur J Neurosci 26:1681–1691.

Zuanazzi A, Noppeney U (2018) Additive and interactive effects of spatial attention and expectation on perceptual decisions. Scientific Reports 8:6732.

Zuanazzi A, Noppeney U (2019) Distinct Neural Mechanisms of Spatial Attention and Expectation Guide Perceptual Inference in a Multisensory World. The Journal of Neuroscience 39:2301–2312.

Appendix A: Original publications

Study 1

Michail G, Keil J (2018) High cognitive load enhances the susceptibility to non-speech audiovisual illusions. *Scientific Reports* 8:11530. doi: 10.1038/s41598-018-30007-6

SCIENTIFIC REPORTS

OPEN

High cognitive load enhances the susceptibility to non-speech audiovisual illusions

Received: 5 April 2018
Accepted: 20 July 2018

Georgios Michail[1] & Julian Keil[1,2]

The role of attentional processes in the integration of input from different sensory modalities is complex and multifaceted. Importantly, little is known about how simple, non-linguistic stimuli are integrated when the resources available for sensory processing are exhausted. We studied this question by examining multisensory integration under conditions of limited endogenous attentional resources. Multisensory integration was assessed through the sound-induced flash illusion (SIFI), in which a flash presented simultaneously with two short auditory beeps is often perceived as two flashes, while cognitive load was manipulated using an n-back task. A one-way repeated measures ANOVA revealed that increased cognitive demands had a significant effect on the perception of the illusion while post-hoc tests showed that participants' illusion perception was increased when attentional resources were limited. Additional analysis demonstrated that this effect was not related to a response bias. These findings provide evidence that the integration of non-speech, audiovisual stimuli is enhanced under reduced attentional resources and it therefore supports the notion that top-down attentional control plays an essential role in multisensory integration.

When one tries to localize a singing bird flitting between the branches of a tree with luxuriant foliage, the combination of auditory and visual input information – as compared to using only auditory or visual information – will probably increase the accuracy and speed of the localisation process. Navigating in an uncertain world, abundant in multisensory objects, requires the constant combination of sensory cues across different modalities, a process known as multisensory integration[1] (MSI). Indeed, a large body of animal and human studies suggest that sensory processing and discrimination is sharpened when multisensory information is provided (animal studies:[2,3], human studies:[4–8]). Recent studies have shown that multisensory perception is not a hardwired routine, but on the contrary, it is influenced by a wide range of neurophysiological processes such as the power and phase of ongoing oscillations and several cognitive factors such as the level of attention and expectations[9].

Regarding specifically the role of attention during the integration of multisensory information, studies point towards a complex relationship that unfolds at different levels of sensory processing[10,11]. The exact nature of the MSI-attention interplay is largely determined by the involved sensory modalities that can, for instance, have different spatio-temporal detection accuracies[7,12–15]. Moreover, the interaction between MSI and attention is influenced by the specific characteristics of the stimuli such as the stimulus intensity (e.g., near- vs. supra-threshold stimuli[16]) and complexity (e.g., speech vs. simple audio-visual stimuli[17,18]. Additionally, this interaction is influenced by the conditions of sensory stimulation such as the noisiness of the background or task-specific requirements narrowing the perceiver's focus on one modality or a specific stimulus feature[11]. The degree of association between the unimodal components of a multisensory signal was also proposed as a factor that determines the extent of attentional effects on multisensory integration[19]. Based on this proposal, the integration of strongly associated unimodal signals (e.g., audio-visual input during natural speech) is less likely to be affected by attentional factors compared to unimodal signals that are weakly associated due to spatial, temporal or semantic incongruencies.

An ongoing debate revolves around the question whether and under which conditions the binding of multisensory stimuli occurs automatically (or pre-attentively) or is influenced by top-down attentional control[20]. Whereas some studies support that MSI is modulated by attention[21–23], others provide evidence that it can also take place in a pre-attentive, automatic way[24–27]. It seems, that the influence that attention exerts on MSI is defined

[1]Department of Psychiatry and Psychotherapy, Multisensory Integration Lab, Charité Universitätsmedizin Berlin, Berlin, Germany. [2]Biological Psychology, Christian-Albrechts-University Kiel, Kiel, Germany. Correspondence and requests for materials should be addressed to G.M. (email: georgios.michail@charite.de)

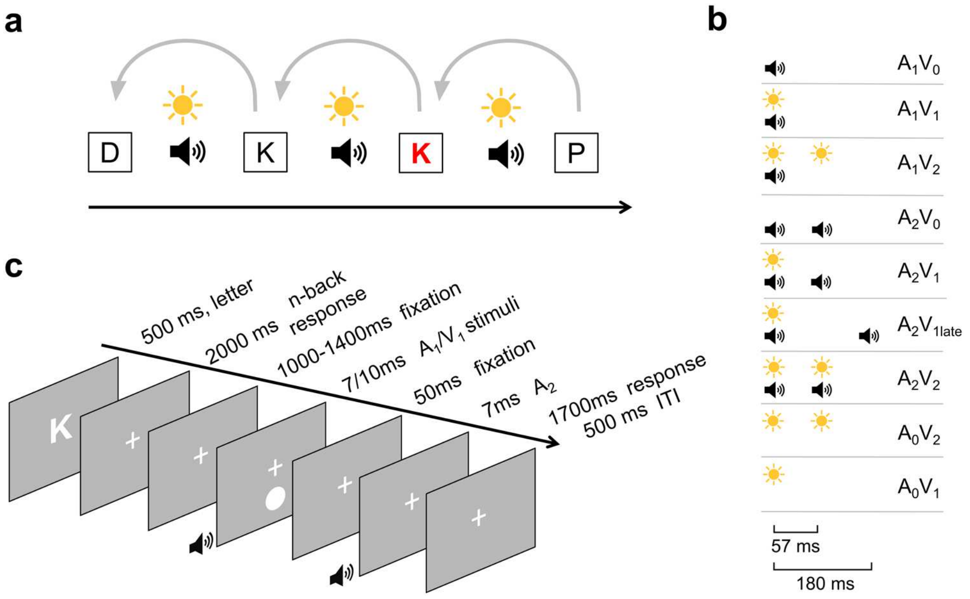

Figure 1. Schematic Illustration of experimental paradigm and material. (**a**) Representation of the dual task design for the 1-back condition. Participants were presented a letter and had to indicate if it matched the letter in the n-th previous trial. After the letter presentation the SIFI audiovisual stimuli were presented and participants had to report the number of perceived flashes. (**b**) The 9 audio-visual stimuli combinations that were used in the experiment. The stimulus onset asynchrony (SOA) was 57 ms for all combinations except the control condition A_2V_{1late} (180 ms SOA). (**c**) Illustration of a single critical A_2V_1 trial depicting the different parts of the trial, the intervals in between these parts, and the duration of the stimuli.

by the combined attentional effect of the bottom-up signalling and the endogenous attentional mechanisms[11]. The relative contribution of the two factors is to a large extent situation-dependent, and thus difficult to define precisely.

One approach to tackle this question is the use of a dual task paradigm in which one task is used to modulate the levels of endogenous resources available for the secondary task. Using this approach, Santangelo and Spence showed that under high perceptual load only audiovisual – and not auditory or visual – cues managed to capture visuo-spatial attention, indicating the effectiveness of multisensory stimulation in orienting spatial attention under high perceptual load[6].

The small number of studies that employed this dual task design to directly explore the effect of limited attentional resources on the integration of multisensory stimuli provided contradicting results[27–29]. Among these, two recent studies reported that audio-visual speech integration, as indexed by the McGurk effect – where a speech sound presented together with incongruent lip movement is perceived as a different, illusory, speech sound – was reduced under high attentional load[28,29]. On the contrary, Zimmer and Macaluso found that visuo-tactile spatial integration was insensitive to load manipulations of working memory and visuo-spatial attention[27]. Given the contradicting findings of these studies as well as their focus on different aspects of MSI (audio-visual speech perception[28,29]; visuo-tactile spatial integration[27]), it appears that several aspects related to the effect of load on MSI have not been sufficiently studied. Importantly, it remains an open question whether the finding of reduced audio-visual speech integration under high attentional load[28,29] is relevant for the binding of simple, non-linguistic, audio-visual information. The increased reliance on sound under load might be to a large extent a speech-specific characteristic as we naturally rely more on sound than on vision for speech recognition. Furthermore, audio-visual speech is suggested to be a specific type of multisensory integration[30,31]. Another limitation of speech as a paradigm for the investigation of MSI–attention interactions is that one cannot exclude participants' strategic use of pre-existing associations related to the semantic content of the stimuli[11,32].

The aim of the present study was to assess the effect of varying levels of cognitive load on the integration of simple, non-linguistic, audio-visual stimuli. We employed a robust audio-visual illusion paradigm, the so-called "Sound-Induced Flash Illusion" (SIFI) in which a single flash presented simultaneously with two auditory beeps is sometimes perceived as two flashes[33] (see Fig. 1). In the SIFI, the degree of audio-visual integration is assessed in terms of the illusion rate. We used an additional, orthogonal n-back task to manipulate the attentional resources that were available for the processing of multisensory input. We assume that increased working memory load requires additional resources, thus limiting resources available for other cognitive processes. This assumption is based on the influential model of working memory proposed by Baddeley and Hitch[34] (and updated later by Baddeley[35]) that involves the interaction of attentional control (performed by the central executive) with the maintenance of information in the storage systems (phonological loop and visuo-spatial sketchpad) and the episodic buffer. A growing body of behavioural and neuroscientific studies suggests that attention and working memory are functionally inter-twined and show extensive neuroanatomical overlap, involving fronto-parietal brain regions[36–40]. Importantly, Gazzaley and Nobre[41], taking into account neurophysiological evidence, propose

n-back level	0-back	1-back	2-back
Accuracy (%)	92.41 (19.18)	92.71 (5.01)	85.26 (11.91)
Reaction time (s)	0.40 (0.03)	0.55 (0.04)	0.70 (0.06)

Table 1. Mean (SD) reaction times and accuracy for the different n-back levels.

that the top-down modulatory mechanism underlying selective attention processes during perceptual processing is also engaged during the different stages of working memory – encoding, maintenance, and memory retrieval. Within this framework, we expect that an increase of working memory load due to the n-back task would present increased demands on cognitive resources thus limiting the resources available to attentional mechanisms for the processing of the upcoming audiovisual stimuli. Interestingly, an fMRI study showed that visuo-tactile spatial integration was independent from both working memory and attentional resources[27]. However, the finding of another study showing that audiovisual cues, as compared to unimodal, were more efficient in biasing access information in visuo-spatial working memory, indicates that multisensory integration can affect working memory performance[42].

To produce the SIFI illusion, participants were presented with a single flash paired with two auditory beeps in rapid succession. Eight other flash-beep combinations (of 0, 1 and 2 flashes and beeps) were used to control for perception of auditory-only, visual-only, and congruent audiovisual stimuli as well as for a response bias in the SIFI illusion (by using the same design as in illusory trials but with increased inter-beep interval) and an alternative illusory phenomenon called "Fusion" illusion – i.e., the illusory perception of a single flash when two flashes are paired with a single beep[43]. In all conditions, participants were asked to report the perceived number of flashes. The cognitive resources available for the SIFI task were manipulated by an n-back task performed prior to the SIFI task, in which participants were asked to indicate if the letter presented in the current trial matched the one presented *n* trials before. Varying the *n* (highest was 2), enabled us to examine the effect that different degrees of cognitive load have on the multisensory integration of simple audio-visual stimuli, as indexed by the SIFI illusion rate. An altered susceptibility to audiovisual illusions under high cognitive load and thus under reduced attentional resources would suggest a regulatory role of attention in audio-visual integration. Additionally, the direction of a possible effect – larger or smaller susceptibility – would provide further insights about the modulatory effect that attention exerts on multisensory integration. Based on the finding of reduced audio-visual speech integration (McGurk effect) under high attentional demands[28], one could expect a similar decline in the SIFI illusion rate under high cognitive load. However, whether this assumption is valid needs to be tested given the differences between the McGurk effect and the SIFI illusion in terms of the nature of the audio-visual stimuli (speech vs. non-speech) and the reported modality (sound vs. vision).

Results

N-back. The n-back task performance was assessed in terms of accuracy and reaction times (RTs). The mean RTs and accuracy, for all the different working memory (WM) load levels are displayed in Table 1. Data are provided throughout the text as mean and, in square brackets, standard deviation (SD), unless otherwise noted.

Our analysis revealed that WM load had a significant effect on accuracy (Friedman's test, $p = 0.002$, $\chi^2 = 12.79$, df = 2). Wilcoxon signed-rank post-hoc tests revealed that the subjects displayed significantly lower accuracy in 2-back compared to 0-back trials (85.26 [11.91] % for 2-back and 92.41 [19.18] % for 0-back, $Z = -2.44$; $p = 0.045$, $r = 0.35$) and 1-back trials (92.71 [5.01] % for 1-back, $Z = -2.43$; $p = 0.045$, $r = 0.35$).

Moreover, we found that reaction times, associated with correct responses to targets, were affected by WM load ($F_{(2, 30)} = 25.89$; $p < 0.001$). Post-hoc paired-samples *t*-tests showed that subjects displayed significantly longer RTs in the 2-back trials compared to both 1-back (mean RT was 0.70 [0.06] s for 2-back and 0.55 [0.04] s for 1-back, t(15) = 3.47, $p = 0.003$, BF = 13.45) and 0-back trials (mean RT was 0.40 [0.03] s for 0-back, t(15) = 6.12, $p < 0.001$, BF = 1220.57). Also, RTs in 1-back trials were longer than in 0-back (t(15) = 4.98, $p < 0.001$, BF = 184.46).

Overall, our data show that with increasing level of difficulty the subjects display lower accuracy and slower RTs. Thus, these results demonstrate the efficacy of the n-back task to modulate the working memory load and limit attentional resources.

Sound-induced Flash illusion. The performance in the SIFI task was assessed in terms of the number of perceived flashes and the reaction times. The mean (SD) percentage of the analysed flash responses in all n-back levels and combinations of audiovisual stimuli is reported along with a summary of the results of the statistical analyses in Table 2. The same information for the RTs is reported in Table 3.

Working memory load manipulation affects SIFI perception. To assess the effect of WM load on the perception of the sound-induced flash illusion, we analysed the differences in the illusion rate i.e. the percentage of "2-flashes" responses in the critical trials (A_2V_1) – between the different WM load levels. One-way repeated-measures ANOVA showed that WM load had a significant effect on the illusion rates ($F_{(3, 45)} = 4.56$, $p = 0.007$), in the direction of increasing illusion rates with increasing WM load as can be seen in Fig. 2a. Post-hoc comparisons revealed a significantly larger illusion perception in 2-back trials compared to no-back (mean illusion rate was 57.08 [27.32] % for 2-back and 41.46 [27.24] % for no-back, t(15) = 3.12, p = 0.043, BF = 7.27). Additionally, the Bayes Factors provided some evidence that the illusion perception was stronger in 2-back trials compared 0-back (mean illusion rate was 49.38 [28.00] % for 0-back, t(15) = 2.63, $p = 0.095$, BF = 3.25), stronger in 1-back trials compared to no-back (mean illusion rate was 51.88 [29.89] % for 1-back, t(15) = 1.96, $p = 0.23$,

AV condition	Flash Response	Mean (SD) percentage of analysed flash responses (%)				Statistical Analysis			
		no-back	0-back	1-back	2-back	*ANOVA*			
						F	$df_{hypothesis}$	df_{error}	*p*
A_2V_1*	2	41.46 (27.24)	49.38 (28.00)	51.88 (29.89)	57.08 (27.32)	4.56	3	45	0.007
						Friedman's Test			
						χ^2	df		*p*
A_0V_1	1	92.2 (8.86)	93.23 (5.46)	93.23 (8.18)	91.67 (12.17)	0.18	3		0.98
A_0V_2	2	92.71 (11.33)	89.06 (14.18)	93.23 (7.59)	91.15 (11.97)	2.07	3		0.56
A_1V_0	0	98.44 (4.53)	94.79 (8.54)	96.35 (5.24)	94.27 (7.89)	7.34	3		0.06
A_1V_1	1	98.44 (3.36)	96.88 (5.99)	96.35 (5.24)	97.40 (5.87)	2.15	3		0.54
A_1V_2*	1	26.04 (28.69)	16.15 (25.18)	19.27 (24.67)	16.67 (23.77)	6.05	3		0.11
A_2V_0	0	97.40 (3.99)	95.83 (6.09)	92.71 (11.33)	93.23 (11.06)	2.55	3		0.47
A_2V_2	2	95.31 (6.78)	96.88 (6.72)	99.48 (2.08)	98.96 (2.85)	8.40	3		0.04
A_2V_{1late}*	2	17.19 (23.27)	25.52 (22.25)	28.13 (29.48)	25.52 (22.04)	2.84	3		0.42

Table 2. Mean (SD) percentage of the analysed flash responses and a summary of the results of statistical analyses (ANOVA - Friedman's test) regarding the effect of WM on the perception of audiovisual stimuli in the SIFI task. *In the incongruent stimuli combinations, the analysed flash response was the "illusory" response (e.g., the 2-flash response in A_2V_1). In all the other control stimuli combinations, the analysis was performed on the percentage of correct responses.

AV condition	Mean (SD) Reaction Time (s)				Statistical Analysis			
	no-back	0-back	1-back	2-back	*ANOVA*			
					F	$df_{hypothesis}$	df_{error}	*p*
A_0V_1	0.624 (0.14)	0.683 (0.10)	0.693 (0.10)	0.714 (0.14)	4.86	3	45	0.005
A_0V_2	0.617 (0.15)	0.719 (0.11)	0.675 (0.12)	0.708 (0.14)	7.28	3	45	<0.001
A_1V_0	0.668 (0.14)	0.781 (0.16)	0.814 (0.19)	0.776 (0.15)	12.87	3	45	<0.001
A_1V_2	0.670 (0.18)	0.733 (0.13)	0.728 (0.16)	0.737 (0.17)	2.80	2.15	32.24	0.072
A_2V_1	0.725 (0.18)	0.761 (0.13)	0.787 (0.13)	0.777 (0.14)	2.51	2.30	34.54	0.089
A_2V_{1late}	0.743 (0.14)	0.853 (0.17)	0.839 (0.17)	0.825 (0.12)	4.57	3	45	0.007
					Friedman's Test			
					χ^2	df		*p*
A_1V_1	0.568 (0.15)	0.643 (0.14)	0.654 (0.13)	0.663 (0.13)	9.38	3		0.025
A_2V_0	0.690 (0.17)	0.822 (0.17)	0.808 (0.19)	0.804 (0.17)	20.63	3		<0.001
A_2V_2	0.594 (0.18)	0.652 (0.14)	0.682 (0.15)	0.687 (0.15)	16.73	3		0.001

Table 3. Mean (SD) RTs and a summary of the results of statistical analyses (ANOVA - Friedman's test) regarding the effect of WM on the RTs after the presentation of AV stimuli in the SIFI task.

BF = 1.17) and stronger in 0-back compared to no-back (t(15) = 2.06, $p = 0.23$, BF = 1.35). We repeated the same one-way repeated measures ANOVA using "normalised" illusion rates (obtained after dividing the percentage of 2-flashes responses in A_2V_1 by A_0V_2) and found a significant main effect of WM load ($F_{(2.25,\ 33.73)} = 3.89$; $p = 0.026$), indicating that the observed effect is not affected by possible individual biases in the perception of two flashes. These findings indicate that the illusory perception of two-flashes when a single flash is presented together with two auditory stimuli is enhanced when the attentional resources are limited.

In addition, we also explored whether WM load affected the magnitude of the SIFI illusion perception in the group of excluded subjects (N = 14). We found that in this highly heterogeneous group there was no significant effect of WM load on the perception of the SIFI illusion (i.e., percentage of 2-flash responses in the A2V1 trials; Friedman's test, p = 0.15, $\chi 2 = 5.38$, df = 3). A separate analysis for each excluded subgroup was not possible due to the small number of subjects in the subgroups.

To examine whether the effect of WM load on the illusion perception was related to a response bias – as compared to an effect on perceptual mechanisms – we performed a similar analysis on the control condition A_2V_{1late} in which the second auditory stimulus was presented with increased latency compared to the A_2V_1 trials (see Fig. 1b). Figure 3 represents participants' illusion rate and the RTs for A_2V_{1late} trials. If the participants' illusion rates ("2-flash" responses) in the critical A_2V_1 trials were based on a reflective response to the number of presented auditory stimuli, we would expect to find an effect of WM load on the "2-flash" responses for the A_2V_{1late}, similar to the effect found in $A_2V_{1.}$ Our analysis revealed that WM load had no significant effect on the percentage "2-flash" responses in the A_2V_{1tlate} trials (Friedman's test, $p = 0.42$, $\chi^2 = 2.84$, df = 3). This result indicates that the observed effect of WM load on the illusion perception (in A_2V_1) was not related to a response bias but was primarily due to changes on perceptual mechanisms linked to varying levels of working memory load.

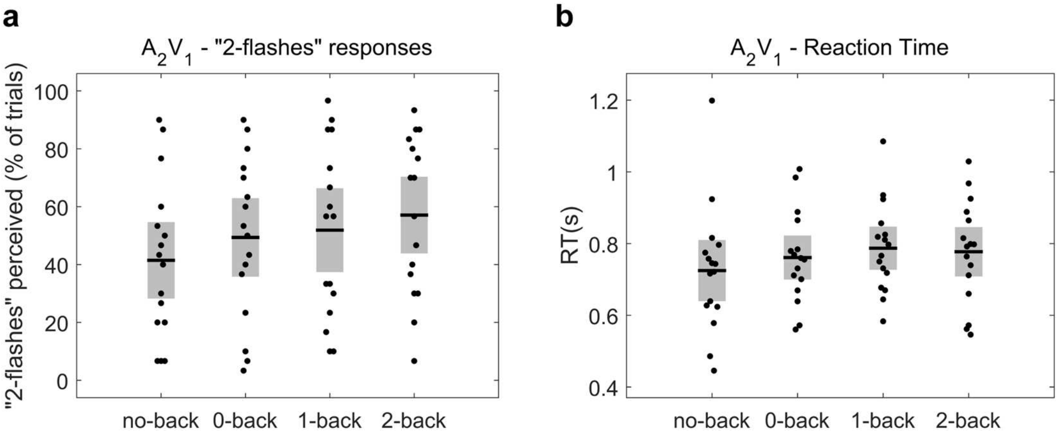

Figure 2. Increased illusion rate under high working memory load in critical A_2V_1 trials. (**a**) The percentage "2-flashes" responses in A_2V_1 trials – i.e., illusion rate – for the different working memory levels. A one-way ANOVA revealed that working memory load had a significant effect on the illusion perception. Post-hoc comparisons showed that illusion perception in 2-back was significantly higher compared to no-back and relatively higher compared to 0-back (**b**) The RTs for reporting the perceived flashes number in A_2V_1 trials for all the working memory load levels. No significant effect of load on the RTs was found. Horizontal black lines denote the mean and grey bars the standard error of the mean.

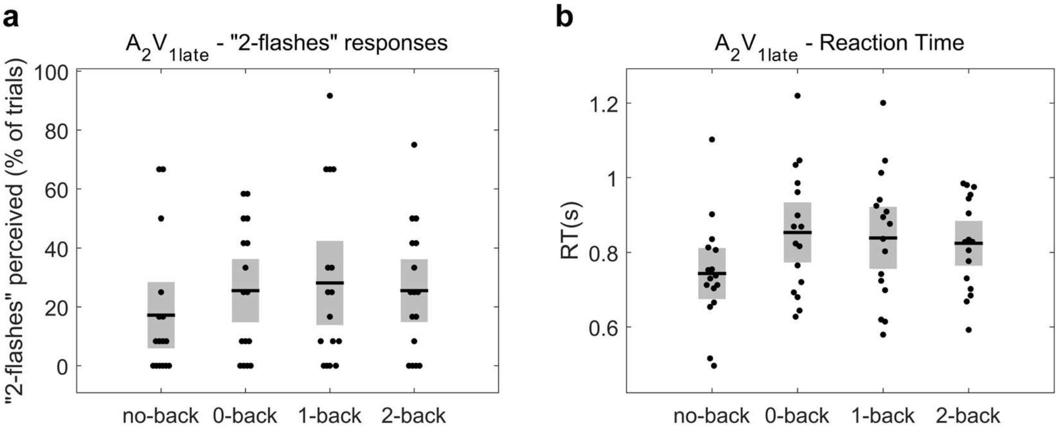

Figure 3. Working memory load doesn't affect illusion rates in control A_2V_{1late} trials. (**a**) The percentage "2-flashes" responses in A_2V_{1late} trials for the different working memory levels. The analysis revealed that working memory load had no significant effect on the illusion perception. (**b**) The RTs for reporting the perceived flashes number in A_2V_{1late} trials for all the working memory load levels. Horizontal black lines denote the mean and grey bars the standard error of the mean.

Additionally, a similar analysis for the A_1V_2 condition – associated with the "Fusion" illusion in which two flashes are 'fused' and perceived as one –, revealed that WM load had no significant effect on the percentage "1-flash" responses (Friedman's test, $p = 0.11$, $\chi^2 = 6.05$, df = 3; Fig. 4). This finding suggests that the effect that WM load has on audio-visual perception might be specifically related to the perceptual mechanisms underlying the SIFI – "Fission" –illusion (one flash perceived as two) that differ from the processes underlying the "Fusion" illusion[43]. The effect of WM on the percentage of correct responses in all the other control conditions was also investigated. The results of the statistical analyses are reported in Table 2. There was no significant effect of WM load on the correct responses in any of the control conditions except the A_2V_2 (Friedman's test, $p = 0.04$, $\chi^2 = 8.40$, df = 3). However, the percentage of "2-flashes" responses was not statistically different between the different n-back levels ($p > 0.05$ in all post-hoc pairwise comparisons). Therefore, and given the small number of A_2V_2 trials (12 per n-back level) – that could lead to inflated percentage differences –, this result should be interpreted with caution and needs to be verified in further studies using larger trial numbers.

Working memory load manipulation and reaction times. At the next step, we first examined whether WM load affected the reaction times of participants when they reported the perceived flashes number, for the audio-visual combinations A_2V_1, A_2V_{1late} and A_1V_2. We found that WM load had no significant effect on the RTs for the critical A_2V_1 trials ($F_{(2.30,\,34.54)} = 2.51$, $p = 0.089$; Fig. 2b), as well as on the RTs for the A_1V_2 trials ($F_{(2.15,\,32.24)} = 2.80$, $p = 0.072$; Fig. 3b). In contrast, WM load had a significant effect on the RTs for the A_2V_{1late} trials

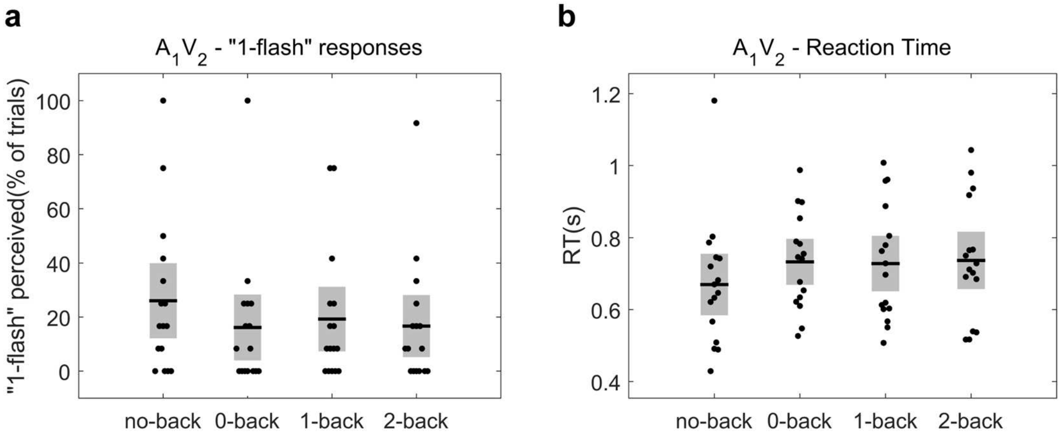

Figure 4. No influence of working memory load on the "Fusion" illusory percept in A_1V_2 trials. (**a**) The percentage "1-flash" responses in A_1V_2 trials – indexing the strength of the "Fusion" percept – for the different working memory levels. Our analysis demonstrated no significant effect of working memory load on the "Fusion" perception. (**b**) The RTs for reporting the perceived flash number in A_1V_2 trials for all the working memory load levels. No significant effect of load on the RTs was found. Horizontal black lines denote the mean and grey bars the standard error of the mean.

($F_{(3, 45)} = 4.57$, $p = 0.007$; Fig. 4b). Post-hoc tests show that in A_2V_{1late} trials the RTs were significantly slower in 2-back compared to no-back, (mean RT was 0.825 [0.12] s for 2-back vs. 0.743 [0.14] s for no-back, t(15) = 3.13, $p = 0.041$, BF = 7.48) and provide some evidence for slower RTs in 1-back compared to no-back (mean RT for 1-back was 0.839 [0.17] s, t(15) = 2.59, $p = 0.082$, BF = 3.05) and in 0-back compared to no-back (mean RT for 0-back was 0.853 [0.17] s, t(15) = 2.89, $p = 0.056$, BF = 4.98).

Interestingly, a similar analysis performed in the other control conditions (A_0V_1, A_0V_2, A_1V_0, A_1V_1, A_2V_0, A_2V_2) showed that WM had a significant effect on RTs in all conditions (mean (SD) RT and summary of the results of statistical analyses can be found in Table 3), in the direction of larger RTs with increasing WM load (e.g., in $A_0V_{2,}$ mean RT was 0.708 [0.14] s for 2-back and 0.617 [0.15] s for no-back, t(15) = 3.48, $p = 0.017$, BF = 13.64). Therefore, these results show that increased WM load resulted in a general slowing of the response after the presentation of audio-visual stimuli except in trials with incongruent audio-visual stimuli (A_2V_1, A_1V_2). This might be related to the fact that in these two conditions the RT was shaped not only by WM load but also by the incongruency between the auditory and visual stimuli, the degree of which was – contrary to WM load – unvarying across n-back levels.

Discussion

In the present study, we analysed the audiovisual integration of simple, non-linguistic stimuli as indexed by the strength of the SIFI effect, under different levels of working memory load. We used an n-back task to manipulate the amount of cognitive resources that were available for the processing of the SIFI stimuli. Our main result was that participants displayed enhanced susceptibility to the SIFI under high working memory load. The absence of such effect in the control condition, in which the inter-beep interval was increased, argues against the possibility of our main result being associated with a response bias. Our finding provides strong evidence that audiovisual integration can be modulated by the amount of available cognitive resources and it therefore argues against a pre-attentive account of multisensory integration.

Previous studies, using a range of multisensory tasks, have demonstrated that cross-modal binding can be immune[24,25], but it can also be sensitive[21,44,45] to spatial attention manipulations. Also, it's unclear whether MSI is affected[46] or not[47] by modality-specific attention. Although these studies examined particular aspects of the role of attention during the integration of multisensory signals (spatial attention, modality-specific attention), the present study addressed a critically different question. We manipulated endogenous attention using a secondary, orthogonal, n-back task and asked whether multisensory integration is affected under conditions of increasingly limited attentional resources.

The present study is the first, to our knowledge, to demonstrate that limiting the available attentional resources prior to the stimulus delivery enhances the integration of simple, non-linguistic, audiovisual signals. Some previous studies using a similar approach reported that visuo-tactile spatial integration[27] and the integration of emotional cues in songs[48] are not affected by increased attentional demands. These studies are not necessarily in conflict with our findings, because their focus and design are in several aspects different to the present study. Zimmer and Macaluso[27] investigated spatial integration of visual-tactile cues, whereas we focus on the temporal integration of audiovisual stimuli. Also, in Thompson *et al.*[48], the complexity of the material (songs) and the level of integration process (emotional cue binding) are quite higher compared to the corresponding features of the present study (low-level binding of simple audiovisual stimuli). There is evidence that multisensory events involving different combinations of sensory modalities or stimulus characteristics activate different brain networks[49], and that the nature of a particular multisensory event affects its susceptibility to attentional manipulations[11,16].

Yet, previous studies on audiovisual temporal integration showed that audiovisual speech perception, as indexed by the McGurk effect, is sensitive to attentional load manipulations[28,29,50]. Interestingly, these studies demonstrated a reduction in the perception of the McGurk effect under high attentional load, a finding that appears to contradict our results. However, the SIFI and McGurk illusions are characterised by distinct temporal integration properties[51] and such differences can account for the discrepancy (see next paragraph). We also assume that the experimental design of the present study has some advantages over previous studies. First, the use of simple, non-linguistic stimuli in the present study, excludes the strategic use of pre-existing associations to which speech is sensitive[11,32]. Second, the design of the secondary task in the previous McGurk studies[28,50] involving the presentation of visual or auditory objects (e.g., shapes superimposed on the faces showing the speech gestures) sometimes temporally overlapping with the McGurk stimuli[28] complicates the interpretation of the results. This is supported by the discrepancy between these studies, in regard to whether the reduction of the McGurk illusion is attributed to the depletion of attentional resources[28] or to modality-specific attention[50]. Importantly, the attentional manipulation in the present study through an n-back task was temporally separated from the primary task (SIFI). This excludes or limits to a great extent the potential interference of secondary task material with the audiovisual processing of primary task stimuli. As such, our experimental design enables the firm conclusion that the enhanced audiovisual integration was induced by the reduced levels of endogenous attentional resources that were available for the processing of the multisensory input.

A mechanism that could account for the present findings relates to the temporal window of integration (TWI), i.e., the maximum temporal asynchrony between two different sensory events that allows their perceptual binding into a singular percept[52]. Previous work has demonstrated that susceptibility to audio-visual illusions such as the SIFI can be predicted by individual differences in the temporal window of integration[51]. The integration window increases with age[53] and it can be recalibrated after exposure to asynchronous stimuli[54,55]. Moreover, the TWI can be adaptively adjusted depending on the task demands[56]. Therefore, it can be assumed that increasing the attentional demands in our experiment might have resulted in an adaptive widening of the individual TWI that in turn led to the enhanced binding of the audiovisual input. This mechanism can also explain the discrepancy between our observation of enhanced SIFI perception under load and the decline of the McGurk effect under increased attentional demands, reported by Alsius *et al.*[28,29]. Previous work has shown that larger TWI is associated with increased susceptibility to SIFI and reduced susceptibility to the McGurk effect[51]. Therefore, a widening of the integration window, induced by high cognitive load, could enhance the illusory perception in SIFI but, on the contrary, diminish the susceptibility to the McGurk effect. What neural mechanism could account for this putative effect? Given the role of alpha oscillations in the temporal sequencing of audio-visual signals[57], and the temporal resolution in visual[58] and SIFI-type audiovisual perception[59,60], it can be hypothesized that a modulation of the alpha oscillations induced by varying attentional demands might have mediated the changes in sensory processing that led to the increased illusory percept. Interestingly, Cecere *et al.*[60] showed that modulating the individual alpha frequency using electrical stimulation resulted in changes in the TWI. Whether varying attentional demands also modulate neural oscillations, and whether this can explain the current findings requires further testing using electrophysiological methods (M/EEG, ECoG).

Our findings may also be explained based on the "attentional load theory", which postulates that when high-level cognitive processes are loaded, the processing of task-irrelevant information is enhanced[61]. Because in our experimental design auditory stimuli are less relevant than visual – since participants are instructed to report the number of flashes – it's possible that under limited attentional resources the auditory input gained a larger sensory weight, which resulted in the enhanced illusion rate. Similarly, the "gating-by-inhibition" hypothesis posits that alpha band oscillations optimize stimulus processing by inhibiting task-irrelevant cortical areas[62]. Increased attentional demands might have interfered with this gating mechanism, coincidentally enhanced neural excitability and increased the crossmodal influence[63].

Taken together, our findings highlight the influence that attention exerts on audiovisual integration and suggest that when attentional resources are depleted, the cross-modal binding of simple, non-linguistic audiovisual signals is enhanced. These results are especially relevant for the understanding of the interplay between attention and multisensory integration because they provide strong evidence against a pre-attentive account of audio-visual temporal integration. Characterizing this interaction at the behavioural level is an essential first step[64]. Further neuroimaging and electrophysiological studies could provide insights about the neural correlates of this interaction and the stage of sensory processing at which attentional effects occur. Further studies could also include lure trials in order to control for the use of familiarity signals during the n-back task[65]. Another interesting question that should be addressed in future investigations is whether there is a different effect of cognitive load on multisensory integration between target and non-target n-back trials. This question couldn't be addressed in the context of the current study, due to the small number of target A_2V_1 trials and the different ratio of target to non-target trials between the different n-back levels. In the current experiment, the use of a fixed inter-beep interval in the critical A_2V_1 trials might have resulted in the extremely high and low SIFI illusion rates we observed in some of the excluded subjects. To alleviate this, future studies could adjust the inter-beep interval individually to account for the inter-individual variability in the temporal window of integration[51,66].

Methods

Subjects. Thirty healthy subjects (10 males, mean age = 29.9 years, SD = 7.8, range = 20–56) participated in this study after providing written informed consent. All participants reported normal hearing, normal or corrected-to-normal vision and absence of any neurological condition. The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the Charité–Universitätsmedizin Berlin.

Task Design. The subjects performed a dual task paradigm (Fig. 1a) that combined a visual verbal n-back task and the SIFI paradigm adapted from Shams *et al.*[67]. The n-back task was used to present increased demands on working memory (WM) and therefore reduce the available cognitive resources for the processing of the SIFI audio-visual stimuli. The experiment consisted of 12 blocks corresponding to 3 blocks for each of the 4 levels of WM load (no-back, 0-back, 1-back, 2-back). Each experimental block contained 10 critical A_2V_1 trials (two auditory and one visual stimulus) and 4 trials for each of the other 8 combinations (see *SIFI audio-visual stimuli*, Fig. 1b). In total, the experiment included for each WM load category, 30 critical A_2V_1 trials and 12 trials for all the other combinations. The order of the blocks was randomized across participants. The experiment was conducted in a sound-attenuated chamber using a portable computer (HP Pavilion 17) and lasted for around 43 min, excluding the short breaks between the blocks. Participants performed 10 practice trials for each WM load category, prior to the main part of the experiment. The Psychophysics toolbox[68] for MATLAB (The Mathworks, Natick, MA, USA) was used for presenting the stimuli and obtaining the responses. Data analyses were performed using MATLAB (The Mathworks, Natick, MA, USA) and SPSS software (SPSS Inc., Chicago, IL, USA).

As illustrated in Fig. 1c, each trial of the 0-, 1- and 2-back blocks started with a letter presented for 500 ms, followed by a window of 2000 ms in which the participants were instructed to indicate if the currently presented letter matched the one presented *n* trials before (1-back and 2-back) or with a predefined letter "X" (0-back). No response was required for non-targets. After a randomized 1000–1400 ms window, participants were presented with a combination of auditory and visual stimuli. In the case of A_2V_2 combination, a pair of – temporally aligned – visual and auditory stimuli was presented followed by another similar pair after a time lag of 57 ms. The visual stimulus was presented for 10 ms and the auditory stimulus for 7 ms. The same timing was followed in all the other combinations (see *SIFI audio-visual stimuli*) except the control condition A_2V_{1late} in which the second auditory stimulus was presented 180 ms after the first auditory stimulus (based on Mishra *et al.*[69]). Directly after the last stimulus, in the response window (1700 ms), the participants indicated the number of perceived flashes (0, 1, 2). The next trial started after an inter-trial interval (ITI) of 500 ms.

The trials of the no-back blocks were structured in the same way, excluding the 'n-back'-related periods (letter presentation and the subsequent response window). A fixation cross was displayed throughout the entire trial length. Participants responded with the right thumb (number of flashes) or index finger (n-back targets) using a handheld gamepad (Logitech Gamepad F310, Logitech, Lausanne, Switzerland).

n-back task stimuli. The stimuli for the n-back task were upper case letters presented in white on a neutral grey background, at the centre of the screen. For each block a (pseudo)random sequence of letters was selected from the set of English consonants. To avoid the use of phonemes as a strategy, vowels were excluded[70]. In the 0-back trials, the target was always the letter "X". To ensure equal difficulty in all the 2-back sequences, we explicitly manipulated the sequences to exclude the occurrence of – potentially confusing – lure trials, that is, trials in which the presented letter is the same with the one presented in the previous trial, but different to the letter presented 2 trials before. In each sequence, 33% of the letters were targets.

SIFI audio-visual stimuli. Nine stimulus combinations were presented (Fig. 1b), consisting of 0, 1 or 2 auditory (A) stimuli combined with either 0, 1 or 2 visual (V) stimuli (A_0V_1, A_0V_2, A_1V_0, A_1V_1, A_1V_2, A_2V_0, A_2V_1, A_2V_{1late}, A_2V_2). The visual stimulus was a white disk subtending a visual angle of 1.6° and was presented at 4.1° centrally below the fixation cross, for 10 ms. The auditory stimulus was a 78 dB (SPL) 1000 Hz sine wave tone and was presented for 7 ms with the use of an amplifier (UR22mkII, Steinberg) and a 6.1 cm long, 4 mm wide tube system (ER30, Etymotic Research).

Data Analysis. The n-back performance was assessed in terms of the accuracy and reaction times (RTs). The accuracy was quantified as the proportion of hits (i.e., correct responses when there was a target letter) minus the misses and false alarms (i.e., responses when there was a non-target letter) over the total number of targets. Regarding the audio-visual stimulation, performance was assessed by estimating, for each combination, the RTs and the proportion of trials when participants reported 0, 1 or 2 perceived flashes.

Previous studies have shown that there is considerable inter-individual variability regarding the perception of the SIFI[71,72]. For the purposes of our study, we focused on subjects that reliably perceived the illusion. Therefore, 8 subjects that didn't perceive the illusion during the critical A_2V_1 trials (i.e., they perceived "2-flashes" in less than 10% or more than 90% of trials[59]) in at least 2 of the 4 conditions (no-back, 0-back, 1-back, 2-back) were excluded from the analysis. Additionally, 6 participants were excluded from the analysis as they markedly failed in perceiving the 2-flashes in the control condition A_0V_2 ("2-flashes" response in less than 60% of trials in the "no-back" blocks). In total, 14 subjects were excluded from the analysis. The final sample size ($N = 16$) is relatively small and limits the external validity of the current findings, however previous SIFI studies showed robust effects using small sample sizes ($N = 8$ in studies by Shams and colleagues[33,67]).

Statistical analysis. The statistical significance of the differences in the evaluated parameters (RT and illusion rate) between the different working memory (WM) load conditions was analysed using a repeated-measures analysis of variance (ANOVA). The Mauchly test was used to verify the assumption of sphericity and the Greenhouse-Geisser correction was applied when necessary to correct for non-sphericity. For these cases, the corrected degrees of freedom and *p*-values are reported. Further analysis of the significant effects was performed using post-hoc paired *t*-tests and the Bayes Factor[73] (BF) as an indicator of the relative evidence. BFs between 1–3 indicate anecdotal support for the alternative hypothesis (H1) while BF between 3–10 and above 10 indicate respectively moderate and strong support for H1. BF = 1 indicates equal support for H1 and null hypothesis (H0) while BF between 1/3–1, 1/10–1/3 and below 1/10, provide respectively anecdotal, moderate and strong support for H0[74].

If the data were not normally distributed (failing the Lilliefors test for normality of distribution at alpha level 0.05) we used the Friedman test – a non-parametric alternative to one-way repeated measures ANOVA –, and post-hoc Wilcoxon signed-rank tests to evaluate differences between conditions. In association with each pairwise Wilcoxon test, we report the effect size (r; $r = Z / n^{1/2}$, Z = Wilcoxon Z-value, n = number of observations).

The Holm-Bonferroni correction[75] was applied for the all the post-hoc pairwise comparisons. An alpha level of 0.05 is used for all statistical tests.

Data availability. The datasets analysed during the current study are available from the corresponding author on reasonable request.

References

1. Stein, B. E. *The New Handbook of Multisensory Processing*. (MIT Press, 2012).
2. Bell, A. H., Meredith, M. A., Van Opstal, A. J. & Munoz, D. P. Crossmodal Integration in the Primate Superior Colliculus Underlying the Preparation and Initiation of Saccadic Eye Movements. *J. Neurophysiol.* **93**, 3659–3673 (2005).
3. Rowland, B. A., Quessy, S., Stanford, T. R. & Stein, B. E. Multisensory Integration Shortens Physiological Response Latencies. *J. Neurosci.* **27**, 5879–5884 (2007).
4. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
5. Battaglia, P. W., Jacobs, R. A. & Aslin, R. N. Bayesian integration of visual and auditory signals for spatial localization. *JOSA A* **20**, 1391–1397 (2003).
6. Santangelo, V. & Spence, C. Multisensory cues capture spatial attention regardless of perceptual load. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 1311–1321 (2007).
7. Spence, C. Crossmodal spatial attention. *Ann. N. Y. Acad. Sci.* **1191**, 182–200 (2010).
8. Frassinetti, F., Bolognini, N. & Làdavas, E. Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp. Brain Res.* **147**, 332–343 (2002).
9. Keil, J. & Senkowski, D. Neural Oscillations Orchestrate Multisensory Processing. *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* 1073858418755352, https://doi.org/10.1177/1073858418755352 (2018).
10. Talsma, D., Senkowski, D., Soto-Faraco, S. & Woldorff, M. G. The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* **14**, 400–410 (2010).
11. Macaluso, E. *et al.* The Curious Incident of Attention in Multisensory Integration: Bottom-up *vs*. Top-down. *Multisensory Res.* **29**, 557–583 (2016).
12. Ward, L. M. Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Can. J. Exp. Psychol. Rev. Can. Psychol. Exp.* **48**, 242–259 (1994).
13. Talsma, D. & Kok, A. Intermodal spatial attention differs between vision and audition: An event-related potential analysis. *Psychophysiology* **39**, 689–706 (2002).
14. Van Damme, S., Gallace, A., Spence, C., Crombez, G. & Moseley, G. L. Does the sight of physical threat induce a tactile processing bias?: Modality-specific attentional facilitation induced by viewing threatening pictures. *Brain Res.* **1253**, 100–106 (2009).
15. Gray, R., Mohebbi, R. & Tan, H. Z. The Spatial Resolution of Crossmodal Attention: Implications for the Design of Multimodal Interfaces. *ACM Trans Appl Percept* **6**, 4:1–4: 14 (2009).
16. Koelewijn, T., Bronkhorst, A. & Theeuwes, J. Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychol. (Amst.)* **134**, 372–384 (2010).
17. Vouloumanos, A., Kiehl, K. A., Werker, J. F. & Liddle, P. F. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J. Cogn. Neurosci.* **13**, 994–1005 (2001).
18. Hugdahl, K., Thomsen, T., Ersland, L., Rimol, L. M. & Niemi, J. The effects of attention on speech perception: an fMRI study. *Brain Lang.* **85**, 37–48 (2003).
19. Santangelo, V. & Macaluso, E. Spatial attention and audiovisual processing. In *The New Handbook of Multisensory Processes* 359–370 (2012).
20. Hartcher-O'Brien, J., Soto-Faraco, S. & Adam, R. Editorial: A Matter of Bottom-Up or Top-Down Processes: The Role of Attention in Multisensory Integration. *Front. Integr. Neurosci.* **11** (2017).
21. Talsma, D., Doty, T. J. & Woldorff, M. G. Selective Attention and Audiovisual Integration: Is Attending to Both Modalities a Prerequisite for Early Integration? *Cereb. Cortex* **17**, 679–690 (2007).
22. Fairhall, S. L. & Macaluso, E. Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *Eur. J. Neurosci.* **29**, 1247–1257 (2009).
23. Klemen, J., Büchel, C., Bühler, M., Menz, M. M. & Rose, M. Auditory working memory load impairs visual ventral stream processing: toward a unified model of attentional load. *J. Cogn. Neurosci.* **22**, 437–446 (2010).
24. Bertelson, P., Vroomen, J., De Gelder, B. & Driver, J. The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept. Psychophys.* **62**, 321–332 (2000).
25. Vroomen, J., Bertelson, P. & Gelder, B. D. The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept. Psychophys.* **63**, 651–659 (2001).
26. Stekelenburg, J. J., Vroomen, J. & de Gelder, B. Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neurosci. Lett.* **357**, 163–166 (2004).
27. Zimmer, U. & Macaluso, E. Processing of multisensory spatial congruency can be dissociated from working memory and visuo-spatial attention. *Eur. J. Neurosci.* **26**, 1681–1691 (2007).
28. Alsius, A., Navarra, J., Campbell, R. & Soto-Faraco, S. Audiovisual Integration of Speech Falters under High Attention Demands. *Curr. Biol.* **15**, 839–843 (2005).
29. Alsius, A., Navarra, J. & Soto-Faraco, S. Attention to touch weakens audiovisual speech integration. *Exp. Brain Res.* **183**, 399–404 (2007).
30. Tuomainen, J., Andersen, T. S., Tiippana, K. & Sams, M. Audio–visual speech perception is special. *Cognition* **96**, B13–B22 (2005).
31. Vatakis, A., Ghazanfar, A. A. & Spence, C. Facilitation of multisensory integration by the 'unity effect' reveals that speech is special. *J. Vis.* **8**(14), 1–11 (2008).
32. Doehrmann, O. & Naumer, M. J. Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Res.* **1242**, 136–150 (2008).
33. Shams, L., Kamitani, Y. & Shimojo, S. Illusions: What you see is what you hear. *Nature* **408**, 788 (2000).
34. Baddeley, A. D. & Hitch, G. Working Memory. in *Psychology of Learning and Motivation* (ed. Bower, G. H.) **8**, 47–89 (Academic Press, 1974).
35. Baddeley, A. The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* **4**, 417–423 (2000).
36. LaBar, K. S., Gitelman, D. R., Parrish, T. B. & Mesulam, M.-M. Neuroanatomic Overlap of Working Memory and Spatial Attention Networks: A Functional MRI Comparison within Subjects. *NeuroImage* **10**, 695–704 (1999).
37. Kane, M. J. & Engle, R. W. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychon. Bull. Rev.* **9**, 637–671 (2002).
38. Awh, E. & Jonides, J. Overlapping mechanisms of attention and spatial working memory. *Trends Cogn. Sci.* **5**, 119–126 (2001).
39. Corbetta, M., Kincade, J. M. & Shulman, G. L. Neural systems for visual orienting and their relationships to spatial working memory. *J. Cogn. Neurosci.* **14**, 508–523 (2002).
40. Awh, E., Vogel, E. K. & Oh, S.-H. Interactions between attention and working memory. *Neuroscience* **139**, 201–208 (2006).

41. Gazzaley, A. & Nobre, A. C. Top-down modulation: bridging selective attention and working memory. *Trends Cogn. Sci.* **16**, 129–135 (2012).
42. Botta, F. *et al*. Multisensory integration affects visuo-spatial working memory. *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 1099–1109 (2011).
43. Andersen, T. S., Tiippana, K. & Sams, M. Factors influencing audiovisual fission and fusion illusions. *Cogn. Brain Res.* **21**, 301–308 (2004).
44. Talsma, D. & Woldorff, M. G. Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J. Cogn. Neurosci.* **17**, 1098–1114 (2005).
45. Senkowski, D., Talsma, D., Herrmann, C. S. & Woldorff, M. G. Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Exp. Brain Res.* **166**, 411–426 (2005).
46. Vercillo, T. & Gori, M. Attention to sound improves auditory reliability in audio-tactile spatial optimal integration. *Front. Integr. Neurosci*. **9** (2015).
47. Helbig, H. B. & Ernst, M. O. Visual-haptic cue weighting is independent of modality-specific attention. *J. Vis.* **8**, 21–21 (2008).
48. Thompson, W., Russo, F. & Quinto, L. Audio-visual integration of emotional cues in song. *Cogn. Emot.* **22**, 1457–1470 (2008).
49. Calvert, G. A. & Thesen, T. Multisensory integration: methodological approaches and emerging principles in the human brain. *J. Physiol.-Paris* **98**, 191–205 (2004).
50. Alsius, A., Möttönen, R., Sams, M. E., Soto-Faraco, S. & Tiippana, K. Effect of attentional load on audiovisual speech perception: evidence from ERPs. *Front. Psychol*. **5** (2014).
51. Stevenson, R. A., Zemtsov, R. K. & Wallace, M. T. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 1517–1529 (2012).
52. van Wassenhove, V., Grant, K. W. & Poeppel, D. Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* **45**, 598–607 (2007).
53. Baum, S. H. & Stevenson, R. A. Shifts in Audiovisual Processing in Healthy Aging. *Curr. Behav. Neurosci. Rep.* **4**, 198–208 (2017).
54. Fujisaki, W., Shimojo, S., Kashino, M. & Nishida, S. Recalibration of audiovisual simultaneity. *Nat. Neurosci.* **7**, 773–778 (2004).
55. Van der Burg, E., Alais, D. & Cass, J. Rapid Recalibration to Audiovisual Asynchrony. *J. Neurosci.* **33**, 14633–14637 (2013).
56. Mégevand, P., Molholm, S., Nayak, A. & Foxe, J. J. Recalibration of the Multisensory Temporal Window of Integration Results from Changing Task Demands. *PLOS ONE* **8**, e71608 (2013).
57. Grabot, L., Kösem, A., Azizi, L. & van Wassenhove, V. Prestimulus Alpha Oscillations and the Temporal Sequencing of Audiovisual Events. *J. Cogn. Neurosci.* **29**, 1566–1582 (2017).
58. Samaha, J. & Postle, B. R. The Speed of Alpha-Band Oscillations Predicts the Temporal Resolution of Visual Perception. *Curr. Biol.* **25**, 2985–2990 (2015).
59. Keil, J. & Senkowski, D. Individual Alpha Frequency Relates to the Sound-Induced Flash Illusion. *Multisensory Res.* **30**, 565–578 (2017).
60. Cecere, R., Rees, G. & Romei, V. Individual Differences in Alpha Frequency Drive Crossmodal Illusory Perception. *Curr. Biol.* **25**, 231–235 (2015).
61. Lavie, N. Distracted and confused?: Selective attention under load. *Trends Cogn. Sci.* **9**, 75–82 (2005).
62. Jensen, O. & Mazaheri, A. Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Front. Hum. Neurosci*. **4** (2010).
63. Lange, J., Keil, J., Schnitzler, A., van Dijk, H. & Weisz, N. The role of alpha oscillations for illusory perception. *Behav. Brain Res.* **271**, 294–301 (2014).
64. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* **93**, 480–490 (2017).
65. Szmalec, A., Verbruggen, F., Vandierendonck, A. & Kemps, E. Control of interference during working memory updating. *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 137–151 (2011).
66. Stone, J. V. *et al*. When is now? Perception of simultaneity. *Proc. R. Soc. Lond. B Biol. Sci.* **268**, 31–38 (2001).
67. Shams, L., Kamitani, Y. & Shimojo, S. Visual illusion induced by sound. *Brain Res. Cogn. Brain Res.* **14**, 147–152 (2002).
68. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
69. Mishra, J., Martinez, A., Sejnowski, T. J. & Hillyard, S. A. Early Cross-Modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion. *J. Neurosci.* **27**, 4120–4131 (2007).
70. Grimes, D., Tan, D., Hudson, S., Shenoy, P. & Rao, R. Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph. *Microsoft Res*. (2008).
71. Mishra, J., Martinez, A. & Hillyard, S. A. Cortical Processes Underlying Sound-Induced Flash Fusion. *Brain Res.* **1242**, 102–115 (2008).
72. Keil, J., Müller, N., Hartmann, T. & Weisz, N. Prestimulus Beta Power and Phase Synchrony Influence the Sound-Induced Flash Illusion. *Cereb. Cortex* **24**, 1278–1288 (2014).
73. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. & Iverson, G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16**, 225–237 (2009).
74. Aczel, B., Palfi, B. & Szaszi, B. Estimating the evidential value of significant results in psychological science. *PLOS ONE* **12**, e0182651 (2017).
75. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* **6**, 65–70 (1979).

Acknowledgements

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through the research grant KE1828/4-1 to J.K.

Author Contributions

G.M. was involved in the design of the study, data acquisition and analysis and the writing of the manuscript. J.K. was involved in the design of the study, data analysis and the writing of the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Study 2

Michail G, Senkowski D, Niedeggen M, Keil J (2021) Memory Load Alters Perception-Related Neural Oscillations during Multisensory Integration. *The Journal of Neuroscience*, 41:1505–1515. doi: 10.1523/JNEUROSCI.1397-20.2020

Study 3

Michail G, Senkowski D, Holtkamp M, Wächter B, Keil J (*submitted*) Early beta oscillations in multisensory association areas underlie crossmodal performance enhancement. *The Journal of Neuroscience*.

Early beta oscillations in multisensory association areas underlie crossmodal performance enhancement

Georgios Michail, Daniel Senkowski, Martin Holtkamp, Bettina Wächter, Julian Keil

Abstract

The combination of signals from different sensory modalities can enhance perception and facilitate behavioral responses. While previous research described crossmodal influences in a wide range of tasks, it remains unclear how such influences drive performance enhancements. In particular, the neural mechanisms underlying performance-relevant crossmodal influences, as well as the latency and spatial profile of such influences are not well understood. Here, we examined data from high-density electroencephalography (N = 30) and electrocorticography (N = 4) recordings to characterize the oscillatory signatures of crossmodal facilitation of response speed, as manifested in the speeding of visual responses by concurrent task-irrelevant auditory information. Using a data-driven analysis approach, we found that individual gains in response speed correlated with reduced beta power (13-25 Hz) in the audiovisual compared with the visual condition, starting within 80 ms after stimulus onset in multisensory association and secondary visual areas. In addition, the electrocorticography data revealed a beta power suppression in audiovisual compared with visual trials in the superior temporal gyrus (STG). Our data suggest that the crossmodal facilitation of response speed is associated with early beta power in multisensory association and secondary visual areas, presumably reflecting the enhancement of early sensory processing through selective attention. This finding furthers our understanding of the neural correlates underlying crossmodal response speed facilitation and highlights the critical role of beta oscillations in mediating behaviorally relevant audiovisual processing.

Significance Statement

The use of complementary information across multiple senses can enhance perception. Previous research established a central role of neuronal oscillations in multisensory perception, but it remains poorly understood how they relate to multisensory performance enhancement. To address this question, we recorded electrophysiological signals from scalp and intracranial electrodes (implanted for presurgical monitoring) in response to simple visual and audiovisual stimuli. We then associated the difference in oscillatory power between the two conditions with the speeding of responses in the audiovisual trials. We demonstrate, that the crossmodal facilitation of response speed is associated with beta power in multisensory association areas during early stages of sensory processing. This finding highlights the importance of beta oscillations in mediating behaviorally relevant audiovisual processing.

Introduction

In everyday life, using complementary information from multiple sensory modalities is often critical to make rapid and accurate perceptual decisions. The synthesis of signals from different senses has been shown to improve perceptual performance, leading to more accurate (Spence and Driver, 2004; Lippert et al., 2007) and faster responses (Hershenson, 1962; Diederich and Colonius, 2004). Previous research has shown that crossmodal interactions are governed by neural oscillations in different frequency bands that can occur at both early and late stages of processing and involve bottom-up and top-down mechanisms (Keil and Senkowski, 2018; Bauer et al., 2020). Despite the considerable progress in characterizing the role of neural oscillations in multisensory processing, it remains unclear how they relate to the behavioral facilitation of responses to multisensory stimuli. In particular, the processing stage at which functionally relevant oscillations unfold during crossmodal behavior facilitation, and whether they reflect top-down or bottom-up influences on sensory processing, are key questions that are not well understood (Bizley et al., 2016).

In relation to the crossmodal facilitation of response times (RTs), electrophysiological studies in humans examining multisensory interactions in evoked brain potentials have suggested a link of RT facilitation with early crossmodal interactions (Giard and Peronnet, 1999; Fort et al., 2002; Molholm et al., 2004; Gondan et al., 2005). However, the proposed association in these studies is based on activity differences between multisensory and unisensory conditions that were not directly linked with the individual gains in multisensory performance enhancement. Thus far, only few studies have examined how neural oscillations relate to crossmodal RT facilitation across individuals (Senkowski et al., 2006; Mercier et al., 2015). In a speeded response paradigm, Senkowski et al. (2006) found a relationship between evoked beta oscillations and shorter RTs for unisensory and bisensory audiovisual stimuli. In an electrocorticography (ECoG) study, Mercier et al. (2015) observed that delta band (<4 Hz) phase alignment in a sensorimotor network was related to crossmodal facilitation of response speed. However, in both studies the modulations in neural oscillations were associated with shorter RTs after both multisensory and unisensory stimulation. Therefore, it cannot be concluded that these brain responses are specific for crossmodal facilitation of RTs. Moreover, the use of speeded responses in these studies, with a mean RT lower than 300 ms for audiovisual trials, indicates that the observed oscillatory activities may reflect motor-related processing. Taken together, while there is some evidence that neural oscillations play a role in crossmodal facilitation of response speed, the specificity of these effects to multisensory processing has not yet been demonstrated. Critically, it remains unclear whether the crossmodal facilitation of response speed is associated with modulations of neural oscillations during early stages of sensory processing.

In two experiments, we examined how individual gains in response speed during crossmodal stimulation relate to neural processing, as reflected in neural oscillations. We investigated oscillatory power in response to unisensory visual and bisensory audiovisual stimuli in experiments in which participants had to indicate the number of perceived flashes. Electrophysiological data were collected independently in healthy individuals (N = 30) using high-density EEG recordings and in patients with drug-resistant focal epilepsy (N = 4) prior to resective surgery, using ECoG recordings. The EEG data analysis revealed that lower early beta band power for audiovisual compared with visual trials in multisensory association and secondary visual regions correlated with crossmodal facilitation in response speed. The ECoG data analysis revealed lower beta power in audiovisual compared with visual trials in the superior temporal gyrus (STG). Our findings suggest that early beta band power in multisensory association cortex plays an important role in crossmodal facilitation of response speed.

Material and Methods

The electrophysiological data from high-density scalp EEG and intracranial ECoG recordings were obtained independently. Throughout the text, the recording sessions to obtain these data are referred to as ‘EEG experiment’ and ‘ECoG experiment’, respectively.

Participants

For the EEG experiment, forty participants (mean age ± standard deviation (SD): 26.6 ± 7.8 years; 19 females) with normal hearing, normal or corrected-to-normal vision and no history of neurological disorders were recruited. Six participants with excessive EEG artefacts (slow wave drifts and muscular artefacts) and four with insufficient trials (less than 30 trials in at least one of the analyzed conditions) were excluded from the analysis. Therefore, a subset of thirty participants (mean age ± SD: 25.5 ± 6.4 years; 17 females) was included in further EEG data analyses.

Four male patients (mean age ± SD: 27.3 ± 4.9 years) with drug-resistant focal epilepsy treated at the Epilepsy-Center Berlin-Brandenburg (Institute for Diagnostic of Epilepsy) in Berlin participated in the ECoG experiment. The patients were implanted with subdural electrodes (n = 66, 50, 40 and 74 for patients 1 to 4, respectively) covering mainly the temporal cortex for presurgical intracranial video-EEG monitoring.

All participants provided written informed consent. The experiments were conducted in accordance with the 2008 Declaration of Helsinki and approved by the ethics committee of the Charité–Universitätsmedizin Berlin (Approval number: EA1/207/15).

Experimental Design

Participants were presented with combinations of auditory and visual stimuli and had to indicate the number of perceived visual stimuli. Stimulus combinations consisted of 0, 1 or 2 auditory (a) stimuli combined with either 0, 1 or 2 visual (v) stimuli. Six stimulus combinations were used in the EEG experiment, (a_0v_1, a_0v_2, a_1v_1, a_2v_0, a_2v_1, a_2v_2), and nine in the ECoG experiment (a_0v_1, a_0v_2, a_1v_0, a_1v_1, a_1v_2, a_2v_0, a_2v_1, a_2v_2 and a_2v_{1late}). The current study focused on the analysis of the visual-only stimulus (a_0v_1, **V**) and the bisensory audiovisual (a_1v_1, **AV**) stimulus combination in which one visual stimulus is presented together with one auditory stimulus (**Figure 1A**). In the EEG experiment, prior to the audiovisual stimulation, participants performed an n-back task (0-back, 2-back). In the current study, we only analyzed the a_0v_1 and a_1v_1 trials and only from the 0-back condition. Further details of the experimental setup can be found in Michail et al. (2021), which analyzed the memory-load effects on the perception of the a_2v_1 trials from the same EEG dataset. The visual (flash) stimulus was a white disk subtending a visual angle of 1.6° and was presented at 4.1° centrally below the fixation cross, for 13.3 ms (EEG) or 16.7 ms (ECoG). The slight difference in visual presentation times is explained by the different refresh rates of the displays used for the EEG and ECoG experiments. The auditory (beep) stimulus was a 78 dB (SPL) 1000 Hz sine wave tone that was presented for 7 ms. In AV trials, auditory and visual stimuli were presented simultaneously.

Stimulus presentation and recording of participants' responses were implemented using the Psychophysics toolbox (Brainard, 1997; RRID:SCR_002881) for MATLAB (The Mathworks, Natick, MA, USA). The EEG experiment was conducted in a dimly lit, electrically shielded, noise-attenuating chamber. Visual stimuli were displayed on a 21-inch CRT screen at a distance of 1.2 m with a 75 Hz refresh rate. The ECoG experiment was conducted at the patient's bedside using a portable computer (HP Pavilion 17) with a 60 Hz screen refresh rate. Auditory stimuli in both experiments were controlled by a USB audio interface (UR22mkII, Steinberg) and delivered through in-ear headphones (ER30, Etymotic Research).

Each trial started with a central fixation cross displayed for a variable duration of 500 to 800 ms (EEG) or 1000 to 1500 ms (ECoG). Then, one of the stimulus combinations was presented. After the presentation of a stimulus, the fixation cross was displayed again and participants had to indicate the number of perceived flashes by a button press (three buttons: 0, 1 or 2). Following the button press or after 1500ms (if no button was pressed), a new trial started. In the EEG experiment, prior to the main task described above, participants performed a verbal visual n-back task (0- and 2-back, for details see Michail et al., 2021). In the current study, we only used trials from the 0-back condition, in which participants had to detect the target letter ‘X’, presented in 33% of all trials. The letter detection task was not related to the V and AV stimuli and should, thus, not have substantially affected the processing of these stimuli. Participants reported the number of flashes with the right thumb using a handheld gamepad (Logitech Gamepad F310, Logitech, Lausanne, Switzerland).

The EEG experiment included 12 blocks (6 blocks for each load level: 0-back and 2-back), each block consisting of 74 trials. The order of blocks was randomized across participants and the duration of experiment was approximately 80 minutes. The ECoG experiment, with a duration of 60 minutes, consisted of 6 blocks, each including 139 trials (due to fatigue, the first participant completed only 4 blocks).

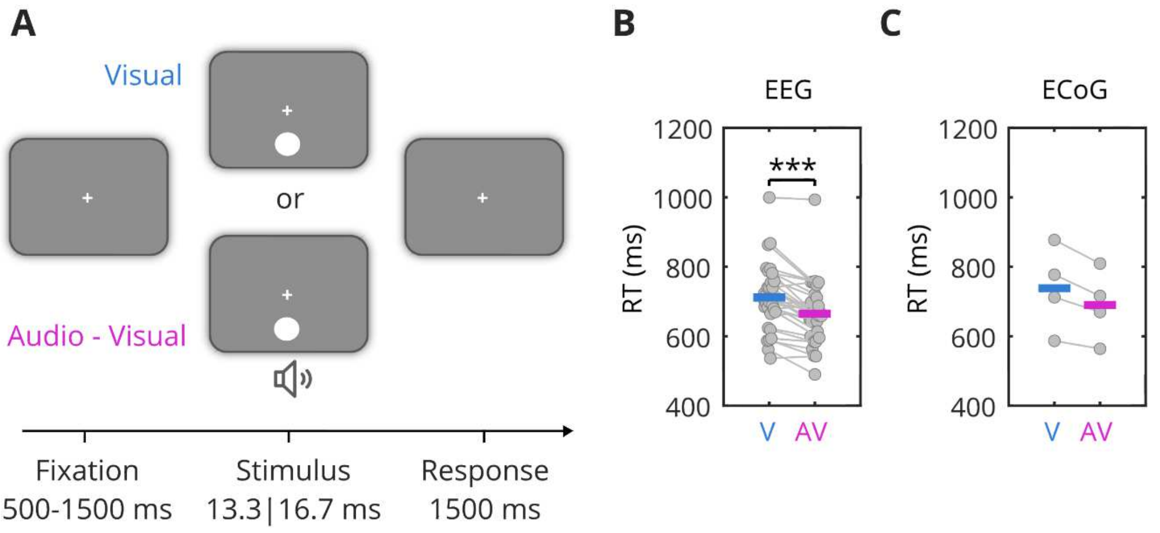

Figure 1. Experimental setup and behavioral results. A. Schematic illustration of the experimental conditions. Participants were presented with a unisensory visual (V) or a bisensory audio-visual stimulus

(AV) and were asked to indicate the number of perceived visual stimuli. **B.** Participants in the EEG experiment responded faster in the AV compared with the V condition. Horizontal bold lines denote the mean. **C.** In the ECoG experiment, participants showed a speeding of responses in the AV condition, similar to the EEG experiment. Within-subject response speed was faster for AV compared with V stimuli in 3 out of 4 participants (significant or trend to significant difference). *** $p < 0.001$

Behavioral data analysis

Behavioral performance was assessed in terms of the percentage of correct responses in the V and the AV condition and the RTs in trials with correct responses.

Data acquisition and preprocessing

High-density EEG was recorded using a 128-channel passive system (EasyCap, Herrsching, Germany) at a sampling rate of 2500 Hz. Two electrodes, at the right lateral canthi and below the right eye, recorded the horizontal and vertical electro-oculograms. Preprocessing was performed with MNE-Python (Gramfort et al., 2014; RRID:SCR_005972) and further data analysis with Fieldtrip (Oostenveld et al., 2011; RRID:SCR_004849) and custom-made Matlab scripts (MathWorks, Natick, MA).

Offline, EEG data were filtered with a zero-phase bandpass finite impulse response (FIR) filter between 1 Hz and 100 Hz using the window design method (“firwin” in SciPy [https://docs.scipy.org/doc/]; Hanning window; 1 Hz lower transition bandwidth; 25 Hz upper transition bandwidth; 3.3 s filter length). A band-stop notch FIR filter from 49 to 51 Hz (6.6 s filter length), was applied to remove line noise. In the next analysis step, data were downsampled to 256 Hz and epoched from -1.5 to 1.5 s relative to the onset of the stimuli. Trials with artefacts (eye blinks, noise, or muscle activity) were removed after visual inspection. Data were then re-referenced to the average of all electrodes and subjected to Independent Component Analysis (ICA) using the Extended-Infomax algorithm (Lee et al., 1999). Components representing eye blinks, cardiac and muscle activity were removed from the data. Next, noisy electrodes were rejected after visual inspection on a trial-by-trial basis and interpolated using spherical spline interpolation (Perrin et al., 1989). Finally, trials with signal exceeding ±150 µV were excluded. On average, across participants, 106.5 (SD 96) trials and 12.1 (SD 4.3) ICA components were removed, and 11.1 (SD 3.6) electrodes were interpolated.

ECoG signals were recorded at a 2048 Hz sampling rate using a 128-channel REFA system (TMSi International, Enschede, The Netherlands). Offline, ECoG data were filtered using a zero-phase bandpass finite impulse response (FIR) filter between 1 Hz and 200 Hz (high pass: order = 6765, -6 dB cutoff frequency = 0.5 Hz; low pass: order = 137, -6 dB cutoff frequency = 225 Hz). A band-stop notch filter was applied at 50 Hz (±1) and its harmonics to filter out line noise. Data were subsequently downsampled to 600 Hz and epoched from -1 to 2.5 s relative to the onset of the stimulus. Electrodes with epileptiform activity or excessive noise were excluded from the analysis. Moreover, trials with an amplitude larger than five times the SD for more than a period of 25 ms (Blenkmann et al., 2019) and trials with artefacts (large slow drifts or excessive noise) identified after visual inspection were removed. Data were then re-referenced to the common average. On average, across participants in the ECoG experiment, 11.4 % (SD 4.3) of the trials and 7.4 % (SD 5.8) of the electrodes were removed.

To determine the locations of the intracranial electrodes, the post-implantation CT was co-registered with the preoperative MRI following the pipeline implemented in FieldTrip for the integrated analysis of anatomical and ECoG data (Stolk et al., 2018).

Time-frequency analysis

Oscillatory power was computed by applying a Hanning taper to an adaptive time window of 4 cycles for each frequency from 2 to 40 Hz, shifted from -1.5 to 1.5 s (EEG) and from -1 to 2.5 s (ECoG), in steps of 10ms. Poststimulus power was baseline corrected using the average power of the prestimulus window from -500 to -100 ms, relative to stimulus onset.

EEG source analysis

Surface-level EEG data were projected into source space to investigate the cortical sources of the correlation between spectral power and RTs, obtained from the sensor level analysis. First, for each participant, the individual T1-weighted MRI (3T Magnetom TIM Trio, Siemens, AG, Germany) was co-registered with the individually digitized EEG electrode positions (Polhemus FastTrak) to a common coordinate system (Montreal Neurological Institute, MNI). This was done by utilizing the digitized headshape information and the fiducial locations (nasion, left and right preauricular points). The co-registered MRI image was then segmented using the SPM12 algorithm and a realistic three-shell (brain, skull, skin) boundary element volume conductor model (BEM) was constructed (Oostendorp and van Oosterom, 1989). Next, the template MNI brain was non-linearly warped onto each participant's anatomical data to obtain a three-dimensional source model (volumetric grid) with a resolution of 10 mm, which was used for the further analysis. To estimate the current density distribution the eLoreta algorithm (Pascual-Marqui, 2007) was used with a lambda regularization parameter set to 1%. To this end, the cross-spectral density (CSD) matrix was calculated using the Fast Fourier Transform (FFT) method for the condition-pooled data. As mentioned in the Introduction, in the current study, we were particularly interested on whether crossmodal RT facilitation is associated with early crossmodal influences. Accordingly, the source analysis focused on the early beta band component (80-200 ms, 13-25 Hz) of the significant cluster obtained from the scalp level correlation analysis. Therefore, CSD was calculated in the time window from 80 to 200 ms relative to stimulus onset**.** Center frequency and spectral smoothing were defined to fit the frequency range of interest; hence, a center frequency of 19 Hz and a smoothing of 6 Hz were used, resulting in a 13–25 Hz range. The current density estimate was normalized to the source estimate for the baseline window (-0.5 to -0.1 s) as follows: (*Poststimulus* - *Baseline*) / (*Poststimulus* + *Baseline*).

Statistical analysis

For the EEG experiment, paired-samples *t*-tests were used to compare behavioral performance, i.e., accuracy and RTs, between V and AV conditions. The corresponding within-subject comparisons in the ECoG experiment were performed using independent-samples *t*-tests.

To compare the EEG spectral power between V and AV conditions, a nonparametric cluster-based permutation test was conducted (cluster-forming alpha = 0.05, dependent t-test, iterations = 1000; Maris and Oostenveld, 2007). The test was applied in the time window from 0 to 500 ms relative to stimulus onset, on frequencies from 2 to 40 Hz. The observed test statistic was evaluated against the permutation distribution in order to test the null hypothesis of no difference between conditions (two-tailed test, alpha = 0.025).

A nonparametric cluster-based permutation test was also applied to assess the correlation between the AV minus V power difference at the sensor level and the RT difference between the two conditions (cluster-forming alpha = 0.05, Spearman’s rank correlation, iterations = 1000). Accordingly, a similar approach was used for the corresponding correlation analysis of the source space data (one-sided cluster-based permutation test, cluster-forming alpha = 0.1, Spearman's rank correlation, iterations = 1000). As mentioned before, the source analysis aimed to further investigate the findings of the sensor level analysis. Therefore, the direction of the one-tailed test was determined by the sensor level results.

With regard to the analysis of the ECoG data, the difference in beta power (averaged across the 13-25 Hz range) between V and AV conditions was assessed for each electrode in the time window from 0 to 500 ms using a nonparametric cluster-based permutation test (cluster-forming alpha = 0.05, independent samples *t*-test, iterations = 1000). Given that the non-symmetric arrangement of grid and strip electrodes prevents the use of spatial clustering algorithms, a more restricted alpha threshold of p = 0.01 was applied.

Results

Behavior

Behavior was assessed in terms of how fast and how accurate participants responded to V and AV stimuli. As depicted in **Figure 1B**, participants in the EEG experiment responded faster in the AV compared with the V condition (mean ± SD: 665 ± 92 ms vs. 712 ± 97 ms; paired samples *t*-test, $t_{(29)} = 6.7$, $p < 0.001$). Similarly, within-subject comparisons for participants in the ECoG experiment revealed significantly faster or a trend for faster responses in AV compared with the V condition in 3 out of 4 participants (**Figure 1C;** independent samples *t*-test, participant #1: 670 ± 97 ms vs. 712 ± 98 ms, $t_{(91)} = -2.1$, $p = 0.038$; #2: 716 ± 157 ms vs. 777 ± 193 ms, $t_{(150)} = -2.2$, $p = 0.033$; #3: 809 ± 182 ms vs. 877 ± 197 ms, $t_{(131)} = -2.1$, $p = 0.041$; #4: 564 ± 88 ms vs. 587 ± 87 ms, $t_{(150)} = -1.6$, $p = 0.11$). Only participant #4 revealed similar performance between conditions (p = 0.11). As this participant responded much faster than the other three participants, it is possible that the absence of a RT facilitation is due to a ceiling effect in performance. In the EEG experiment, while responses were more accurate in AV than V trials (98.1 ± 2.3 % vs. 92.0 ± 8.9 %, t(29) = -3.9, p < 0.001), participants showed in general high accuracy (>90%), suggesting that the task was easy to perform. Similarly, responses in the ECoG experiment were also highly accurate (V: 92.7 ± 5 %, AV: 93 ± 6.9 %; individual accuracies: participant #1: V=90.4%, AV= 88.5%; #2: V=96.2%, AV= 100%; #3: V=86.8%, AV= 85.9%; #4: V=97.4%, AV= 97.4%). Taken together, behavioral data from both EEG and ECoG experiments revealed that participants responded faster when the visual stimulus was combined with a task-irrelevant auditory stimulus than when the visual stimulus was presented alone.

Audiovisual stimulation induces increased EEG theta power

In the first step we analyzed the difference in EEG oscillatory power between the AV and V condition in the window from 0 to 500 ms, on frequencies from 2 to 40 Hz, using only correct trials. As illustrated in **Figure 2**, the nonparametric cluster-based permutation test revealed stronger theta power increase in the AV compared with the V condition, over medio-frontal and occipital electrodes in the time window from 0 to 400 ms relative to stimulus onset (*p* = 0.003).

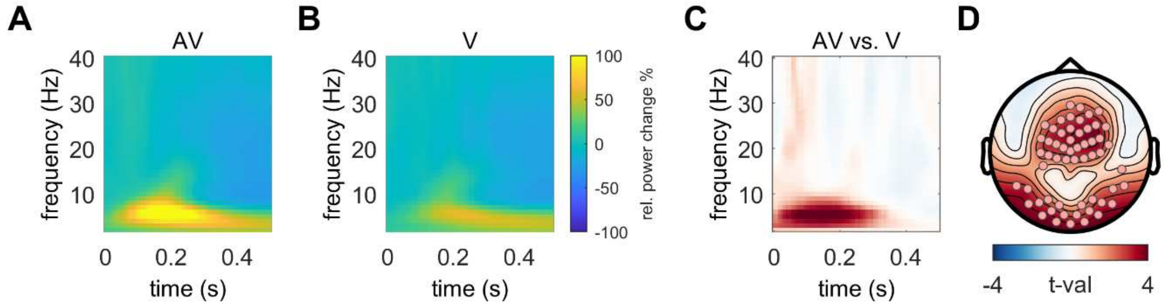

Figure 2. Oscillatory power difference between AV and V trials. The cluster-based analysis of EEG oscillatory power revealed higher theta power in AV compared with V trials in medio-frontal and occipital electrodes. **A-B.** TFRs of oscillatory power modulation after AV and V stimulation, averaged across

electrodes with high contribution to the cluster (i.e., with a total number of significant time-frequency samples at or above the mean). **C.** TFR of AV-V power difference (in t-values), averaged across electrodes with high contribution to the cluster and masked based on the temporal and spectral extent of the cluster. Higher values indicate stronger power for AV compared with the V condition. The color scale refers only to unmasked t values. **D.** Topographic map showing the spatial distribution of the difference in the cluster's time-frequency window. Electrodes with high contribution to the cluster are highlighted with dots.

Early beta power in association areas correlates with crossmodal facilitation of response speed

We next examined whether differences in EEG oscillatory power between the AV and the V condition correlated with the crossmodal facilitation of RTs. (**Figure 1B**). For this analysis only correct trials were used. A nonparametric cluster-based permutation test revealed one significant cluster (p = 0.001) showing a negative correlation between the RT difference (Δ RT $_{V\text{-}AV}$) and the power difference (Δ Power $_{AV\text{-}V}$) over mainly parieto-occipital and frontal scalp regions (**Figure 3A-B**). The cluster comprised two components, one in the early beta band activity (strongest effect at 80-200 ms, 13-25 Hz) and a second one in the late alpha band activity (strongest effect at 250-400 ms, 8-12Hz). To confirm the finding of the cluster-based analysis, a Spearman's rank correlation was performed between the RT facilitation (V minus AV) and the AV minus V power difference in the cluster (**Figure 3C;** rho = -0.81, p < 0.001). A comparison of the power in the cluster between the V and AV conditions revealed no significant difference between the two conditions. As mentioned in the *Introduction,* a central aim of the current study was to identify potential crossmodal effects at early processing stages. Therefore, the corresponding correlation analysis for the source activity focused on the early beta band activity (80-200 ms, 13-25 Hz). This analysis revealed a significant negative correlation of the AV minus V beta power difference in areas of the right inferior parietal and extrastriate occipital cortex with the crossmodal RT facilitation (nonparametric cluster-based permutation test, *p* = 0.001).

Taken together, our analysis revealed that the lower the early, parieto-occipital beta power in the AV compared with the V condition the faster participants responded in the AV vs. V condition. Moreover, the source localization of this correlation suggests the involvement of multisensory association areas and secondary visual cortex during the crossmodal RT facilitation.

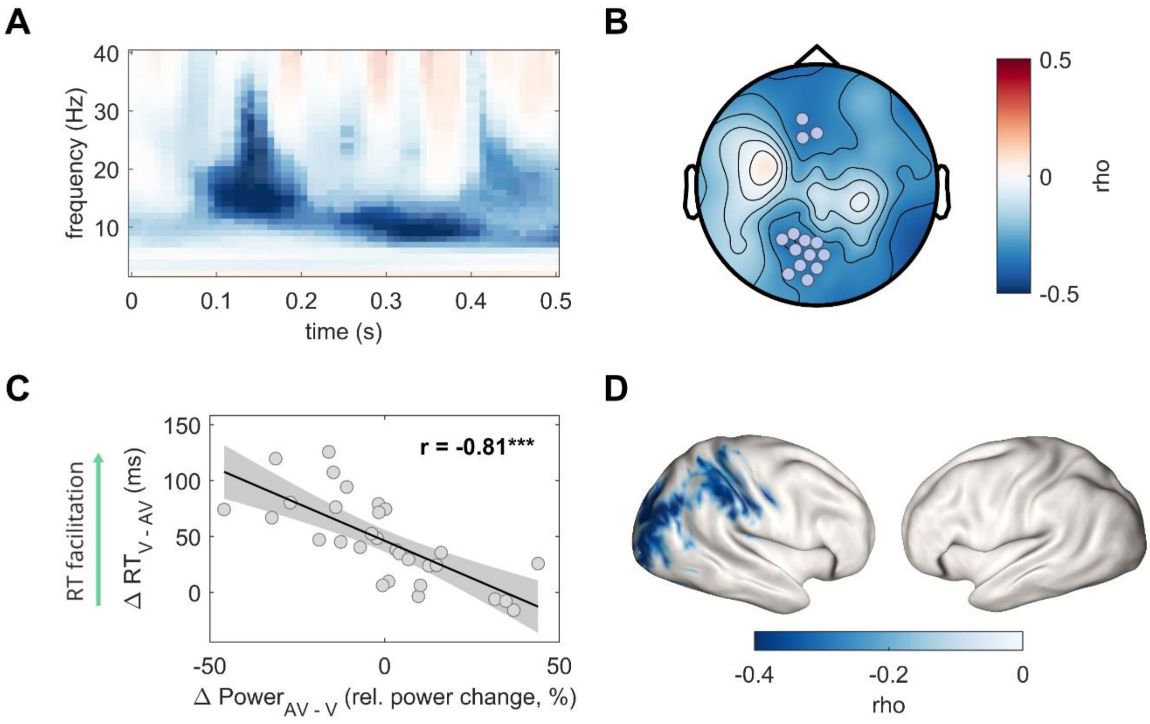

Figure 3. Correlation between AV minus V power difference and the crossmodal RT facilitation. The cluster-based correlation analysis revealed that crossmodal RT facilitation was associated with reduced beta power at 80-200 ms and reduced alpha power at 250-400 ms in mainly parieto-occipital electrodes, with the earlier beta effect being localized in inferior parietal and extrastriate occipital areas. **A.** TFR of the correlation (in rho values) between the AV minus V power difference and the V minus AV RT difference, averaged across electrodes with the highest contribution to the cluster (i.e., with a total number of significant time-frequency samples at or above the 75th percentile) and masked based on the temporal and spectral extent of the cluster. Lower values (blue) indicate that crossmodal RT facilitation correlates with smaller AV minus V power difference. The color scale refers only to unmasked rho values.

B. Topographic map showing the distribution of the correlation between AV minus V power difference and the crossmodal RT facilitation. Electrodes with the highest contribution to the cluster are highlighted with dots. **C.** Scatterplot depicting the correlation between the individual power difference (A minus V) in the cluster and the crossmodal RT facilitation (i.e., V minus AV). The lower the power in the cluster for the AV compared with the V condition the larger the crossmodal RT facilitation. Black lines represent the best-fitting linear regression and shaded areas the 95% confidence interval. **D.** Correlation in source space between the early beta band power difference (AV minus V, 80-200 ms, 13-25 Hz) and the crossmodal RT facilitation. Lower AV vs. V beta power in inferior parietal and extrastriate visual areas correlated with the crossmodal RT facilitation.; *** $p < 0.001$

ECoG beta power in the superior temporal gyrus is lower in audiovisual compared with visual-only trials

To further examine the role of beta power during crossmodal processing, we compared beta power modulations between the unisensory V and bisensory AV conditions in four participants implanted with intracranial electrodes covering mainly the temporal cortex (**Figure 4**). As reported above, these participants displayed shorter RTs for the AV compared with the V condition (**Figure 1C**). Our primary interest in this study, was to investigate early crossmodal influences on neural oscillations. Therefore, based on the outcome of the EEG data analysis – which linked early beta power modulations with crossmodal RT facilitation – the ECoG data analysis focused on the time course of the beta band power (13-25 Hz). This analysis revealed that, consistently for all four participants, beta power in the superior temporal gyrus (STG) starting at approximately 130 to 150 ms poststimulus was significantly lower in the AV compared with the V condition (nonparametric cluster-based permutation test, $p < 0.01$). Interestingly, the reverse pattern was observed for very early beta power (< 100 ms) in few electrodes in participant #1 (STG) and participant #4 (rolandic operculum, middle frontal gyrus). In these electrodes, beta power in the first 100 ms after stimulus onset was significantly higher in AV compared with V trials. **Table 1** provides an overview of the statistical results and the MNI coordinates of the electrodes at which significant effects were observed. These results provide further evidence that beta band power modulations in multisensory association areas, and especially in STG, reflect early crossmodal influences that might play a critical role in crossmodal RT facilitation.

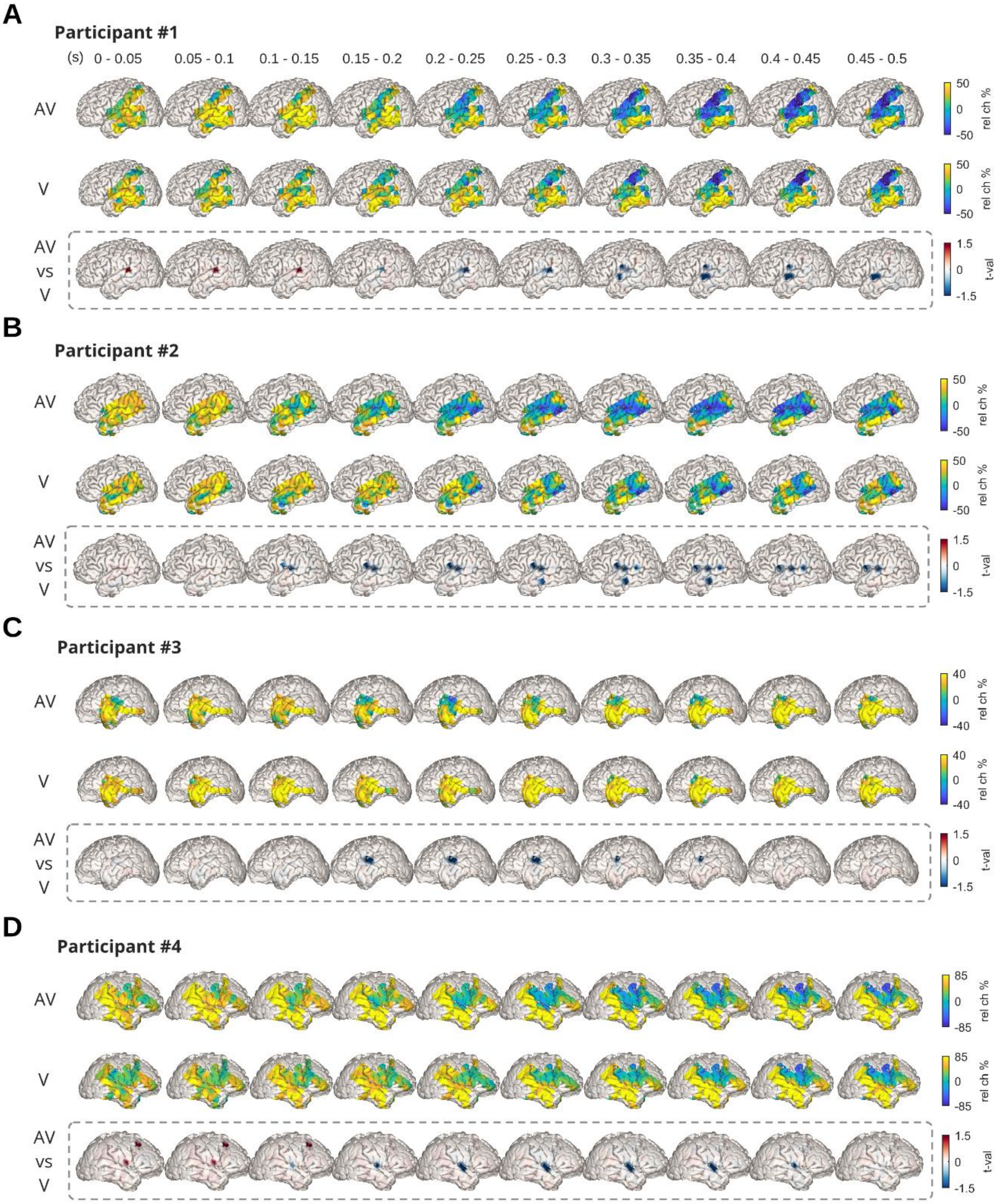

Figure 4. Intracranial (ECoG) beta band power in response to AV and V stimuli. The comparison of ECoG beta band power between the AV and the V condition showed that, consistent across participants,

beta power in STG starting at approximately 150 ms after stimulus onset was significantly lower in AV compared with the V condition. **A-D.** For each participant, the first two rows display the beta band (13-25 Hz) power modulation after AV and V stimulation in the time window from 0 to 500 ms after stimulus onset. The third row (highlighted with a dotted line) shows the beta band power difference (in *t*-values) between AV and V conditions. Larger values (red) indicate stronger power for AV compared with the V condition.

Table 1. Comparison of beta power between V and AV trials in ECoG experiment

Elec	**MNI Coordinates**			**Region (AAL atlas)**	**Statistical results**			
					AV > V		*AV < V*	
					interval (s)	$p_{cluster}$	*interval (s)*	$p_{cluster}$
Participant #1								
A28	-66.4698	-32.2866	14.2284	Superior temporal gyrus, L	0.01 - 0.12	.003		
							0.19 - 0.28	.005
A29	-66.1779	-21.6054	13.3358	Superior temporal gyrus, L			0.24 - 0.34	.006
B16	-63.8651	-12.4889	19.5746	Postcentral gyrus, L			0.31 - 0.44	.002
A23	-63.7925	-3.8424	-0.7619	Superior temporal gyrus, L			0.34 - 0.49	.002
A22	-66.4068	-13.4439	0.1332	Middle temporal gyrus, L			0.38 - 0.50	.007
Participant #2								
A21	-68.0743	-18.6511	12.6049	Superior temporal gyrus, L			0.13 - 0.50	.001
A30	-64.8222	-2.9299	16.5722	Postcentral gyrus, L			0.15 - 0.34	.001
A06	-69.7984	-19.5751	-14.4308	Middle temporal gyrus, L			0.30 - 0.39	.005
A11	-68.2180	-43.3035	14.4621	Superior temporal gyrus, L			0.35 - 0.43	.007
A31	-60.2779	6.0567	10.7324	Inferior frontal gyrus, opercular, L			0.39 - 0.49	.004
Participant #3								
A06	-68.2460	-9.0777	-4.1051	Middle temporal gyrus, L			0.16 - 0.29	.001
A11	-64.9019	-0.7296	1.0793	Superior temporal gyrus, L			0-17 - 0.37	.002
Participant #4								
A24	64.2172	4.7504	8.4491	Rolandic operculum, R	0.00 - 0.09	.007		
E03	44.2111	29.7326	46.4387	Middle frontal gyrus, R	0.02 - 0.12	.008		
A23	66.3881	-3.1604	2.5210	Superior temporal gyrus, R			0.15 - 0.41	.001
A27	62.8264	3.1071	-7.3572	Superior temporal pole, R			0.22 - 0.37	.001

Discussion

In this study, we analyzed EEG and ECoG data to elucidate the neural correlates of the crossmodal RT facilitation, as manifested in the speeding of behavioral responses to visual stimuli by the addition of task-irrelevant auditory information. We showed that reduced beta power in the AV compared with V trials correlated with individual crossmodal RT gains. This effect occurred around 80-200 ms poststimulus in parieto-occipital electrodes and was localized in secondary visual and multisensory association areas. Moreover, the ECoG data analysis showed that beta power in the STG, which is a key multisensory association area, is reduced in AV compared with V trials, starting approximately 150 ms after stimulus onset. These findings provide evidence that beta band power modulations in multisensory association and secondary visual cortex during early visual sensory processing reflect the crossmodal facilitation of response speed.

Despite evidence of crossmodal influences occurring during both early and late multisensory processing and in both primary sensory and higher-order cortical areas (Macaluso and Driver, 2005; Koelewijn et al., 2010; Talsma et al., 2010; Keil and Senkowski, 2018), it is not well understood how such interactions enable the multisensory facilitation of performance. A central question regards the processing stage and the level of cortical hierarchy at which information from one modality influences another modality, in particular when such multisensory influences facilitate performance (Bizley et al., 2016).

Our finding that crossmodal RT facilitation was linked with oscillatory power modulations at 80-200 ms poststimulus suggests that the auditory signal influenced early visual sensory processing to enhance performance. This result is consistent with a large body of primate and human electrophysiological studies demonstrating multisensory interactions at early processing stages (Giard and Peronnet, 1999; Molholm et al., 2002; Schroeder and Foxe, 2005; Talsma and Woldorff, 2005; Lakatos et al., 2007; Kayser et al., 2010; Mercier et al., 2013). Moreover, our finding is in line with EEG studies providing direct evidence of early crossmodal responses underlying multisensory behavior facilitation in tasks using simple audiovisual stimuli (Thorne et al., 2011; Van der Burg et al., 2011; Starke et al., 2020). On the contrary, other studies using more complex stimuli have shown that sound-induced improvements of visual motion and visual object categorization were associated with late single-trial EEG activity starting at 300 ms (Kayser et al., 2017; Franzen et al., 2020). This divergence in the latency of performance-relevant crossmodal influences is consistent with evidence of multisensory integration taking place during both sensory encoding and decision formation (Mercier and Cappe, 2020) and is likely attributed to stimulus complexity, in accordance with the adaptive engagement of integrative mechanisms depending on task-specific characteristics (van Atteveldt et al., 2014; Bizley et al., 2016). In this framework, our data argue that under conditions of low stimulus complexity, multisensory RT facilitation is linked with crossmodal influences at early processing stages.

Critically, the crossmodal RT facilitation in our study was associated with power modulations in the beta band (13-25 Hz). The correlation between crossmodal beta power modulation and RT facilitation was observed in parieto-occipital electrodes and was localized in inferior parietal and extrastriate occipital regions. We propose that the performance-relevant beta power suppression in the audiovisual compared with the visual condition reflects the enhancement of early visual processing through top-down attentional control originating from multisensory association and secondary visual cortex. This proposal is consistent with growing evidence on the role of beta oscillations in conveying feedback influences on low-level visual areas (Buschman and Miller, 2007;

Kerkoerle et al., 2014; Bastos et al., 2015; Michalareas et al., 2016; Richter et al., 2017; Limanowski et al., 2020). Moreover, evidence of feedback influences in the alpha-beta band modulating feedforward gamma band processing (Spaak et al., 2012; Richter et al., 2017) suggests that feedback signals in the low-frequency range (i.e., in the alpha-beta range), originating from association areas can directly modulate the feedforward stream of information during early sensory processing (Bressler and Richter, 2015). Our proposal is further supported by research showing that the suppression of low-frequency activity is associated with more efficient sensory processing of task-relevant signals (Klimesch et al., 2007; Jensen and Mazaheri, 2010), possibly by enhancing the feedforward communication through gamma band coherence (Hahn et al., 2019). In multisensory settings, previous studies provided evidence implicating beta power in the audiovisual redundant target effect (Senkowski et al., 2006), the integration of incongruent or noisy audiovisual speech stimuli (Schepers et al., 2013; Roa Romero et al., 2015), crossmodal influence on pain (Senkowski et al., 2011; Mancini et al., 2013), and the impact of working memory load on audiovisual illusory perception (Michail et al., 2021). Moreover, previous research demonstrated the involvement of beta band functional connectivity between primary and higher-order association areas in multisensory perception (Kayser and Logothetis, 2009; Hipp et al., 2011; Keil et al., 2014). Interestingly, a crossmodal (AV minus V) theta power enhancement over medio-frontal and occipital regions was not related to performance enhancement, suggesting that crossmodal theta power modulations might not be directly relevant for behavior. In this context, we argue that the functionally relevant beta band suppression in secondary visual and multisensory association areas – driven by the task-irrelevant auditory stimulus – enhanced early sensory representations of the visual stimulus through top-down attentional control of feedforward information processing.

Additionally, the analysis of the ECoG data revealed that beta power was reduced in the STG in the AV compared with the V condition. Previous work has established the critical role of the STG in multisensory perception, acting as a convergence hub for inputs from multiple modalities (Calvert et al., 2000; Beauchamp et al., 2004; Barraclough et al., 2005; Balz et al., 2016; Ozker et al., 2017; Karas et al., 2019; Mégevand et al., 2020). Moreover, previous studies using illusory audiovisual paradigms demonstrated that beta band power suppression was associated with audiovisual mismatch evaluation and top-down influences on audiovisual integration, induced by working memory load (Roa Romero et al., 2015; Michail et al., 2021). In accordance with these studies, the beta band suppression in the STG might reflect an auditory-driven feedback signal to improve visual processing through selective attention. This notion is consistent with the temporal and spatial profile of the observed tight relationship between beta oscillations in the EEG data and the crossmodal RT facilitation. It also in line with neuroimaging and electrophysiological evidence showing anatomical and functional connections in the beta band between STG and primary sensory areas (Noesselt et al., 2007, 2010; Cappe et al., 2009; Kayser and Logothetis, 2009; Keil et al., 2014). Therefore, this finding, together with the sources of the correlation between EEG beta power and RT facilitation, suggest an important role of multisensory association areas during behaviorally relevant early crossmodal processing.

Thus far, only few studies have investigated the oscillatory signatures of crossmodal RT facilitation using similar audiovisual stimuli as the current study (Senkowski et al., 2006; Mercier et al., 2015). Contrary to present findings, one previous study found that the audiovisual RT facilitation was associated with increased evoked beta power in left frontal and right occipital electrodes (Senkowski et al., 2006). This inconsistent finding might be explained by differences in the task instructions. In the current study participants had to report on features of the visual stimulus, whereas in Senkowski et al. (2006) participants made speeded responses upon stimulus detection independent of modality. This resulted in markedly faster RTs in Senkowski et al. (2006) compared with the current study (mean RTs to AV trials: 255 ms vs. 665 ms, respectively). Thus, the beta modulations in that previous study were possibly related to motor processes, whereas, in the present study, there is an additional perceptual aspect. Using a similar speeded detection task, an ECoG study has linked crossmodal RT facilitation with local phase alignment and phase synchronization between auditory and motor cortex in the delta band (Mercier et al., 2015). The use of a speeded detection task in these studies makes it difficult to disentangle the oscillatory activities associated with audiovisual interactions in sensory and non-sensory stages of information processing. Further investigations are required to differentiate the contributions of beta power and functional connectivity at the level of sensory processing, decision-making and motor response.

One limitation of our study is the small sample size in the ECoG experiment, which prevented us from performing similar analyses as in the EEG experiment. In addition to that, the heterogeneity between participants in the cortical grid coverage, further constrained the ability to perform analyses across participants to obtain statistically robust results at the group level. Thus, future ECoG studies, recruiting larger participant cohorts and possibly with a more diverse cortical grid coverage could provide insights into the role of other regions in crossmodal performance enhancement.

Altogether, our data suggest that beta power in multisensory association areas is related to the crossmodal facilitation of response speed. This beta power modulation presumably reflects the earliest stage of behaviorally relevant audiovisual feedback processing in higher multisensory areas, starting around 80 ms after stimulus presentation. Thus, the present findings highlight the important role of beta oscillations in mediating behaviorally relevant crossmodal influences between the auditory and visual modalities.

References

Balz J, Keil J, Roa Romero Y, Mekle R, Schubert F, Aydin S, Ittermann B, Gallinat J, Senkowski D (2016) GABA concentration in superior temporal sulcus predicts gamma power and perception in the sound-induced flash illusion. NeuroImage 125:724–730.

Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of Visual and Auditory Information by Superior Temporal Sulcus Neurons Responsive to the Sight of Actions. Journal of Cognitive Neuroscience 17:377–391.

Bastos AM, Vezoli J, Bosman CA, Schoffelen JM, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. Neuron 85:390–401.

Bauer AKR, Debener S, Nobre AC (2020) Synchronisation of Neural Oscillations and Cross-modal Influences. Trends Cogn Sci 24:481–495.

Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of Auditory and Visual Information about Objects in Superior Temporal Sulcus. Neuron 41:809–823.

Bizley JK, Jones GP, Town SM (2016) Where are multisensory signals combined for perceptual decision-making? Current Opinion in Neurobiology 40:31–37.

Blenkmann AO, Collavini S, Lubell J, Llorens A, Funderud I, Ivanovic J, Larsson PG, Meling TR, Bekinschtein T, Kochen S, Endestad T, Knight RT, Solbakk AK (2019)

Auditory deviance detection in the human insula: An intracranial EEG study. Cortex 121:189–200.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Bressler SL, Richter CG (2015) Interareal oscillatory synchronization in top-down neocortical processing. Current Opinion in Neurobiology 31:62–66.

Buschman TJ, Miller EK (2007) Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. Science 315:1860–1862.

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Current Biology 10:649–657.

Cappe C, Rouiller EM, Barone P (2009) Multisensory anatomical pathways. Hearing Research 258:28–36.

Diederich A, Colonius H (2004) Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. Perception & Psychophysics 66:1388–1404.

Fort A, Delpuech C, Pernier J, Giard MH (2002) Dynamics of Cortico-subcortical Cross-modal Operations Involved in Audio-visual Object Detection in Humans. Cerebral Cortex 12:1031–1039.

Franzen L, Delis I, De Sousa G, Kayser C, Philiastides MG (2020) Auditory information enhances post-sensory visual evidence during rapid multisensory decision-making. Nature Communications 11:5440.

Giard MH, Peronnet F (1999) Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. Journal of Cognitive Neuroscience 11:473–490.

Gondan M, Niederhaus B, Rösler F, Röder B (2005) Multisensory processing in the redundant-target effect: A behavioral and event-related potential study. Perception & Psychophysics 67:713–726.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS (2014) MNE software for processing MEG and EEG data. NeuroImage 86:446–460.

Hahn G, Ponce-Alvarez A, Deco G, Aertsen A, Kumar A (2019) Portraits of communication in neuronal networks. Nature Reviews Neuroscience 20:117–127.

Hershenson M (1962) Reaction time as a measure of intersensory facilitation. Journal of Experimental Psychology 63:289–293.

Hipp JF, Engel AK, Siegel M (2011) Oscillatory Synchronization in Large-Scale Cortical Networks Predicts Perception. Neuron 69:387–396.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci 4:186.

Karas PJ, Magnotti JF, Metzger BA, Zhu LL, Smith KB, Yoshor D, Beauchamp MS (2019) The visual speech head start improves perception and reduces superior temporal cortex responses to auditory speech Pasternak T, King AJ, Mahon B, eds. eLife 8:e48116.

Kayser C, Logothetis NK (2009) Directed Interactions Between Auditory and Superior Temporal Cortices and their Role in Sensory Integration. Front Integr Neurosci 3:7.

Kayser C, Logothetis NK, Panzeri S (2010) Visual Enhancement of the Information Representation in Auditory Cortex. Current Biology 20:19–24.

Kayser SJ, Philiastides MG, Kayser C (2017) Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. NeuroImage 148:31–41.

Keil J, Müller N, Hartmann T, Weisz N (2014) Prestimulus Beta Power and Phase Synchrony Influence the Sound-Induced Flash Illusion. Cereb Cortex 24:1278–1288.

Keil J, Senkowski D (2018) Neural Oscillations Orchestrate Multisensory Processing. The Neuroscientist 24:609–626.

Kerkoerle T van, Self MW, Dagnino B, Gariel-Mathis MA, Poort J, Togt C van der, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. PNAS 111:14332–14341.

Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res Rev 53:63–88.

Koelewijn T, Bronkhorst A, Theeuwes J (2010) Attention and the multiple stages of multisensory integration: A review of audiovisual studies. Acta Psychologica 134:372–384.

Lakatos P, Chen C-M, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. Neuron 53:279–292.

Lee TW, Girolami M, Sejnowski TJ (1999) Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources. Neural Computation 11:417–441.

Limanowski J, Litvak V, Friston K (2020) Cortical beta oscillations reflect the contextual gating of visual action feedback. NeuroImage 222:117267.

Lippert M, Logothetis NK, Kayser C (2007) Improvement of visual contrast detection by a simultaneous sound. Brain Research 1173:102–109.

Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. Trends in Neurosciences 28:264–271.

Mancini F, Longo MR, Canzoneri E, Vallar G, Haggard P (2013) Changes in cortical oscillations linked to multisensory modulation of nociception. European Journal of Neuroscience 37:768–776.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods 164:177–190.

Mégevand P, Mercier MR, Groppe DM, Zion Golumbic E, Mesgarani N, Beauchamp MS, Schroeder CE, Mehta AD (2020) Crossmodal Phase Reset and Evoked Responses Provide Complementary Mechanisms for the Influence of Visual Speech in Auditory Cortex. J Neurosci 40:8530–8542.

Mercier MR, Cappe C (2020) The interplay between multisensory integration and perceptual decision making. NeuroImage 222:116970.

Mercier MR, Foxe JJ, Fiebelkorn IC, Butler JS, Schwartz TH, Molholm S (2013) Auditory-driven phase reset in visual cortex: Human electrocorticography reveals mechanisms of early multisensory integration. Neuroimage 79:19–29.

Mercier MR, Molholm S, Fiebelkorn IC, Butler JS, Schwartz TH, Foxe JJ (2015) Neuro-Oscillatory Phase Alignment Drives Speeded Multisensory Response Times: An Electro-Corticographic Investigation. The Journal of Neuroscience 35:8546–8557.

Michail G, Senkowski D, Niedeggen M, Keil J (2021) Memory Load Alters Perception-Related Neural Oscillations during Multisensory Integration. J Neurosci 41:1505–1515.

Michalareas G, Vezoli J, van Pelt S, Schoffelen JM, Kennedy H, Fries P (2016) Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. Neuron 89:384–397.

Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory Visual–Auditory Object Recognition in Humans: a High-density Electrical Mapping Study. Cerebral Cortex 14:452–465.

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. Cognitive Brain Research 14:115–128.

Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual Temporal Correspondence Modulates Human Multisensory

Superior Temporal Sulcus Plus Primary Sensory Cortices. J Neurosci 27:11431–11441.

Noesselt T, Tyll S, Boehler CN, Budinger E, Heinze H-J, Driver J (2010) Sound-Induced Enhancement of Low-Intensity Vision: Multisensory Influences on Human Sensory-Specific Cortices and Thalamic Bodies Relate to Perceptual Enhancement of Visual Detection Sensitivity. J Neurosci 30:13609–13623.

Oostendorp TF, van Oosterom A (1989) Source parameter estimation in inhomogeneous volume conductors of arbitrary shape. IEEE Transactions on Biomedical Engineering 36:382–391.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.

Ozker M, Schepers IM, Magnotti JF, Yoshor D, Beauchamp MS (2017) A Double Dissociation between Anterior and Posterior Superior Temporal Gyrus for Processing Audiovisual Speech Demonstrated by Electrocorticography. Journal of Cognitive Neuroscience 29:1044–1060.

Pascual-Marqui RD (2007) Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. arXiv:07103341 [math-ph, physics:physics, q-bio] Available at: http://arxiv.org/abs/0710.3341.

Perrin F, Pernier J, Bertrand O, Echallier JF (1989) Spherical splines for scalp potential and current density mapping. Electroencephalography and Clinical Neurophysiology 72:184–187.

Richter CG, Thompson WH, Bosman CA, Fries P (2017) Top-Down Beta Enhances Bottom-Up Gamma. J Neurosci 37:6698–6711.

Roa Romero Y, Senkowski D, Keil J (2015) Early and late beta-band power reflect audiovisual perception in the McGurk illusion. Journal of Neurophysiology 113:2342–2350.

Schepers IM, Schneider TR, Hipp JF, Engel AK, Senkowski D (2013) Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing. NeuroImage 70:101–112.

Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, ‘unisensory’ processing. Current Opinion in Neurobiology 15:454–458.

Senkowski D, Kautz J, Hauck M, Zimmermann R, Engel AK (2011) Emotional Facial Expressions Modulate Pain-Induced Beta and Gamma Oscillations in Sensorimotor Cortex. J Neurosci 31:14542–14550.

Senkowski D, Molholm S, Gomez-Ramirez M, Foxe JJ (2006) Oscillatory Beta Activity Predicts Response Speed during a Multisensory Audiovisual Reaction Time Task: A High-Density Electrical Mapping Study. Cereb Cortex 16:1556–1565.

Spaak E, Bonnefond M, Maier A, Leopold DA, Jensen O (2012) Layer-Specific Entrainment of Gamma-Band Neural Activity by the Alpha Rhythm in Monkey Visual Cortex. Current Biology 22:2313–2318.

Spence C, Driver J (2004) Crossmodal Space and Crossmodal Attention. Oxford University Press.

Starke J, Ball F, Heinze HJ, Noesselt T (2020) The spatio-temporal profile of multisensory integration. European Journal of Neuroscience 51:1210–1223.

Stolk A, Griffin S, Meij R van der, Dewar C, Saez I, Lin JJ, Piantoni G, Schoffelen J-M, Knight RT, Oostenveld R (2018) Integrated analysis of anatomical and electrophysiological human intracranial data. Nat Protoc 13:1699–1723.

Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. Trends in Cognitive Sciences 14:400–410.

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.

Thorne JD, Vos MD, Viola FC, Debener S (2011) Cross-Modal Phase Reset Predicts Auditory Task Performance in Humans. J Neurosci 31:3853–3861.

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

van Atteveldt N, Murray MM, Thut G, Schroeder CE (2014) Multisensory Integration: Flexible Use of General Operations. Neuron 81:1240–1253.

Appendix B: Erklärung über den Eigenanteil

Erklärung gemäß § 7 Abs. 3 Satz 4 der Promotionsordnung über den Eigenanteil an den veröffentlichten oder zur Veröffentlichung vorgesehenen eingereichten wissenschaftlichen Schriften im Rahmen meiner publikationsbasierten Arbeit

I.

Name, Vorname:	Michail, Georgios
Institut:	Fachbereich Erziehungswissenschaft und Psychologie der Freien Universität Berlin
Promotionsfach:	Psychologie
Titel:	M.Sc.

II. Nummerierte Aufstellung der eingereichten Schriften:

1. Michail G, Keil J (2018) High cognitive load enhances the susceptibility to non-speech audiovisual illusions. *Scientific Reports* 8:11530.

2. Michail G, Senkowski D, Niedeggen M, Keil J (2021) Memory Load Alters Perception-Related Neural Oscillations during Multisensory Integration. *The Journal of Neuroscience*, 41:1505–1515.

3. Michail G, Senkowski D, Holtkamp M, Wächter B, Keil J (*submitted*) Early beta band power in multisensory association areas reflects the crossmodal facilitation of response speed. *The Journal of Neuroscience*.

III. Darlegung des eigenen Anteils der Schriften:

zu II. 1.: Konzeption (in Teilen), Versuchsdesign (überwiegend), Datenerhebung (vollständig), Datenauswertung (überwiegend), Ergebnisdiskussion (überwiegend), Erstellen des Manuskriptes (überwiegend).

zu II. 2.: Konzeption (in Teilen), Versuchsdesign (überwiegend), Datenerhebung (vollständig), Datenauswertung (überwiegend), Ergebnisdiskussion (überwiegend), Erstellen des Manuskriptes (überwiegend).

zu II. 3.: Konzeption (in Teilen), Versuchsdesign (überwiegend), Datenerhebung (überwiegend), Datenauswertung (überwiegend), Ergebnisdiskussion (überwiegend), Erstellen des Manuskriptes (überwiegend).

IV. Namen und Anschriften nebst E-Mail der jeweiligen Mitautor*innen

Zu II.1:

Julian Keil
Christian-Albrechts-Universität zu Kiel
Biologische Psychologie
Olshausenstraße 62
24118 Kiel
keil@psychologie.uni-kiel.de

Zu II. 2:

Daniel Senkowski
Psychiatrische Universitatsklinik
der Charite im St. Hedwigs Krankenhaus
Große Hamburger Str. 5-11
10115 Berlin
daniel.senkowski@charite.de

Michael Niedeggen
Fachbereich Erziehungswissenschaft und Psychologie
Freie Universität Berlin
Habelschwerdter Allee 45
14195 Berlin
niedegg@zedat.fu-berlin.de

Julian Keil, s.o.

Zu II. 3:

Daniel Senkowski, s.o.

Martin Holtkamp
Epilepsie-Zentrums Berlin-Brandenburg
Abteilung für Epileptologie
Herzbergstraße 79
10365 Berlin
martin.holtkamp@charite.de

Bettina Wächter
Epilepsie-Zentrums Berlin-Brandenburg
Abteilung für Epileptologie
Herzbergstraße 79
10365 Berlin
b.waechter@keh-berlin.de

Julian Keil, s.o.

Datum, Unterschrift des Doktoranden

Ich bestätige die von Georgios Michail unter III. abgegebene Erklärung

Name: Julian Keil Unterschrift: __________________________

Name: Daniel Senkowski Unterschrift: __________________________

Name: Michael Niedeggen Unterschrift: __________________________

Name: Martin Holtkamp Unterschrift: __________________________

Name: Bettina Wächter Unterschrift: __________________________

Appendix C: Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit eigenständig und ohne unerlaubte Hilfe verfasst habe, und dass Ideen und Gedanken aus Arbeiten anderer entsprechend gekennzeichnet wurden.

Ort, Datum

Unterschrift

Appendix D: Curriculum vitae

Ausbildung

seit 09/2018	Promotion (Psychologie) *Freie Universität Berlin*
10/2013 – 09/2015	Master-Studium in Klinische Neurowissenschaft (MSc) *Charité Universitätsmedizin Berlin*
09/2010 – 06/2013	Bachelor-Studium in Biologie (BSc), Nebenfach Psychologie *Universität Zypern (UCY), Nikosia, Zypern*

Forschungserfahrung

03/2017 – 05/2021	Wiss. Mitarbeiter, AG Multisesensorische Integration | Charité Universitätsmedizin Berlin
10/2015 – 12/2016	Wiss. Mitarbeiter, AG Neurophysik | Charité Universitätsmedizin Berlin

Lehrerfahrung

02/2019	EEG-Daten Analyse mit FieldTrip Toolbox (Workshop) *Springschool „Kognitive Neurowissenschaften", Christian-Albrechts-Universität, Kiel*
06/2018 – 05/2021	Lernen & Gedächtnis (Seminar) *Charité Universitätsmedizin Berlin*
04/2018 – 05/2018	Grundlegende Fragen der kognitiven Neurowissenschaften und der biologischen Psychologie (Seminar) *Christian-Albrechts-Universität, Kiel*

Publikationen

Michail G, Senkowski D, Holtkamp M, Wächter B, Keil J (*submitted*) Early beta oscillations in multisensory association areas underlie crossmodal performance enhancement.

Michail G, Toran Jenner L, Keil J (2021) Prestimulus alpha power but not phase influences visual discrimination of long-duration visual stimuli. Eur J Neurosci. doi: 10.1111/ejn.15169

Michail G, Senkowski D, Niedeggen M, Keil J (2021) Memory Load Alters Perception-Related Neural Oscillations during Multisensory Integration. J Neurosci 41:1505–1515. doi: 10.1523/JNEUROSCI.1397-20.2020

Moran JK*, **Michail G***, Heinz A, Keil J, Senkowski D (2019) Long-Range Temporal Correlations in Resting State Beta Oscillations are Reduced in Schizophrenia. Front Psychiatry 10:517. doi: 10.3389/fpsyt.2019.00517 *shared first authors

Michail G, Nikulin VV, Curio G, Maess B, Herrojo Ruiz M (2018) Disruption of Boundary Encoding During Sensorimotor Sequence Learning: An MEG Study. Front Hum Neurosci 12:240. doi: 10.3389/fnhum.2018.00240

Michail G, Keil J (2018) High cognitive load enhances the susceptibility to non-speech audiovisual illusions. Scientific Reports 8:11530. doi: 10.1038/s41598-018-30007-6

Michail G, Dresel C, Witkovský V, Stankewitz A, Schulz E (2016) Neuronal Oscillations in Various Frequency Bands Differ between Pain and Touch. Front Hum Neurosci 10:182. doi: 10.3389/fnhum.2016.00182