Neurofunctional Correlates of Audiovisual Binding in fMRI, EEG and EEG-Guided fMRI

Inaugural-Dissertation

Zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

Justus-Liebig-Universität Gießen

FB 06 – Psychologie und Sportwissenschaft

Otto-Behaghel-Str. 10F

35394 Gießen

vorgelegt von

Matthias Bischoff

2009

Gutachter

PD Dr. Gebhard Sammer (Psychologie und Sportwissenschaft, Zentrum für Psychiatrie, Gießen)

Prof. Dr. Frank Bremmer (Neurophysik, Marburg)

Zusammenfassung

Ziel dieses Dissertationsprojektes war es, hämodynamische und elektrophysiologische Korrelate der perzeptuellen Bindung modalitätsübergreifender auditorisch-visueller Stimuli im menschlichen Gehirn zu identifizieren. Wie multisensorische Konvergenzareale und die zeitliche Korrelation von Arealen an der Bindung beteiligt sind, ist eine aktuelle Frage der Forschung zur Multisensorik. Im Fokus des Interesses dieser Arbeit stand daher der Zusammenhang zwischen diesen Korrelaten in der funktionellen Magnetresonanztomographie (fMRT) und dem gleichzeitig erhobenen Elektroenzephalogramm (EEG). Neben den separaten Analysen beider Signale wurde in einer kombinierten Analyse das fMRT-Signal vom gleichzeitig erhobenen EEG geleitet, d.h. das fMRT-Signal wurde durch das quantifizierte EEG-Signal parametrisch moduliert.

Die perzeptuelle Ebene der modalitätsübergreifenden Bindung auditorisch-visueller Stimuli wurde mit Hilfe des Bauchredner-Effekts erfasst. Aufgabe der Probanden war es, die zeitliche Kongruenz sowie die Raumposition auditorischer Stimuli anzugeben, während sie visuelle und auditorische Stimuli geringer Komplexität beobachteten. Die Stimuli erschienen und verschwanden entweder zu einander synchron oder asynchron. Beim Bauchredner-Effekt ruft die synchrone Darbietung räumlich abweichender auditorisch-visueller Stimuli eine Verschiebung der Lokalisierung des subjektiven Ursprungs des auditorischen Stimulus in Richtung des visuellen Stimulus hervor. Modalitätsübergreifenden Bindung wurde ins Experiment eingeführt durch zeitlich kongruente auditorisch-visuelle Stimuli, die den Bauchredner-Effekt hervor riefen. Zeitlich inkongruente auditorische und visuelle Stimuli, die keine Verschiebung der Lokalisierung hervorriefen, dienten als bimodale Kontrolle.

In einem Netzwerk multisensorischer Konvergenzareale wurde mit modalitäts-übergreifender Bindung korrelierende hämodynamische Aktivierung gefunden in der Insula, dem Sulcus parieto-okzipitalis (POS) und dem Sulcus temporalis superior (STS). An parietalen Elektroden wurde eine elektrophysiologisches Korrelat modalitätsübergreifender Bindung ermittelt, in Form von Änderungen in der neuronalen Synchronisation im Theta- und Alpha-Band. Die Zeit-Frequenz-Darstellungen zeigten unterschiedliche Synchronisations-muster für die kongruente und die inkongruente Stimulation. Die perzeptuelle Bindung modalitätsübergreifender auditorisch-visueller Stimuli spiegelte sich also in der hämodynamischen und der elektrophysiologischen Hirnaktivität wider.

Auf Ebene der einzelnen Durchgänge wurde das quantifizierte EEG-Korrelat genutzt, um das hämodynamische Signal in zur Vorhersage des Verlaufs des hämodynamischen Signals eingesetzt. Durch die Modulierung des hämodynamischen Signals konnte der Zusammenhang zwischen den betrachteten kortikalen Arealen und der EEG-Dynamik im Netzwerk der modalitätsübergreifenden Bindung aufgeklärt werden. Dieser Zusammenhang ist von besonderem Interesse im Kontext der beiden Hypothesen zur modalitätsübergreifenden Bindung; einmal entsteht die Bindung in hierarchisch aufgebauten Konvergenzarealen, als zweite Hypothese entsteht die Bindung über die zeitliche Korrelation zwischen Arealen. Als wichtigstes Ergebnis der EEG-geleiteten fMRT-Analyse zeigte sich ein klar abgegrenztes Gebiet als verbunden mit der erhöhten Synchronisation während der perzeptuellen Bindung modalitätsübergreifender auditorisch-visueller Stimuli. Bezieht man diese Aktivierung im STS auf die in der Literatur postulierten Projektionen von und zu unisenorischen Arealen, dann können die beiden Hypothesen der modalitätsübergreifenden Bindung zusammengeführt werden für die auditorisch-visuelle Bindung. Der im STS gefundene

Zusammenhang weist darauf hin, dass der STS als modulierendes Konvergenzareal der auditorisch-visuellen Bindung fungiert.

Abstract

This thesis aimed to find hemodynamic and electrophysiological correlates of the perceptual binding of crossmodal audiovisual stimuli in the human brain. How the multisensory convergence areas and the temporal correlation (synchronization) of areas are involved in binding is currently a major discussion in multisensory research. Therefore, the question of central interest in the present work was the relationship between the correlates in the functional magnet resonance imaging (fMRI) and the simultaneously acquired electroencephalogram (EEG). Additionally to the separated analyses of both signals in a further analysis fMRI was examined guided by the EEG, i.e. the fMRI signal was parametrically modulated by the quantified EEG signal.

The perceptual binding of crossmodal audiovisual stimuli was addressed in a paradigm using the ventriloquism-effect. Participants were asked to report the temporal congruency and location of auditory stimuli when observing both visual and auditory stimuli of low complexity that were switched on/off either synchronously or asynchronously. In synchronously presented spatially disparate audiovisual stimuli the ventriloquism-effect is indicated by a localization shift of the subjective origin of the auditory stimulus toward the visual stimulus. Crossmodal binding was introduced with temporally congruent audiovisual stimuli eliciting the ventriloquism-effect, whereas temporally incongruent audiovisual stimuli without a localization shift were the bimodal control.

In a network of multisensory convergence areas hemodynamic activity correlating with crossmodal binding was found in the insula, the parieto-occipital sulcus (POS), and the superior temporal sulcus (STS). At parietal electrodes an electrophysiological correlate of crossmodal binding was identified in neural synchronization changes of

theta and alpha band activity. The time-frequency representations showed different patterns of synchronization for the congruent and the incongruent stimulation. The perceptual binding of crossmodal audiovisual stimuli was reflected in hemodynamic and in electrophysiological brain activity.

On a single trial basis, the quantified EEG correlate was used to predict the hemodynamic signal in the EEG-guided fMRI analysis. This modulation of the hemodynamic signal gave information about the association of cortical areas and EEG-dynamic within the framework of crossmodal binding. The association is of special interest in the context of the two hypotheses of binding, binding by hierarchically organized convergence areas and binding by temporal correlation. As the main result of the EEG-guided fMRI analysis a very distinct area associated with increased synchronization was related to the perceptual binding of crossmodal audiovisual stimuli. When this activation in the STS is related to projections from or to unisensory areas as discussed in the literature, the two hypotheses of crossmodal binding could be combined for audiovisual binding. The association revealed in the STS provides evidence for the STS to serve as the modulating convergence area of audiovisual binding.

Acknowledgements

Contents

Chapter 1

General Introduction

The fast and comprehensive detection and interpretation of events in the outside world is of obvious benefit to any advanced organism. The human senses have evolved to specialized systems, each one sensitive to small inputs in its own domain while capable of processing complex information as well. The various senses cover a broad range of possible sources of information. However, the senses do not function as fully separated modules although they have mostly been examined as independent in perception research. The ability to combine their information gives enormous advantages; the integration of several sensory inputs increases speed as well as accuracy of detection and provides additional information of the external event. Therefore, structuring sensory information from multiple modalities into one coherent representation of the environment in perception is one of the most important functions of the brain. This structuring of internal representations of external objects or events is called binding. This thesis is concerned with the binding of audiovisual stimuli. The perception of bound audiovisual events is examined in hemodynamic and electrophysiological signals. A question of central interest is the relationship between these simultaneously acquired signals revealed by a combined analysis of functional magnet resonance imaging (fMRI) guided by the electroencephalogram (EEG). The following chapter provides an introduction to the experimental studies. First, the characteristics and distinctions of binding are illustrated in the context of multisensory integration. Next, the approach of applying multisensory perceptual effects to examine the perceptual level of binding is introduced briefly, and a summary of the established knowledge on

the areas and dynamical aspects of brain activity involved in crossmodal binding is given. Finally, an overview of methodological aspects of combining fMRI and EEG along with several considerations of the experimental design leads to the outline of the addressed research questions.

Binding

The question behind the binding process is how separated units of information related to an object or an event processed by the multitude of human senses are integrated into one coherent and conscious representation in perception. In philosophy the idea that a form of synthesis is constantly necessary to structure representations in perception can be traced back to the Critique of pure reason by Immanuel Kant (Kant & Guyer, 1998; Revonsuo and Newman, 1999). In the neuroscientific field, discussions were initiated in the context of correlating neural activity (von der Malsburg, 1981), but also in the context of so-called convergence zones in which the "fragments" of perception are brought to a coherent experience (Damasio, 1989), and in the context of the feature integration theory of attention by Treisman (1996).

Different types of binding can be distinguished regarding the characteristics that have to be 'bound'. The categorization by Treisman (1996) illustrates the fundamental range of the problem: 'Property binding' relates to the properties (or features) as shape or color of an object. In 'part binding' the parts of an object are separated from the background and the unity has to be preserved in spite of partial occlusion. 'Range binding' assures that small variations in a dimension do not reduce the unity, e.g., different shades of a color define one surface. The surface-defining properties are bound with those defining the shape of an object (e.g. orientation, curvature, and closure) in

'hierarchical binding'. The properties of an object influence the interpretation of each other in a 'conditional binding', e.g. the direction of motion depends on depth information. These five types of binding have to be stabilized over time by 'temporal binding'– a precondition of movement perception –and object features are bound to space by 'location binding'.

The following thesis distinguishes between two conceptions of binding as proposed by Revonsuo (1999). The distinction focuses on the perceptual quality of the result rather than on the elements to be bound. Revonsuo (1999) distinguishes "stimulus-related binding" which includes all processes necessary to represent stimulus entity, from "consciousness-related binding" which addresses the unifying merging in perception. The first conception approaches the problem of how features of an object or event are assembled correctly in the multitude of events constantly perceived. The types of binding categorized by Treisman (1996) describe the necessary binding processes for this conception. The second conception approaches the problem of how the features of an object or event are integrated into one unity in conscious perception or experience in terms of Damasio (1989). Thus, the decisive difference between both questions is on the phenomenal level which is comprehended only in the second conception, the consciousness-related binding.

Consciousness-related binding is possible even in the absence of a unified external stimulus and thus can constitute an illusory or hallucinatory unity in consciousness (Revonsuo 1999). Also, experiments in which ambiguous stimuli are being perceived in alternating variants, while these variants do not include physical alternations, demonstrate that physical features do not determine perception absolutely. Treisman (1996) argues the strongest evidence that neural correlates reflect the binding process is

given, when they coincide with perceived changes in ambiguous figures or with an attention capture. Capture and illusion effects can be found when more than one modality is involved in binding and a crossmodal interaction of modalities is necessary. With perceptual effects the level of consciousness-related binding is accessible in multisensory integration. Multisensory integration comprises all the mechanisms by which multimodal stimuli are converged (Stein & Stanford, 2008). At the level of single cells, multisensory integration results in an enhancement or depression of the neural response compared to the most effective component stimuli (Stein & Stanford, 2008). While the term "crossmodal" is equally used (Stein & Stanford, 2008) as "bi-", "tri-" or "multimodal" (Calvert, 2001) referring to the number of modalities, the crossing of modalities implies that the information of one sense might be changed when it is merged into one coherent representation of an event. The crossmodal percept can vary qualitatively, i.e. the spatial, temporal, or semantic information is different from the information of the single senses involved, as will be explained later. Therefore, on the behavioral level perceptual illusions can serve as indicators for crossmodal interaction.

The widening of the binding question to multiple modalities is also obvious in another way. Initially, the binding problem was studied in neuroscience in the field of feature integration in visual object recognition (e.g., Singer 1999). The discovery of the dorsal and ventral pathway of visual processing (Mishkin & Ungerleider, 1982) raised the question of how information of both streams is reunited. What information is processed along the pathways is conceptualized differently. Whereas Mishkin and Ungerleider (1982) coined the terms "where-pathway" and "what-pathway", attributing the localizing or spatial aspects of object recognition to the dorsal stream and the feature processing to the ventral stream, Goodale and Milner (1992) focus on the function of

the pathways. In their concept, the role of the dorsal stream is one of action control while the ventral stream processes object recognition. For both concepts, the integration of object features perceived in multiple senses is an obvious extension of the binding question. In the meantime, ventral and dorsal pathways are known for the auditory, somatosensory, and visual modality (Schroeder & Foxe, 2002).

Perceptual effects

Recently, crossmodal perception effects have been used to investigate crossmodal integration. One approach is to compare spatially or temporally congruent and incongruent stimulation conditions, other studies compare the occurrence and the absence of an illusion during physically identical trials (for a review, see Senkowski, Schneider, Foxe & Engel, 2008). An explanation for crossmodal effects is provided with the model of Maximum Likelihood Estimation by Ernst and Banks (2002). The interaction of modalities during crossmodal effects depends on the precision of the sensory information. The precision of stimuli in different modalities is estimated constantly for each modality, and each modality's contribution to the percept is weighted. For example, in the ventriloquism effect (e.g., Howard & Templeton, 1966), a stimulus from one modality that offers better spatial information dominates the localization of a bimodal stimulus which has weaker spatial resolution in the second modality. The model of Maximum Likelihood Estimation states that this effect is reversed if the relation of the spatial precision of stimuli is reversed. This has been demonstrated by Alais and Burr (2004). When visual stimuli were blurred, the visual stimuli were mislocalized by an auditory capture. In both versions of the capture effect the integration of sensory information leads to a percept shifted in space without the

stimuli being perceived as conflicting or separate. Also, the spatial information of one modality is not totally neglected. Therefore the ventriloquism effect does not demonstrate a total dominance of one modality, but an interaction between modalities during the merging of a percept. In this thesis, events of crossmodal binding were characterized by the occurrence of a ventriloquism effect which served as an indicator on the behavioral level. Using this indicator, brain activation during crossmodal binding was examined in brain areas known to be involved in multisensory integration.

Areas of multisensory integration

Multisensory integrating neurons were first detected in the superior colliculus (Stein & Meredith, 1993). Since then, a multitude of different brain areas have been found to show multisensory processing (for a review, see Calvert & Thesen, 2004; Stein & Stanford, 2008). A reason for the heterogeneity of results is the variance of experimental approaches. Important factors are the modalities involved, the manipulated experimental parameter (spatial, temporal or semantic content), and the paradigm used (passive stimulation, speeded responses, attention tasks, matching tasks, or percept-related illusion tasks). The scope of these factors may lead to the question "Is neocortex essentially multisensory?" as one review is titled (Ghazanfar & Schroeder, 2006). Another reason for the heterogeneous results may be the use of different analytic approaches to identify brain areas involved in multisensory processes (Calvert, 2001; Calvert & Thesen, 2004; Laurienti, Perrault, Stanford, Wallace & Stein, 2005). Recent studies provide evidence that multisensory information is also processed in sensor-specific and primary sensory areas which for a long time have been assumed to be purely unisensory (for a review, see Amedi, von Kriegstein, van Atteveldt, Beauchamp

& Naumer, 2005; Ghazanfar & Schroeder, 2006; Driver & Noesselt, 2008). In addition, multisensory processing has been found in frontal as well as subcortical areas other than the superior colliculus (Senkowski et al., 2008). However, the exact function is still unclear, even for the "classical" set of heteromodal cortical integration areas (fig. 1) which were implicated by a number of neuroimaging studies.

The superior temporal sulcus (STS) is activated during the integration of audiovisual speech presumably based on the detection of shared phonetic features (Callan, Callan, Kroos & Vatikiotis-Bateson, 2001; Calvert, Campbell & Brammer, 2000; Macaluso, George, Dolan, Spence & Driver, 2004;Raji, Uutela & Hari, 2001), by spatially nonpredictive auditory precues (McDonald, Teder-Sälejärvi, Di Russo & Hillyard, 2003), during auditory and visual object processing, as well as visual speech processing (Amedi et al., 2005). High resolution fMRI has identified areas in the STS sensitive to videos and sounds of tools (e.g., a hammer hammering). The STS may therefore have an integrative function in object identification (Beauchamp, Argall, Bodurka, Duyn & Martin, 2004) less restricted than attributed by Calvert et al. (2000) for the STS and speech processing. The superior colliculus (SC) and the insula/claustrum are activated by simultaneous onsets of stimuli across different modalities (Bushara, Grafman & Hallett, 2001; Calvert, Hansen, Iversen & Brammer, 2001). The insula/claustrum are also involved in the transfer and binding of unimodal information and visuo-tactile form integration (Amedi et al., 2005). Parts of the inferior and superior parietal lobes, the intraparietal sulcus (IPS), and the parieto-occipital sulcus (POS) are involved in the detection of multisensory stimuli that correspond in spatial location or visuo-motor aspects (Bremmer et al., 2001; Bushara et al., 2001; Callan et al., 2001; Calvert et al., 2001; Fort, Defpuech, Pernier & Giard, 2002; Molholm et al., 2002; Macaluso et al.,

2004). In visual-tactile object-recognition, the IPS is considered to be involved in object shape analysis and recognition (Amedi et al., 2005). The inferior parietal lobe is activated by synchronous auditory and visual speech stimuli at different locations (Macaluso et al., 2004).
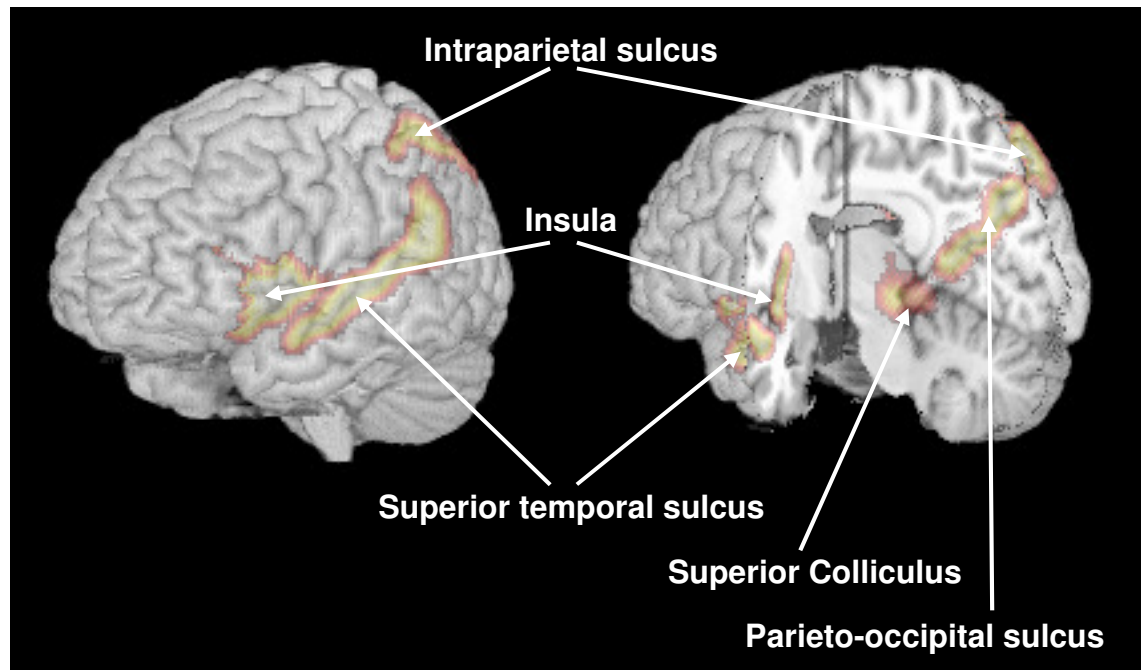
Figure 1. Convergence areas involved in audiovisual multisensory integration. For details of area definition see methods section in chapter 2.

Dynamical brain activity

In their pioneering work on single-cell activity of neurons in the superior colliculus of the cat, Stein and Meredith (1993) assumed a hierarchical order of areas where sensory signals from different modalities are transferred in overlapping spatial maps in subcortical regions and merged in higher convergence areas. Senkowski, Schneider, Foxe, and Engel (2008) name three problems such a pure convergence model cannot solve. First, multisensory interactions are found in primary cortices, so that at least feedback connections would have to exist. Second, the convergence stream would have

to be fast and flexible to enable rapid recombination of crossmodal stimuli into novel percepts. Third, a feedforward convergence model would lose the low-level information in the stream of processing towards a non-compositional high-level representation, but low-level information can remain accessible (Senkowski, Schneider, Foxe & Engel, 2008). As a solution, dynamical aspects of brain activity were suggested to play a role in multisensory integration (Singer, 1998; for a review see Senkowski, Schneider, Foxe & Engel, 2008).

The so-called temporal correlation hypothesis states that the synchronization of distributed neural signals in networks of brain areas is a key mechanism for the assignment of information and multisensory integration (Singer & Gray, 1995; Singer, 1999; Driver & Spence, 2000; Engel et al., 2001; Bauer, 2008; Driver & Noesselt, 2008; Senkowski, Schneider, Foxe & Engel, 2008). Concurrent activity in areas not adjacent or close to one another is observed by oscillatory responses in the EEG (Milner, 1974; von der Malsburg, 1981). The various frequency bands of such synchronized oscillations have been related to a variety of cognitive and sensorimotor processes. Most studies on multisensory integration have focused on synchronization in the gamma band (for a review, see Senkowski, Schneider, Foxe & Engel, 2008). Synchronization of cell assemblies in the gamma frequency range was found during feature integration in the visual cortex of the cat (Gray, König, Engel & Singer, 1989; Singer & Gray, 1995). This coherent electrophysiological mass activity may reflect unimodal binding; similarly synchronized EEG of unisensory areas has been proposed to reflect the mechanism of crossmodal integration (for a discussion, see Engel & Singer, 2001). However, coherence in gamma band frequencies is linked to a variety of processes. Gamma band synchrony was found not only in visual but also in auditory,

somatosensory, and olfactory processing, during many different tasks as feature integration, surface segregation, binocular rivalry, visual motion, attentional processing, sensorimotor integration, movement preparation, and memory encoding, delay and retrieval (for a review, see Senkowski, Schneider, Foxe & Engel, 2008). Similar to the effects on gamma synchronization, changes in lower frequencies in response to multisensory processing are known. Stimulation in two modalities evokes a power increase in the theta frequency band (Basar, 1999; Basar, Basar-Eroglu, Karakas & Schürmann, 2001). The bimodal gain was only reflected in the theta band, even though alpha band responses to stimulation in either single modality were recorded (Schürmann and Basar, 1999). Analyses of the event-related (de-)synchronization processes (ERD/ERS; Pfurtscheller & Aranibar, 1979; Pfurtscheller & Lopes da Silva, 1999) are more time sensitive. ERD/ERS can be described by amplitude changes in relation to a reference interval. Event-related band power changes have been found in response to stimulation and during task processing. An increase in event-related theta power is proposed to reflect the binding of a memory element with temporal information, e.g., object "x" appeared after object "y" (Klimesch, 1999). The event-related increase in theta has been found during the successful encoding and retrieval of episodic memory, while processing of semantic information in working memory is not related to changes in theta (Klimesch, 2006). Klimesch (2006) concludes that the increase in event-related theta power is a correlate of the creation of this 'episodic code', the association of the time information relative to the situational context, i.e., the sequence of events. The theta band responses have been found to be concurrent with gamma band responses in human EEG; phase coupling between (frontal) theta and gamma has been reported during memory maintenance in a memory task (Schack, Vath, Petsche, Geissler &

Moller, 2002). The authors attributed this to a functional linkage between a prefrontal network operating in gamma oscillations and a cortico-limbic network operating in theta oscillations. Exactly which intra-area or inter-area processes are observable in which frequency band is far from clear. All the more, the hypothesis of crossmodal binding by hierarchically organized dedicated areas and the hypothesis of temporal correlation as the driving mechanism of binding can be combined by introducing sets of transmodal nodes (Mesulam, 1998). These nodes are proposed to act as epicenters for convergence across areas; neural synchronization in these transmodal nodes could affect synchronization of other areas.

Combined fMRI and EEG

One approach to investigate the relationship of brain areas and neuronal dynamics is combining the high spatial resolution of functional magnetic resonance imaging (fMRI) with the high temporal resolution of electroencephalography (EEG; for a review, see Menon & Crottaz-Herbette, 2005). The relation between neuronal activity and the cerebral blood flow is labeled neurovascular coupling. A linear correlation between the stimulus-driven local field potential (LFP) activity– the presynaptic or input activation – and the hemodynamic response was found by Logothetis, Pauls, Augath, Trinath and Oeltermann (2001) in anesthetized monkeys, validating the assumptions of neurovascular coupling. A stronger correlation was shown for the BOLD-signal and LFPs than between multi-unit or single-neuron activity– the postsynaptic or output activation –and the BOLD-response (Logothetis & Pfeuffer, 2004). Since EEG reflects spatially summed and volume-conducted LFPs, a combined EEG and fMRI analysis allows a high temporal and spatial resolution of a neuronal process. However, it may

not be registered when subthreshold membrane potentials of neurons are altered without eliciting spikes by feedback and modulatory processes (Logothetis & Wandell, 2004).

The spatial scale of scalp-recorded EEG is several centimeters compared to about 3mm in fMRI or even 0.5mm at very high field strengths for human imaging; in animal studies 0.25mm were achieved with small surface coils in a tradeoff between resolution and volume coverage (Menon & Crottaz-Herbette, 2005; Logothetis, 2008). Moreover, the scalp recorded EEG is not only damped by the skull, the recorded sum signal can be influenced by sites distal from the electrodes. While technically the temporal resolution of fMRI can also reach 40ms per slice, the time to acquire a volume covering the whole brain is 1 to 3 s due to artifacts and physiological factors (Logothetis, 2008). In EEG, the temporal resolution is 1 ms, signal fluctuations (EEG-components) are identified in terms of 100ms (Menon & Crottaz-Herbette, 2005). The combination of EEG and fMRI gains explanatory power, especially when inter-trial variance is not eliminated by averaging. Combining nonsimultaneous fMRI and EEG acquisition spares the elaborate correction of the MR-scanner artifact in the EEG signal and also, the signal-to-noise-ratio is better for data obtained outside the scanner (Allen, Josephs, & Turner, 2000; Mandelkow, Halder, Boesiger, & Brandeis, 2006). However, attentional, motivational, as well as learning effects (familiarity with task) can differ between the two sessions and only brain response which is time locked to the stimulus or response can be examined in nonsimultaneous combination (Menon & Crottaz-Herbette, 2005). Also, arousal levels which may differ during fMRI due to the environment inside a scanner affect brain activation during cognitive tasks (Matsuda et al., 2002). Acquiring fMRI and EEG data simultaneously in the scanner assures that both data signals reflect the same neuronal processes in as far as the task is performed with the same ability,

strategy, and arousal (Menon & Crottaz-Herbette, 2005). The individual and inter-trial variance in crossmodal binding of audiovisual events with a ventriloquism effect hold valuable information, so that the single trial analysis of fMRI and EEG signals is highly advantageous.

Experimental design

The present projects addressed the consciousness-related aspects of multisensory integration, i.e. the crossmodal binding resulting in a unified percept. Therefore the paradigm was required to include a form of indicator for perceptual binding, in order to distinguish successful integration of multisensory stimuli from failed stimulus manipulation. Furthermore the paradigm had to be adapted to the constraints examined by both signals. Considering these premises, the audiovisual events were arranged in mini-blocks of eight seconds, providing enough time of stimulation for the BOLD-signal to accumulate sufficiently, cover a higher number of stimuli in the EEG, and still verify the manipulation with the behavioral responses. Blocking audiovisual events is advantageous for increasing the signal-to-noise ratio in fMRI; however, it prevents the parsing of specific component processes. On the other hand a slow event-related design– characterized by intertrial-intervals long enough for the BOLD-signal to subside back to baseline –would have prolonged the duration of the experiment very unfavorably unless only very few trials had been recorded. To obtain reliable components in EEG typically 30–100 stimuli trials are needed (Menon & Crottaz-Herbette, 2005). In a fast event-related design with inter-trial-intervals of a few seconds, recording the behavioral indicator (the subjects responses) would be disproportionately time consuming and interfering with the BOLD-response induced by the stimulation.

Thesis outline

This thesis included the conduction of three projects to investigate brain activity during crossmodal binding. Following the concept of consciousness-related binding, the perceptual level of audiovisual integration was focused with the help of the audiovisual ventriloquism effect. In the first project, based on the known candidate areas of multisensory integration, integration, hemodynamic responses to crossmodal binding were examined in a defined set of regions using fMRI. A hemodynamic correlate was found in temporal and parietal areas. In the second project, brain dynamics were transformed to time-frequency representations in order to analyze low-frequency patterns during binding. An electrophysiological correlate in the theta and alpha band range of time-frequency representations of the electroencephalogram was established.

The most innovative contribution to the understanding of neural correlates of crossmodal binding are the results of the combined analysis of the activation localization and the dynamical data in the third project. The relationship of both simultaneously measured correlates was revealed in a combination of fMRI and EEG. A distinctive activation in the right superior temporal sulcus was revealed when the hemodynamic response was parametrically modulated by the quantified single trial energy weighted with the time-frequency pattern of binding.

In chapters 2 to 4 the projects are presented as independent manuscripts intended for publication in scientific journals. Some redundancy is accepted to allow for the intelligibility as separated parts. The closing chapter 5 provides a review of the findings from all three projects.

Chapter 2

Utilizing the Ventriloquism Effect to Investigate Audiovisual Binding

Audio-visual binding - as subset of crossmodal integration - describes the combination of information across both these senses to the subjective unified perception of a bound object. We investigated audio-visual binding by using the ventriloquism effect (localization of a sound is biased towards and by a simultaneous visual stimulus) to act as an indicator for perceived binding. Simple visual and auditory stimuli were presented synchronously or asynchronously. fMRI was recorded during task performance (n = 19 subjects) in order to reveal activation in areas discussed to be involved in multisensory processing in the literature.

Contrasting trials with reported ventriloquism effect vs. the no-binding condition revealed activation in the right insula, superior temporal sulcus and parieto-occipital sulcus. Implementing the ventriloquism effect allows us to relate these activations to consciousness-related processes, which probably are different from stimulus-driven multisensory integration in subcortical areas.

Introduction

Many perceptual events are not restricted to a single modality but constructed from information of two or more senses, which is experienced as a coherent and unified percept. Multisensory input can increase the detectability of external stimuli, disambiguate their discrimination, speed up responsiveness and facilitate memory retrieval processes (Stein & Meredith, 1993). New information can emerge by input integration, as is the case with the the perception of taste, which results from a combination of gustatory and olfactory input (O'Hare, 1991). Moreover, taste can be influenced by auditory cues (Zampini & Spence, 2004).

Hemodynamic and electromagnetic studies show that multisensory integration mechanisms are distributed across distinct neuronal networks (reviews Calvert, 2001; Calvert & Thesen, 2004; Amedi et al., 2005; Beauchamp, 2005). Several specific networks have been proposed to be specifically involved in different aspects of information processing, e.g. for coincidence in time, or correspondence in space or content (Calvert & Thesen, 2004). Recent studies provide evidence that such information also is processed in primary sensory areas, which for a long time have been assumed to be purely unisensory (Amedi et al., 2005). The brain forms subjective unified perceptions of multisensory objects from unisensory cortical representations. Unisensory cortical representations are distributed over the cortex, and object features are generated from unimodal sensory information along different pathways. The integration of object features usually is referred to as 'crossmodal binding' within multisensory integration research (Revonsuo, 1999). When focusing on the percept of binding, it is necessary to distinguish between stimulus- and consciousness-related

binding. Stimulus-related binding includes all processes necessary to represent stimulus entity– conscious as well as unconscious ones. Consciousness-related binding occurs when a unified perception is reported in the absence of a unified external stimulus and thus constitutes an illusory or hallucinatory unity in consciousness (Revonsuo, 1999). Tasks of consciousness-related binding offer the opportunity to investigate the correlate of the 'binding-percept', if a perceptual check of conditions can be used. Evidence of the occurrence of consciousness-related binding can be provided by perceptual illusory effects which depend on the manipulation of perceived stimulus features. Such an illusion is the ventriloquism effect (Howard & Templeton, 1966). It describes the mislocation of an auditory stimulus towards a simultaneous visual stimulus. For example, an eponymous ventriloquist synchronizes the movements of a puppet's head and mouth with his own speech while avoiding movements of his own head or lips. This effect was suggested to result from 'visual capture' of the speech information of the dominant visual sense. Recently, Alias and Burr (2004) offered support for an alternative explanation, ascribing the effect to the higher spatial resolution of the visual information as suggested by Ernst & Banks (2002). Presenting blurred visual stimuli, which were difficult to locate, simultaneously with sounds that could be better located, they showed a converse effect: The blurred visual stimuli were mislocalized towards the clear auditory ones. Deblurring of the visual stimuli shifted back localization towards the visual stimuli, indicating 'visual capture'.

This and previous evidence suggests that the ventriloquism effect does not depend on the direction of deliberate visual attention (Bertelson, Vroomen, deGelder & Driver, 2000) or automatic visual attention (Vroomen, Bertelson & deGelder, 2001). In EEG the ventriloquism effect evokes a mismatch negativity closely resembling the mismatch

negativity evoked by an actual sound shift, indicating that the perceived and the actual sound shift are comparable (Stekelenburg, Vroomen & de Gelder, 2004). Psychophysical results also show that the ventriloquism effect is an audio-visual perceptual effect (Bertelson & Aschersleben, 1998). In conclusion, the ventriloquism effect can be defined as a shift of localization which results from an integration process that involves binding. Thus, the ventriloquism effect can be implemented as a perceptual check when investigating the percept of consciousness-related binding.

The neuronal correlate of consciousness-related binding is not identified so far. A set of regions has been found to be involved in multisensory integration processes. The superior colliculus and the insula/claustrum are activated by simultaneous onsets of stimuli across different modalities (Bushara et al., 2001; Calvert, et al., 2001). However, the insula/claustrum was also suggested to be involved in the transfer and binding of unimodal information, and visuo-tactile form integration (Amedi et al., 2005). Parts of the inferior and superior parietal lobes, the intraparietal sulcus (IPS) and the parieto-occipital sulcus (POS) are involved in the detection of multisensory stimuli that correspond in spatial location or visuo-motor aspects (Bushara et al., 2001; Callan et al. 2001; Calvert et al,. 2001; Fort et al., 2002; Molholm et al., 2002; Macaluso et al., 2004). In visual-tactile object-recognition, the IPS is considered to be involved in object shape analysis and recognition (Amedi et al., 2005). In a study investigating audio-visual speech with PET, activation was found in the right inferior parietal lobe evoked by synchronous auditory and visual speech stimuli at different locations, i.e. the condition of a ventriloquism effect (Macaluso et al., 2004). Activations in the superior temporal sulcus (STS) were observed during the integration of audio-visual speech, presumably based on the detection of shared phonetic features (Callan et al., 2001;

Calvert et al., 2000; Macaluso et al., 2004; Raji et al., 2001). STS is also suggested to be involved in dynamic aspects of visual object processing, visual speech processing and auditory object processing (Amedi et al., 2005). Performing EEG source localization with inverse dipole modelling, ERPs in STS were found to be modulated by spatially nonpredictive auditory precues (McDonald et al., 2003). Areas in the STS sensitive to videos and sounds of tools (for example a hammer hammering) have been identified by high-resolution fMRI; the STS therefore may have a more general integrative function in object identification (Beauchamp et al., 2004).

A reason for the heterogeneity of results may be the use of different analytic approaches to identify brain areas involved in multisensory processes (Calvert, 2001; Calvert & Thesen, 2004; Laurienti et al., 2005). We consider the comparison of a congruent (synchronous) with an incongruent (asynchronous) condition to be adequate for analyzing multisensory integration. By manipulating the correspondence of stimuli (e.g. temporal synchrony or spatial coincidence), conditions of different crossmodal congruency can be investigated. While both conditions of a congruency task are bimodal, activation-differences in response to congruent and incongruent stimuli can clearly be ascribed to crossmodal interaction. Occurrence of the ventriloquism effect functions as a filter in this comparison to address consciousness-related binding.

The aim of this study was to investigate the functional significance of cortical areas previously associated with audio-visual multisensory processing for crossmodal consciousness-related binding. Consciousness-related binding was recognized by the occurrence of the ventriloquism effect in an audio-visual task. Two visual stimuli were presented, one of which was presented synchronously with an auditory stimulus. A localization shift of the auditory stimulus towards the synchronously presented visual

stimulus (ventriloquism effect) was interpreted as indicating consciousness-related binding to an audio-visual object. In contrast, the absence of a ventriloquism effect after presentation of an asynchronous visual-auditory stimulus combination was supposed to indicate unbound object-features. Thus this study investigated brain activation specific to the percept of a "bound" object in contrast to unbound stimuli. Hemodynamic brain activation was measured with functional magnetic resonance imaging (fMRI) in order to identify brain regions involved in consciousness-related processing of bound audio-visual stimuli. Of trials consisting of synchronously presented stimuli, those with occurrence of a ventriloquism effect were selected for comparison with a second bimodal audio-visual condition, consisting of asynchronously presented stimuli. The comparison of congruent versus incongruent permitted the attribution of activation differences to integration processes without applying the criterion of super-additive strength of hemodynamic responses as - for example - in a conjunction bimodal versus unimodal contrast.

In conclusion, this study identifies activation patterns during consciousness-related binding and thus helps to elucidate the role of neuronal multisensory integration processes for perceptual binding.

Methods

Participants

Twenty-five participants were recruited for the study, nineteen were included in the analysis (12 female, age= 20.25 – 34.83; mean=26.19; S.D.=5.021). Five subjects had to be excluded due to lacking the ventriloquism effect, one for technical reasons.

All participants were naive as to the purpose of the experiment. None of them had a history of or a current neurological or psychiatric illness, all reported normal or

corrected-to-normal vision (MR-suitable glasses) as well as normal hearing. All participants provided written informed consent.

Visual stimulation

Visual Stimuli consisted of pictures showing two vertically aligned grey discs (RGB 200,200,200). The diameter of each disc was 240 pixels. A fixation cross (length 50 pixel, RGB 200, 200, 200,) was placed in the center between both discs. Each disc was presented at three different rates, 6 times in 8 seconds, 8 times in 8 seconds, and 10 times in 8 seconds (0.75 Hz, 1.0 Hz and 1.5 Hz). Discs were erased after 200ms leaving the screen black with the exception of the fixation cross. The duration of each trial was 8 seconds. Depending on the experimental condition, discs on both sides appeared simultaneously or with an onset asynchrony (360ms, 200ms and 120ms).

Visual stimuli were displayed on a back-projection screen (1280 x 1024 pixels) behind the scanner. Subjects could watch the screen with the help of a dual-mirror, which was mounted to the head coil. The angle of vision was approximately 11° vertical and 18° horizontal (each disc occupying 4°).

Auditory stimulation

Auditory stimuli were sinusoidal tones with a frequency of 440 Hz. The duration of each tone was 120 ms. The tones were presented at the same rates as the visual stimuli (0.75, 1.00, 1.25 Hz). In each trial, the frequency of tone presentation matched the presentation of one, both or none of the discs. If the presentation of the tone and a disc were matched, the tone started with an onset-asynchrony of 40ms and was also ended 40ms before the disc disappeared in order to secure a dominance of the visual stimulus in terms of the ventriloquism effect by higher spatial information (Ernst & Banks 2002,

Alais & Burr, 2004). All the auditory stimulation was transmitted in mono mode on stereo headphones resulting in perception of the tone as being spatially centered.

The auditory stimuli were transmitted to the subjects' ears using MR-compatible electrodynamic high quality sound stereo headphones. In these headphones the permanent magnet is removed and is substituted by the static B0-field of the MR-scanner. This technique enables the use of electrodynamic headphones instead of using low fidelity standard air headphones. A disadvantage of such a system is that the gain level of the channels depends on the location of the earphone in the scanner. We mounted the headphones to the head fixation clamps, which in turn are part of the patient table. This configuration assured both a fixed placement in the MR-scanner and a good fit of the headphones to the subject's head. Stereo-balance and gain were calibrated for each subject during a series of test-tones. The balance of the headphones was individually adjusted, until the participants judged the sound as originating from the central point. An fMRI-sequence was run during these adjustments to match all the conditions during the subsequent task.

Task

Table 1

Combinations of visual and auditory stimulation

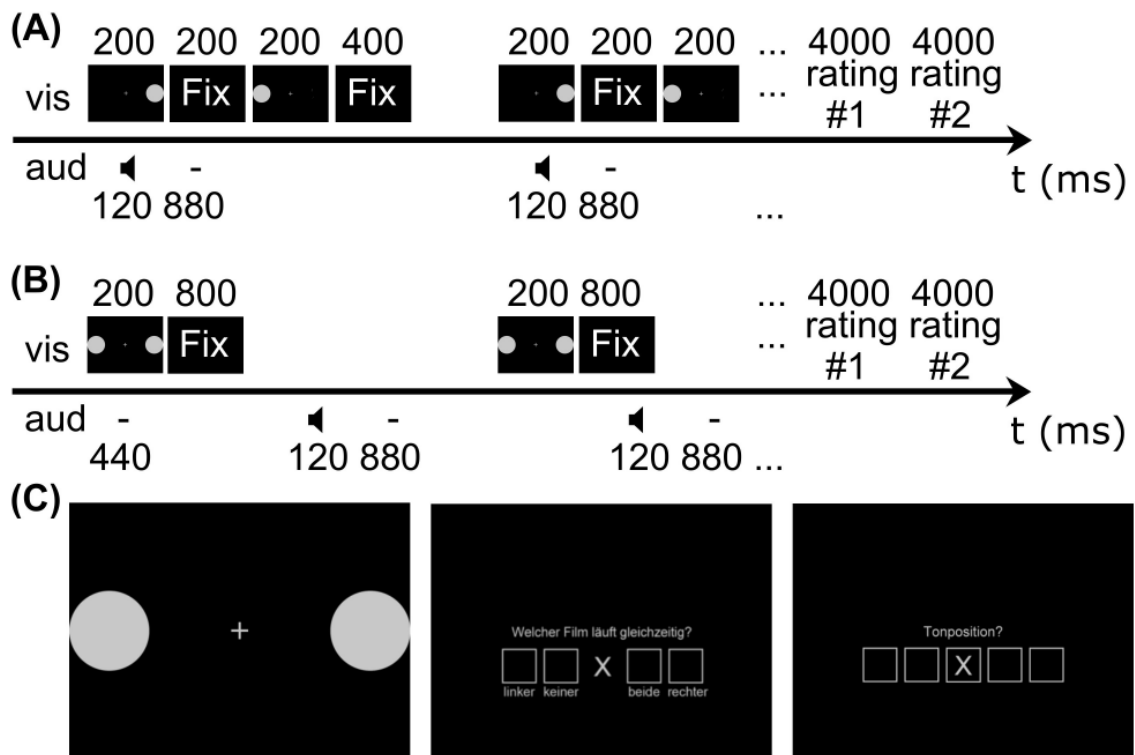| Visual stimulation		Auditory stimulation
Timing		Position
Disc on the left side - tone	Disc on the right side - tone	
synchronously	asynchronously	central
asynchronously	synchronously	central
asynchronously	asynchronously	central

Figure 1. Examplary timecourse for repetition frequency of 1Hz and screens of the experiment. The durations of stimuli are quoted in ms. Stimulation is given for 8 seconds in all trials, followed by two ratings for 4 seconds each. (A) Experimental condition, auditory stimuli onset is 40ms after visual stimuli onset, (B) control condition; vis = visual stimuli, aud = auditory stimuli, Fix = fixation screen.

In table 1 the three combinations of visual and auditory stimulation used in this study are listed. The first and second combination refer to synchronous presentation of one of the disc-stimuli left or right of the fixation-mark and the tone-stimulus, the second disc-stimulus is presented asynchronously, and appears without a tone respectively (trials of interest, experimental control). These configurations form the experimental condition, during which the occurrence of the ventriloquism effect signals binding induced by synchrony of the audio-visual stimulation. In the third combination both disc-stimuli are delayed with regard to the tone-onsets. This configuration is the

control condition, which represents audio-visual perception without integration of the modalities.

Each trial consisted of one disc-tone-combination which was presented for 8 seconds. After this period a line of four boxes appeared on the screen. The subject's task was to check one of the boxes by moving a cursor with two buttons to the box indicating which of the visual stimuli has been presented synchronously to the auditory stimulus and confirming the choice with a third button. The answer boxes not chosen were dimmed out to show that the response was locked. Possible responses were 'left', 'none', 'both', 'right'. Responses had to be given during a fixed interval of 4 seconds duration. Subsequently, this scale was erased and another scale appeared on the screen, on which subjects were asked to indicate where the tone had originated. For this purpose the scale consisted of 5 unlabeled boxes. Again, the response interval was 4 seconds. All responses were performed with the index finger of the right hand. Subjects were instructed not to report their expectation but to concentrate on perception. During the following inter-trial-interval only the fixation cross remained visible. The time between two trials was 21+/- 1.5 seconds; stimulus onsets were jittered in relation to the BOLD-measurement.

Every experimental run consisted of 48 trials. Twenty four trials belonged to the experimental condition (synchronous), the remaining 24 trials belonged to the control condition (asynchronous). The presentation order was pseudorandomized. Stimulus presentation and response acquisition was controlled by 'Presentation'-software (Version 0.76, Neurobehavioral Systems, Albany). Prior to the fMRI-session, subjects practiced up to 10 sample trials of the task to become familiar with the keypad and to ensure that the instructions were understood correctly.

FMRI acquisition

fMRI was measured using a 1.5 Tesla Siemens Symphony whole body scanner with Quantum-gradients. Functional images were obtained using a single shot T2*-weighted gradient-echo planar imaging (EPI) sequence. Thirty slices were acquired in descending fashion to cover all parts of the brain. Slice thickness was 4 mm with 1 mm gaps. Other MR-scanning parameters were: repetition time TR = 3000 ms, echo time TE = 59 ms, flip angle = 90°, field of view FOV = 192x192mm, voxel size = 3x3x5 mm, in-plane image resolution was 64 x 64 voxels.

An anatomical MRI was acquired using a T1-weighted, three-dimensional MPR sequence. This whole brain 3D-image was used for control of the normalizing procedure and for the purpose of documentation.

Behavioral Data analysis

To calculate if a ventriloquism effect had occurred, only trials in which synchrony between visual and auditory stimuli was reported correctly, were retained (i.e., participants chose the correct visual stimulus that was synchronous to the sound in the experimental condition, or chose 'none' in the control condition). This was done, because we wanted to exclude effects of 'temporal ventriloquism' (shift of perceived time by spatial identical stimuli; Spence & Squire, 2003; King, 2005). A ventriloquism effect was considered if participants chose one of the two boxes for the sound-localization question in the same direction (left or right) as the synchronous visual stimulus. Thus for a trial of the control condition to be correctly answered in terms of the study, the central sound-position had to be reported. If a response was not given, the trial was considered to be 'incorrect'.

Effects and interactions of condition and repetition frequency were analyzed with a (2x3) ANOVA for the correct answers of each question ("synchrony of stimuli" and "localization of sound"). The ratings of sound localization were used for kappa coefficient calculation. In this study, Kappa represents the agreement between a fictive 'ideal rater', who only gives correct responses, and the subject's responses. The ideal rater always indicates the sound to be localized on the same side as the synchronous visual stimulus. In the control condition, the ideal rater always checks the central box to indicate central sound localization. Both response alternatives indicating left sound localization were pooled into one left response, both right alternatives were pooled respectively. The ratings of each subject were compared to the ideal rater's responses, resulting in a 3 (ideal rater's ratings: left, center right) x 3 (subject's ratings: left, center, right) matrix of ratings.

Kappa estimates the agreement between two raters, adjusted for the amount of agreement that could be expected due to chance alone (Cohen, 1960). Kappa is less than or equal to 1, where 1 implies perfect agreement.

FMRI analysis

FMRI data were analyzed using SPM2 (Wellcome Institute of Neurology at University College London, UK. http://www.fil.ion.ucl.ac.uk/spm). The first five images of each EPI-recording session were discarded to account for the time needed for the magnetic field to achieve a steady state. The EPI-data were preprocessed including movement correction and correction of field image distortion, normalization into the MNI-reference space and smoothing (FWHM = 8 mm).

At the 1st level, a general linear model was applied to the data of each subject. Regressors and contrasts were labeled as 'correct' in synchronous stimuli, when

synchrony and a localization shift in the direction of the visual stimulus synchronously presented to the auditory one were reported. In asynchronous trials, 'correct' indicated asynchrony and localization of the tone in the center was reported. 'Incorrect' labels synchronous stimuli, reported as asynchrony without a localization shift or a shift in the wrong direction. Regressors were 'asynchronous stimuli, correct', 'synchronous stimuli at the left, correct', 'synchronous stimuli at the right, correct', 'asynchronous stimuli, incorrect', 'synchronous stimuli at the left, incorrect', 'synchronous stimuli at the right, incorrect', 'response', 'rest', 'asynchronous stimuli, response missing', 'synchronous stimuli at the left, response missing' and 'synchronous stimuli at the right, response missing'.

The movement parameters computed during the realignment procedure were included in the models. Trials with synchronously presented stimuli (left and right side) were termed 'experimental'; trials of asynchronously presented stimuli were labeled 'control'. Contrasts of interest were 'correct experimental vs. correct control', 'experimental vs. control', 'incorrect experimental vs. correct control', 'correct experimental vs. incorrect experimental' and 'correct control vs. incorrect control'. These contrasts of interest were used for group analyses. Statistical parametric maps were computed using the kappa coefficients of participants as simple regressor. For all voxels one-sample t-tests were calculated and corrected for familywise error (Friston, Frith, Liddle & Frackowiak, 1991; Worsley, Evans, Marrett, & Neelin, 1992).

Results were computed based on region of interest (ROI) analyses, limiting the correction for multiple testing to a-priori specified regions of the brain. Regions were selected for ROI-analysis, which have been proposed to be involved in multisensory integration processes in previous studies (review Calvert &Thesen, 2004, Amedi et al.

2005). The ROI included in the current study were bilateral insula, superior temporal sulcus (STS) – without the temporal pole – superior colliculi (SC), intraparietal sulcus (IPS) and parieto occipital sulcus (POS).

Masks for these regions were created using the MARINA-tool ("MAsks for Region of INterest Analysis", by Walter et al., 2002). This program allows the creation and editing of masks in Analyze format based on the anatomical parcellation of the brain published by Tzourio-Mazoyer et al. (2002). It provides predefined masks based on the coordinates of brain structures from the SPM toolbox AAL (automated anatomical labeling; Tzourio-Mazoyer et al., 2002). Voxelsize is 2x2x2 mm. To create a ROI-mask for a gyrus, masks of adjacent areas were loaded; voxels belonging to the mask had a value of one. The first one was loaded and smoothed (FWHM=5mm), the mask was thresholded and voxels belonging to the mask with a probability <.15 were deleted. The same was done for the adjacent area. Using another threshold (>.95) the intersecting part was selected and the outer parts removed manually. The resulting mask was filled with ones, smoothed (FWHM=5mm) and voxels belonging to the mask with a probability <.25 were deleted. This was done for left and right hemispherical masks of IPS, STS and POS separately to account for hemispherical differences in size and shape. The masks of left and right insula were taken directly from the AAL-set. The mask for the superior colliculi was adapted from Nielsen and Hansen (2002). Alignment of the masks was checked by visual inspection and by projection on a MNI-template using MRIcro V.1.39 (Rorden & Brett, 2000).

Results

Behavioral Responses

Synchrony of stimuli was recognized correctly by participants on average in 22.05/24 trials (92%) in the experimental condition (min=18; max=24; S.D. = 1.72). In the control condition the mean number of correctly identified asynchronous trials was 22.84/24 trials (95%; min=17; max=24; S.D.=1.71). The 2x3 (conditions x repetition frequencies) ANOVA resulted in no main effect of repetition frequency or condition, but an interaction "frequency * condition" ($F[2, 36]=4,154$; p=0,024; Mauchly-test for sphericity p = .468). Stimuli presented with a frequency of 0.75Hz were answered correctly 7.05/8 trials (88%) in the experimental condition (min=4; max=8; S.D.=1.08) and 7.84/8 (98%) in the control condition (min=6; max=8; S.D.=0.50); the difference was significant (p=0.001). Stimuli presented with a frequency of 1.00Hz were processed correctly in 7.42/8 trials (93%) in the experimental condition (min=6; max=8; S.D.=.69) and 7.47/8 (93%) in the control condition (min=5 ; max=8 ; S.D.=.84). Stimuli presented with a frequency of 1.25Hz were answered correctly 7.58/8 trials (95%) in the experimental condition (min=6; max=8; S.D.=.69) and 7.53/8 (94%) in the control condition (min=4 ; max=8 ; S.D.=1.02). Overall, 434 trials of the control condition and 419 trials of the experimental condition were correct.

Sound was localized correctly (tone localized in direction of left/right visual stimulus) in the experimental condition in 17.11/24 trials (71%; min=12; max=22; S.D.=2.87). In the control condition 12.68/24 trials (53%) were answered correctly (tone localized in center) in mean (min: 3; max: 23; S.D.: 4.80). In the ANOVA (2 conditions x 3 repetition frequencies) the main effect of condition showed significance

(F[1,18]=17,56; p=0,001). The main effect of repetition frequency and the interaction "frequency * condition" did not reach a level of significance.

Both questions were answered correctly in 16/24 trials in the experimental condition (67%, min=11; max=21; S.D.=2.87), and in 11.79/24 trials in the control condition (49%, min=3; max=22; S.D.=4.70). In the ANOVA (2 conditions x 3 repetition frequencies) the main effect of condition showed significance in this analysis, too (F[1,18]=14,89; p=0,001). The main effect of repetition frequency and the interaction "frequency * condition" did not reach a level of significance.

FMRI-Data ROI-Analyses

Regions of interest analyses were conducted for the insula, STS, SC, IPS and POS. The analysis of the contrast 'correct experimental vs. correct control' showed activations in the right insula, superior temporal sulcus and parieto-occipital sulcus. No activation was found in the superior colliculi and intraparietal sulci.

Contrasting 'experimental vs. control' showed activation only in the ROI-mask of the left superior temporal sulcus. Contrasting 'incorrect experimental vs. correct control' did not reveal activation in any ROI. Contrasting 'correct experimental vs. incorrect experimental' showed activation in bilateral parieto-occipital sulcus. Contrasting 'correct control vs. incorrect control' showed activation in bilateral insula and intraparietal sulcus.

Table 2

Maxima of activations in ROI-masks of candidate areas (analysis at voxel-level). Included are first maxima in ROIs, mask size in voxel, T-values and corresponding p corrected for familywise error (FWE).

Side	Structure/mask	mask size (voxel)	contrast: 'correct experimental vs. correct control' MNI-coordinates x	y	z	T	$p_{FWE-corrected}$	contrast: 'experimental vs. control' MNI-coordinates x	y	z	T	$p_{FWE-corrected}$
L	IPS	478	-27	-60	45	3.28	0.322	-30	-75	48	2.97	0.444
R	IPS	458	36	-63	57	3.08	0.386	33	-57	57	3.79	0.151
L	POS	541	-6	-72	27	5.67	0.008	-9	-72	36	3.59	0.235
R	POS	586	3	-72	36	4.43	0.075	15	-54	3	2.68	0.634
L	STS	671	-48	-45	15	5.35	0.018	-48	-48	21	5.56	0.011
R	STS	866	51	-27	-9	4.83	0.061	48	-18	-12	3.41	0.423
L	Insula	786	-36	-21	24	2.63	0.755	-39	-21	21	3.47	0.351
R	Insula	732	36	-27	18	5.83	0.007	36	-21	18	3.93	0.187
L/R	SC	463	-3	-36	6	3.06	0.375	9	-21	-12	2.79	0.481

Side	Structure/mask	mask size (voxel)	contrast: 'correct experimental vs. incorrect experimental' MNI-coordinates x	y	z	T	$p_{FWE-corrected}$	contrast: 'correct control vs. incorrect control' MNI-coordinates x	y	z	T	$p_{FWE-corrected}$
L	IPS	478	-24	-69	33	3.44	0.249	-48	-42	51	3.58	0.227
R	IPS	458	36	-78	42	3.13	0.344	30	-60	45	4.44	0.060
L	POS	541	-15	-48	12	4.47	0.063	-3	-84	42	3.25	0.372
R	POS	586	12	-63	12	4.38	0.075	12	-48	3	4.01	0.143
L	STS	671	-63	-45	3	2.72	0.666	-45	-6	-12	2.07	0.920
R	STS	866	51	-39	0	2.37	0.872	45	-51	21	4.18	0.166
L	Insula	786	-36	-3	-3	2.94	0.585	-36	21	-9	3.60	0.319
R	Insula	732	39	6	6	3.36	0.374	30	24	-12	5.82	0.008
L/R	SC	463	no suprathreshold voxels					-6	-27	0	3.79	0.155

Abbrevations. L: left, R: right, IPS: intraparietal sulcus, POS: parieto-occipital sulcus, STS: superior temporal sulcus, SC: superior colliculi.
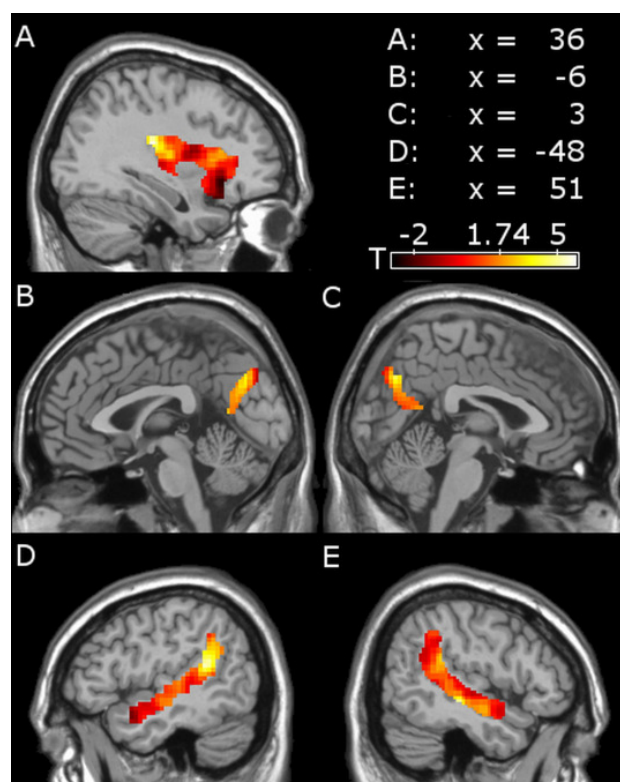
Figure 2. ROI-analyses; activation of contrast 'correct experimental vs. correct control'. Statistical parametric maps of ROI-masks are rendered onto a template single subject brain. Included are masks, in which a maximum with corresponding p<.1 corrected for familywise error has been found (see table 2). (A) right insula, (B) left POS, (C) right POS, (D) left STS, (E) right STS. Slices were chosen by x-coordinate of maxima. T-value of 1.74 is the threshold for uncorrected p<.05 at voxel-level.
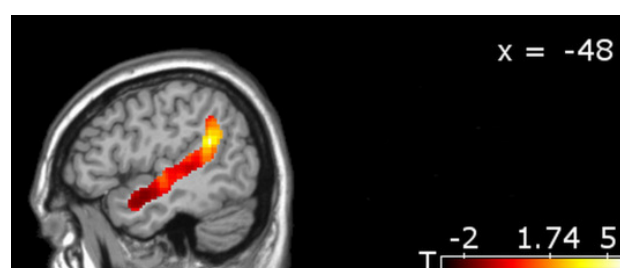

Figure 3. ROI-analyses; activation of contrast 'experimental vs. control'. Statistical parametric maps of ROI-masks are rendered onto a template single subject brain. Included is the mask, in which a maximum with corresponding p<.1 corrected for familywise error has been found (see table 3). Left STS; slice was chosen by x-coordinate of maxima. T-value of 1.74 is the threshold for uncorrected p<.05 at voxel-level.

fMRI-Data whole head analysis

No significant activation at a corrected level was found, when contrasting 'correct experimental vs. correct control', as well contrasting 'experimental vs. control'. Comparing both contrasts at an uncorrected level shows clearly differing activation patterns.

Table 3

Maxima of activations in fMRI-whole-head analysis of contrast 'correct experimental vs. correct control', height threshold T=3.65; p<0.001 (uncorrected); extent threshold k=9 voxels. No maximum reaches significance at a corrected level; included are first maxima in clusters.

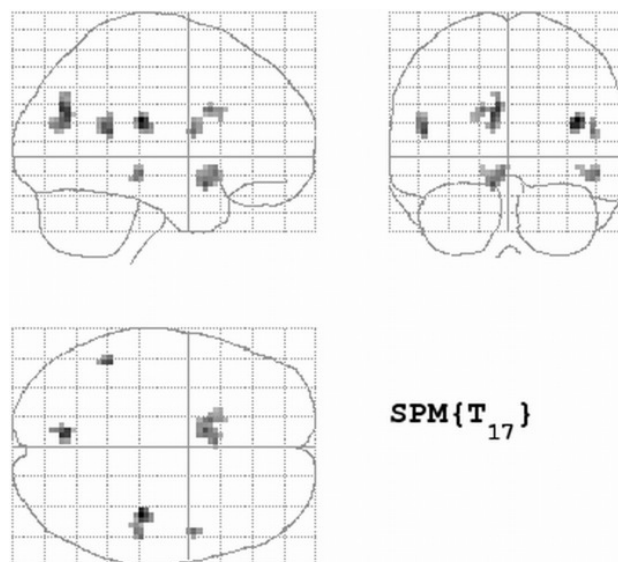| Side	Structure	MNI-coordinates			T
		x	y	z	
R	Rolandic operculum	39	-27	21	6.54
L	Parieto occipital sulcus	-6	-72	27	6.67
L	Superior temporal sulcus	-48	-45	15	5.35
L	Olfactory lobe	-9	9	-15	5.26
L	Caudate	-15	12	27	4.90
R	Superior temporal sulcus	51	-27	-9	4.83
R	Rolandic operculum	48	3	15	4.82

$SPM\{T_{17}\}$

Figure 4. Whole-head analysis; activation of contrast 'correct experimental vs. correct control'. T-value of 3.65 is threshold for uncorrected p<.001 at voxel-level. Included are clusters with a minimal size of k=9.

Table 4

Maxima of activations in fMRI-whole-head analysis of contrast 'experimental vs. control', height threshold T=3.65; p<0.001 (uncorrected); extent threshold k=9 voxels. No maximum reaches significance at a corrected level; included are first maxima in clusters.

| Side	Structure	MNI-coordinates			T
		x	y	z	
L	Cerebellum	-36	-69	-48	6.53
L	Cerebellum	-45	-69	-48	4.66
L	Cerebellum	-33	-60	-54	4.63
R	Parietal superior	33	-57	63	6.47
L	Caudate	-15	12	27	6.07
R	Caudate	21	-30	21	5.91
L	Middle frontal gyrus	-21	12	36	5.62
L	Inferior parietal gyrus	-36	-36	36	5.25
L	Caudate	-18	30	0	5.15
L	Cerebellum	-18	-81	-27	5.09
L	Cerebellum	-27	-78	-21	4.37
L	Cerebellum	-18	-87	-39	3.77
R	Thalamus	6	-6	21	4.49
R	Caudate	15	6	27	4.34
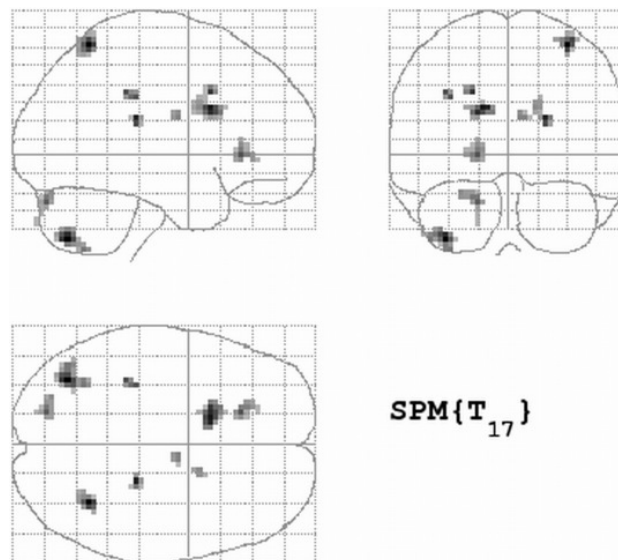
Figure 5. Whole-head analysis; activation of 'experimental vs. control'. T-value of 3.65 is threshold for uncorrected p<.001 at voxel-level. Included are clusters with a minimal size of k=9.

Discussion

We investigated audio-visual binding with fMRI by comparing brain activation evoked by synchronous presentations of audio-visual stimuli, where a ventriloquism effect indicated binding, with activation evoked by unbound asynchronously presented audio-visual stimuli. Thus, we focused on multisensory integration areas associated with consciousness-related binding. The ventriloquism paradigm was included in the experimental design to discriminate trials with crossmodal consciousness-related binding from trials without binding of audio-visual stimuli. If the subjectively perceived spatial localization of a sound is shifted by a simultaneously presented visual stimulus, the perceptual interaction indicates the emergence of a bound audio-visual object. This study-design represents a consciousness-related approach, because the trials of interest were selected by the occurrence of a phenomenal effect instead of on the basis of physical characteristics of the stimuli.

Behavioral data show that both the experimental condition (synchronously presented stimuli) and the control condition (asynchronously presented stimuli) induced perceptual phenomena as intended by the study. For synchronously and asynchronously presented stimuli the hit rate was above 90%. Subjects rated synchrony or asynchrony above chance at all presentation rates, the lowest observed hit rate was 88%. As indicated by subjects' responses, the ventriloquism effect was successfully induced in 71% of the experimental trials; in the control condition, 52% of sounds were localized correctly. Thus, a sufficient number of correct experimental and correct control trials remained for the statistical evaluation. Especially for asynchronously presented stimuli, the hit rate of sound localization was below our expectations. It has to be taken into account that the noise caused by the MR-scanner during EPI-scanning may have

increased the difficulty of the localization task compared to a sound proof lab environment usually available for crossmodal experiments. However, the only disadvantage of selecting trials based on the occurrence of the ventriloquism effect is that some trials are lost for analysis. Subjects were not equally successful in locating the sound as required by the task. Because of this variability, the Kappa coefficient was included in the analysis of group effects. Subjects showing a high number of correct responses more likely have perceived auditory, visual or bound audio-visual objects as intended by the experimental logic. Thus, the statistical contrast-images of those subjects must contain a higher amount of effect-variance. Controlling for the subjects' success in perceiving the ventriloquism effect as indicated by Kappa in the analysis of hemodynamic activity improves the detection of activation associated with consciousness-related binding by reducing experimental noise.

Correlates of crossmodal consciousness-related binding were investigated by comparing hemodynamic activation associated with correct experimental trials and correct control trials. Hemodynamic activation was investigated in predefined regions of interest (ROI). For synchronous audio-visual stimulation, where an ventriloquism effect occurred, activation was found in the upper part of the right insula, bilaterally in the upper part of the parieto-occipital sulcus (POS), the upper posterior part of the left superior temporal sulcus (STS) and the middle part of right STS. Activation in the left STS was also found, when experimental (synchronous) condition and control (asynchronous) condition were compared including all trials – without selecting trials with correct behavioral responses. The insula has been reported to be sensitive to congruency of stimulus onset between different modalities and unimodal binding processes. In this study, the insula responded to the bound synchronous stimuli, but no

activation was seen when the perceptual check data were discarded. This supports the view that the insula is involved in binding rather than responding to stimulus synchrony alone. The bilateral activation in the parieto-occipital sulcus (POS) may represent the spatial aspect of the task, if POS is processing spatial information of the audio and the visual modality. This is in correspondence with findings of mismatch negativity associated with both the ventriloquism effect and actual spatial sound shifts (Stekelenburg et al., 2004). Thus, illusionary ventriloquism effect and real sound displacement induce similar brain processes. The functional significance of POS and IPS was described to be associated with processing of spatial congruence (identity of location) of multisensory stimulation (Bushara et al., 2001; Fort et al., 2002; Molholm et al., 2002). However, we did not find activation in the IPS in our main analysis. With PET, activation was solely found in the right inferior parietal lobe, when visual and auditory speech stimuli were presented in opposite hemifields (Macaluso et al., 2004). Since IPS is not specific for speech, the IPS activation was expected in our study. The comparison of correct and incorrect control trials revealed activation in the insula and the IPS. Trials of the control condition were classified as incorrect, if audio-visual asynchrony was not recognized or if localization of the sound was not in accord with the sound source. Hence, this activation in the IPS may reflect sound localization processes.

In the middle part of the right STS activation was found, when the binding check at a phenomenal level via the ventriloquism effect was included in analysis. This is in correspondence with findings relating STS activation to object identification (e.g. Beauchamp et al., 2004). The activation in right STS was not seen in the analysis of conditions defined by physical characteristics. Without using a perceptual check for binding, Macaluso et al. (2004) found activation in left STS only, when synchronous

audiovisual speech was analyzed irrespective of the relative location of the stimuli. This provides evidence that the multisensory integration process in the STS is associated with the perception of a multisensory bound object.

The comparison of correct with incorrect experimental trials appears to indicate activation specific to the localization shift of the ventriloquism effect, because in correct and incorrect trials the stimuli consisted of identical physical characteristics. However, contrary to correct responses, no model explains the origin of incorrect responses. Confusion about sound localization or erroneous button pressing could be sources of incorrect responses. The number of experimental trials (29%) incorrectly responded to in localization was less than half as big as the number of correct experimental trials (71%). The lack of an appropriate model for incorrect responses is one reason why contrasting incorrect experimental trials with correct control trials did not show activation in ROIs. No activation was seen in the superior colliculi (SC) in any comparison. However, SC-activation was not necessarily expected, because the study focused on consciousness-related cortical processes. Both conditions implemented in the study were bimodal. Differences in activation between synchronous and asynchronous stimulation have been reported only in some neurons of SC (Stein & Meredith, 1993). SC could have been activated in both conditions, or the difference was too small to detect with fMRI. SC is small in size and is vulnerable to movement artefacts, which makes signal detection more difficult.

In previous studies, the areas selected for ROI-analyses have been thought to be involved in multisensory integration processes. The contrast of correct experimental trials versus correct control trials and the contrast of experimental versus control trials were the main comparisons in this study. Whole-head analyses of these two contrasts

were conducted to see if the regions of interest cover all task-relevant effects, or if different brain areas should be considered in further research. The whole-head analysis of correct experimental trials versus correct control trials at the level of uncorrected tests indicated the same regions, found to be significant in the ROI analysis with the exception of the caudate areas that were not subjected to a ROI analysis. The statistical parametric maps of the whole-head analysis of the comparison synchronous versus asynchronous (i.e. all trials of the experimental condition versus all of the control condition) does not correspond to the results of ROI-analyses. This indicates that an analysis aimed at isolating processes of consciousness-related binding benefits from utilizing the ventriloquism effect. Omitting the ventriloquism effect for analysis introduces hemodynamic activity into brain activation maps, the sources of which remain unclear in the context of the experiment.

Based on the results of this study, the major conclusion is that utilizing the ventriloquism effect as a perceptual indicator of consciousness-related binding reveals a pattern of areas involved in the processing of a bound percept. This consciousness-related process probably is different from stimulus-related multisensory integration mechanisms, especially those in subcortical areas, e.g. the SC. The pattern of activation can not be seen if only physical characteristics of stimuli are used to define conditions. The psychological characteristic of 'Gestalt' is critical for investigating the perception of binding. Further work should concentrate on the investigation whether activation associated with audio-visual binding as described in the current study can also be identified with other combinations of modalities (e.g. tactile-visual processing) and thus be attributed to general characteristics of crossmodal consciousness-related binding.

Chapter 3

Indicators of Audiovisual Binding in Low-Frequency EEG

Crossmodal binding is the merging of feature information of different sensory streams into the unified representation of an object in perception. The present study investigated correlates of audiovisual binding in the electroencephalogram (EEG). Participants had to report the synchrony and location of auditory stimuli when observing both visual and auditory stimuli that were presented either synchronously or asynchronously. The ventriloquism effect was used as an indicator of audiovisual binding. This multisensory perception effect describes an auditory localization shift toward synchronously presented visual stimuli. It recently has been shown that the ventriloquism effect can be used to indicate the integration of crossmodal stimuli on a behavioral level. EEG recording was accompanied by simultaneous functional magnetic resonance imaging (fMRI). Results showed characteristic differences in EEG synchronization between synchronously and asynchronously presented stimuli. Quantified on a single-trial level these differences distinguished successfully between binding and no-binding trials. Our data suggest that crossmodal binding is associated with changes in the lower frequency EEG at parietal electrodes.

Introduction

One major question in neuroscience is how features of an event are integrated across modalities into one coherent representation in mind. Crossmodal binding can be considered as a subcategory of multisensory integration. Whereas multisensory integration comprises all processes by which crossmodal stimuli are converged (Stein & Stanford, 2008), binding specifically assigns the stimuli to a unified event (Revonsuo, 1999).

Simultaneous stimulation in multiple modalities is associated with both electrophysiological and hemodynamic processes (see, for single-cell measurements, Stein & Meredith, 1993; for EEG in response to audiovisual stimulation, Basar, 1999; Basar et al., 2001; for fMRI in response to audiovisual stimulation, Bushara et al., 2001; Calvert et al., 2001). Hemodynamic correlates of binding have been found in different brain areas, depending on which senses were stimulated and which parameter was manipulated in the specific crossmodal task (i.e., spatial, temporal, or semantic content; see, for review, Calvert & Thesen, 2004; see, also, Teder-Salejarvi, Di Russo, McDonald, & Hillyard, 2005). Synchronization of cell assemblies in the gamma frequency range found during feature integration in the visual cortex of the cat (Gray et al., 1989; Singer & Gray, 1995) indicates that dynamic aspects of brain activity reflect the underlying mechanism of binding (Singer, 1998). Similar to changes of EEG dynamic within the visual cortex with unimodal binding, EEG coherence within brain areas and across brain areas as well as combinations of intra-area and inter-area coherence have proved to be associated with crossmodal binding (for a discussion, see Engel & Singer, 2001, for a review, see Senkowski et al., 2008). The frequency range of observed oscillations corresponds with the distance of areas involved in a task (von

Stein, Rappelsberger, Sarnthein, & Petsche, 1999; von Stein & Sarnthein, 2000). Sakowitz, Quirogaa, Schürmann, and Basar (2005) found multisensory-specific components in all frequency bands. The most pronounced component was an early-appearing (< 100 ms) theta rhythm over fronto-centro-parietal sites. Increases in theta power have also been reported as being evoked by audiovisual stimulation (Basar, 1999; Basar et al., 2001). Noesselt et al. (2007) and Bonath et al. (2007) investigated the neural basis of the ventriloquism illusion with fMRI and source localization of separately measured ERP differences between audio-visual and unimodal stimulation. Dipolar sources were found in the auditory cortex of the planum temporale. Relatively greater activation was found in the auditory cortex contralateral to the subjectively shifted sound position in fMRI (Bonath et al., 2007). Analyzing effective connectivity of fMRI activations, effects of congruency (audiovisual temporal correspondence) were found in the contralateral STS and in primary visual and auditory areas (Noesselt et al., 2007).

An explanation for crossmodal effects is provided with the model of Maximum Likelihood Estimation of multisensory integration by Ernst and Banks (2002). The precision of stimuli in different modalities is estimated constantly for each modality, and each modality's contribution to the merged percept is weighted. For example when localization is relevant, a stimulus from one modality offering more precise spatial information dominates a less spatially precise stimulus from another modality. Therefore, presenting an auditory and a visual stimulus in proximity may result in a localization shift away from the auditory stimulus towards the visual stimulus, because the visual stimulus offers higher spatial precision. This phenomenon has been named the ventriloquism effect (Howard & Templeton, 1966). It serves as an indicator for the integration of crossmodal information in the present study. If spatial precision is

manipulated by blurring the visual stimulus, the ventriloquism effect can be reversed (Alais & Burr, 2004).

In the present study, we investigated neuronal correlates of crossmodal binding using simultaneously measured EEG and fMRI. Instead of defining 'binding trials' by stimulus configuration or experimental design, the ventriloquism effect was used as a behavioral indicator of crossmodal binding. The objective of this study was to identify a parameter in the EEG signal indicating audiovisual binding that could be validated by fMRI data for a better spatial localization.

Methods

Participants

EEG data were analyzed in 19 healthy participants (12 female; mean age = 26.19, SD = 5.0). All gave written informed consent for their participation before commencing the study.

Stimulation

Visual stimuli were two vertically aligned gray discs, one at the left-most the other in the right most position on a black computer screen. There was also a gray fixation cross at the center of the screen. During each trial (duration = 8 s), the discs flashed for 200 ms at one of three different rates: 6, 8, or 10 times in 8 s (equivalent to 0.75 Hz, 1.0 Hz, and 1.25 Hz). In the "congruence" condition, the discs appeared in alternation on the left and on the right with an onset lag between the left and the right disc of 560, 400, or 320 milliseconds. In the "incongruity" condition, both discs appeared synchronously.

Auditory stimuli were sinusoidal tones (440 Hz, duration 120 ms) presented via headphones at the same rates as the visual stimuli. In the congruence condition, the tone always appeared simultaneously with one of the discs, i.e. on one side (left/right). Each

auditory stimulus began with an onset lag of 40 ms after the onset of that visual stimulus to which it was linked to. In the incongruity condition, the auditory stimuli were lagged to the visual stimuli on both sides (710, 640, or 540 ms). Stereo balance and gain were calibrated separately for each participant during presentation of a series of test sounds.

Task

After the stimulation period, participants indicated which of the visual stimuli had been presented synchronously to the auditory stimuli: those on the left or those on the right side. Subjects indicated on a five-point scale to whether the sound came from far left, left, center, right or far right. Each response had to be confirmed with an extra button press within 4 seconds. Main effects and interaction effects on performance were analyzed with a 2 (condition) x 3 (repetition frequency) ANOVA. The fixation cross remained visible during all inter-trial intervals (duration 21 +/- 1.5 s). Every experimental run consisted of 24 'binding' and 24 'no-binding' trials in pseudo randomized order, totaling to 192 synchronous and 192 asynchronous audiovisual events. Stimulus presentation and response acquisition were controlled by Presentation software (Neurobehavioral Systems, Albany, CA). Participants performed up to 10 practice trials prior to the experimental session.

EEG-Data recording

EEG data were recorded from 29 sintered Ag/AgCl ring electrodes (10-20 system, plus FC1, FC2, CP1, CP2, FC5, FC6, CP5, CP6, TP9, TP10) mounted on an electrode cap (BrainCap, Falk-Minow Services, Germany) with an additional reference electrode between Fz and Cz. Input impedance was strictly kept below 5 kOhm. Two additional channels were used to record the electrocardiogram, a further channel was used for recording of the vertical electrooculogram. Signals were recorded with a BrainAmp-MR

32-channel amplifier (BrainProducts, Munich, Germany) using a sampling rate of 5 kHz, filters were set to 0.016-250 Hz . Trigger signals for stimulation and responses were aquired along with EEG, ECG, and CEOG using BrainVision Recorder software (BrainProducts, Munich, Germany). All recordings were performed inside a 1.5 Tesla whole-body MR-scanner (Siemens Symphony) during functional EPI-scanning.

EEG preprocessing and parameter extraction

MR-gradient artifacts and cardioballistic artifacts were corrected according to Allen et al.'s (2000; Allen, Polizzi, Krakow, Fish, & Lemieux, 1998) template methods with postcorrection downsampling to 250 Hz and low-pass filtering at 30 Hz. Eye-blink artifacts and eye-movement artifacts were controlled by visual inspection after performing an independent component analysis (infomax algorithm, in BrainVision Analyzer, Brainproducts, Munich, Germany). All EEG channels were re-referenced to their common average and sampled down further to 125 Hz. Digital passband filters between 3.5 Hz / 24 dB and 14.0 Hz / 24 dB were applied. Time-frequency representations of segmented epochs across frequencies were acquired in increments of 0.125 Hz through continuous wavelet transformation (CWT) using the Morlet mother wavelet. In order to detect single-trial spectral changes, event-related synchronization/synchronization (ERD/ERS) was computed for each wavelet transform, with the 500 ms before stimulation onset being used as the reference phase. The ERD/ERS representations of two groups of electrodes were pooled by averaging. The first group consisted of electrodes P3, P7, CP1, and CP5; the second, P4, P8, CP2, and CP6. Accordingly to the acquisition time of one single EPI-volume EEG-segments covered the time span between 0 ms and 3000 ms.

For each condition, the grand average of the ERD/ERS spectrograms was computed to represent the "typical" activation. Only congruence trials in which the ventriloquism

effect occurred were included. In the incongruity condition, sounds had to be localized without a shift in order to be included in further analyses. The difference between 'binding' and 'no-binding' grand averages are suggested to represent the effect of the crossmodal binding process resulting in coherent perception of that event. The spectrograms of the differences were smoothed over time using a Gaussian filter (FWHM = 1.248 s). The maximum was set to 1 for creation of a prototypical difference spectrogram.

In a next step, this difference spectrogram was used to quantify the binding-related EEG activity for each trial. For each group of electrodes, the single-trial ERD/ERS spectrograms were multiplied with the difference spectrogram. All amplitudes of the resulting spectrogram were summed up to the final parameter value. It should be noted that energy in single-trial spectrograms corresponding to negative values in the difference spectrogram lowered the overall sum. Therefore, a deviant activation pattern with high power in all frequencies did not result in a high parameter value. All trials were included regardless of the accuracy of reported answers. The mean difference for all binding and all incongruity trials was tested for statistical significance using a *t*-test. Alpha threshold was set at $p < .05$.

Results

Behavioral responses
Behavioral responses were analyzed on the level of means over all 19 participants. In the congruence condition, audiovisual stimuli were detected correctly as synchronous in 92% of trials (22.05 out of 24 trials, $SD = 1.72$). Stimuli were correctly reported to be asynchronous in the incongruity condition in 95% of trials (22.84 out of 24 trials; $SD = 1.71$). The 2 (conditions) x 3 (repetition frequencies) ANOVA revealed no main effect, but a significant Frequency x Condition interaction, $F_{(2, 36)} = 4.15$, $p = .024$,

Mauchly's test for sphericity $p = .468$). A statistically significant difference ($p < .001$) was found for stimuli presented with a frequency of 0.75Hz. When presented at this frequency, synchrony or asynchrony was detected correctly in 88% of trails (7.05 out of 8 trials, $SD = 1.08$) in the congruence condition and in 98% of trials (7.84 out of 8, $SD = 0.50$) in the incongruity condition.

In the synchronous condition, sounds were localized without a shift in 29% of trials (6.89 out of 24 trials, $SD = 2.87$). In the asynchronous condition, 53% of sounds were correctly localized (12.68 out of 24 trials, $SD = 4.80$). The main effect of condition was significant, $F(1, 18) = 17.56$, $p < .001$). Neither an effect of repetition frequency nor an interaction effect was found.

EEG analysis

The EEG signal associated with binding was parameterized for single-trials in three steps. The grand averages of time-frequency representation were the basis for the spectrograms of differences between 'congruence' and 'incongruity'. The time-frequency representations of the differences were applied to all single-trial spectrograms to extract the sum of power changes over time, which is the parameter.

For both groups of electrodes (left, right) and for all conditions, the grand averages of ERD/ERS there was evidence for increased synchronization in the theta band (see figure 1). In relation to the reference phase, the maximum gain of synchronization was 0.579; the maximum of desynchronization, - 0.120.
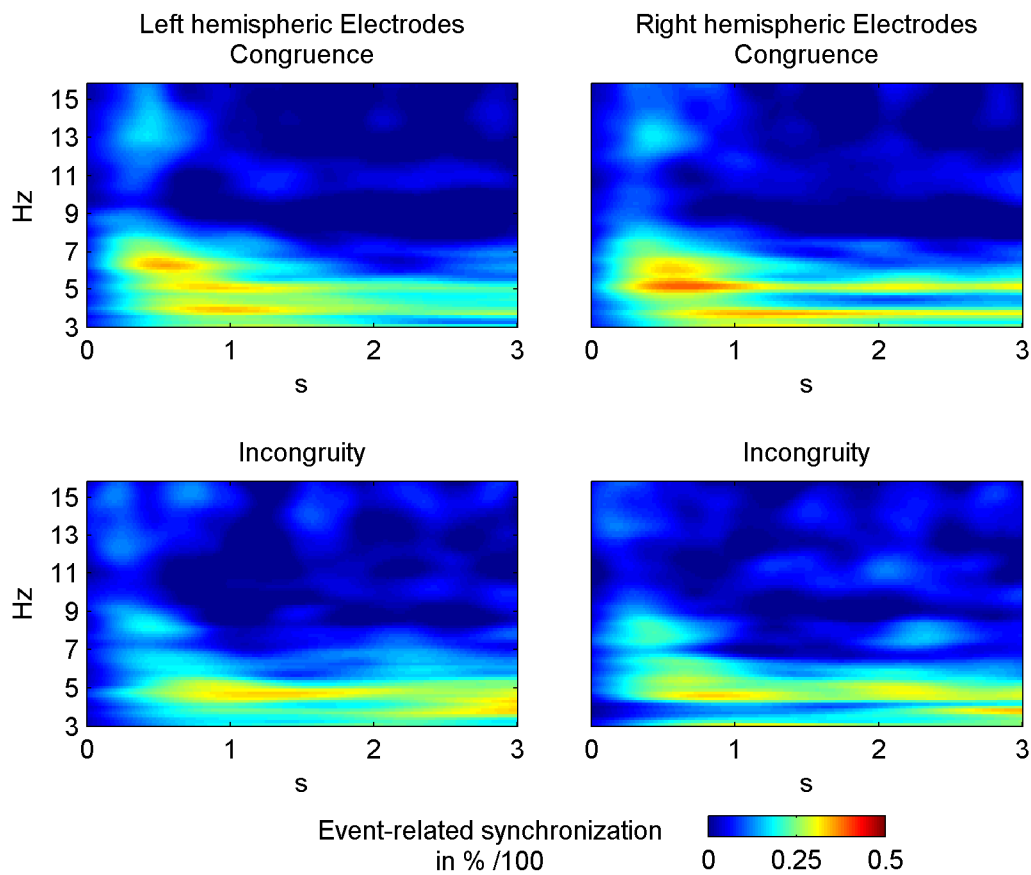
Figure 1. Grand average time-frequency representations of synchronization change (ERD/ERS, 500 ms baseline or pre-stimulus interval) for the conditions 'congruence' and 'incongruity', relating to the 500 ms pre-stimulus interval. Signals were pooled for two groups of electrodes; one left-hemisphere group at positions P3, P7, CP1, and CP5; one right-hemisphere group at positions P4, P8, CP2, and CP6.

After subtracting the grand average of the 'incongruity' condition from the 'congruence' condition the differences were most pronounced between 4 and 7 Hz. They ranged from - 0.259 to 0.213 in the left-hemisphere (LH) electrodes and from - 0.364 to 0.245 in the right-hemisphere (RH) electrodes before smoothing (figure 2).
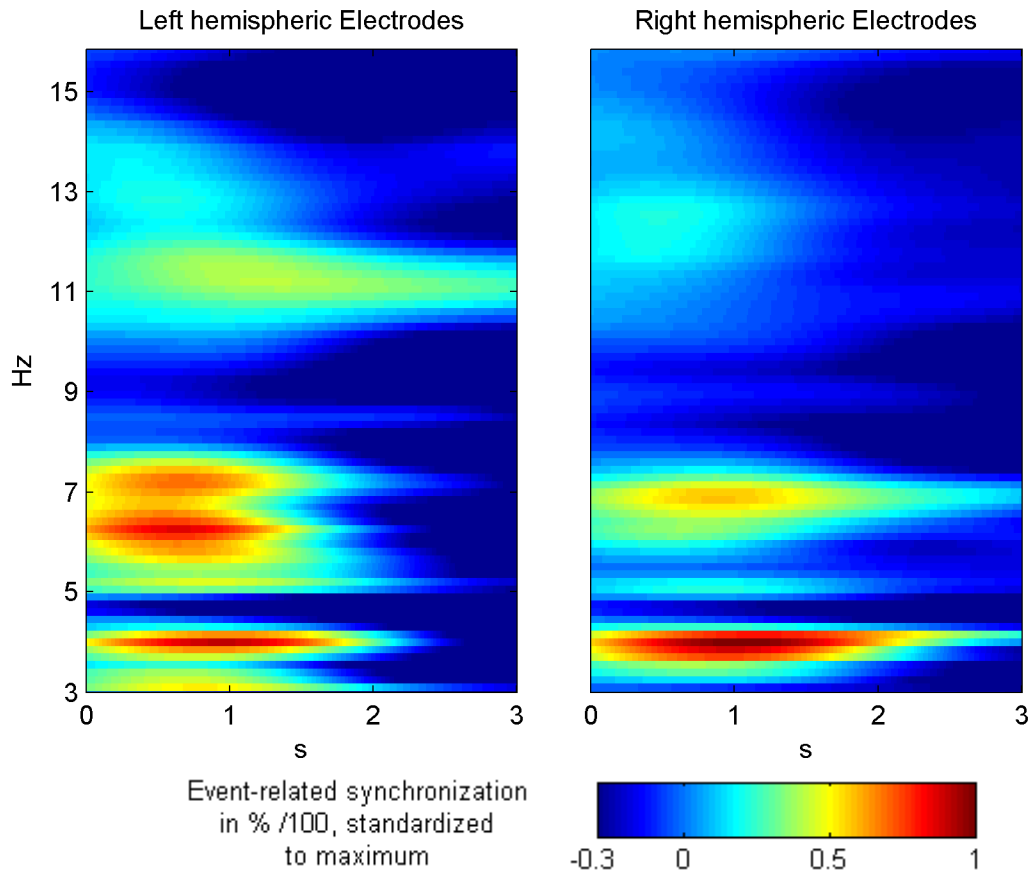
Figure 2. Time-frequency representations of the differences in changes during crossmodal binding ('congruence' minus 'incongruity'; max set to 1). Highest difference corresponds to ~23% higher synchrony.

Parameter values were extracted by multiplying the spectrograms of single trials with the prototypical difference spectrograms, resulting in one single value for the LH group of electrodes and one value for the RH group of electrodes for each trial. Pairwise *t* tests comparing binding and no-binding resulted in statistically significant differences for both pools (LH group of electrodes P3, P7, CP1, CP5: $M = 188.41$, $SD = 339.10$, $t[18] = 2.42$, $p = .026$; RH group of electrodes P4, P8, CP2, CP6: $M = 219.93$, $SD = 192.29$, $t[18] = 4.99$, $p < .001$.

Discussion

The present study is designed to extract correlates of crossmodal binding from lower frequencies of the EEG recorded inside an MR scanner during EPI-scanning. The subjects were asked to localize simple audiovisual events, which were presented synchronously and asynchronously. The ventriloquism effect, a shift in the subjective localization of synchronous audiovisual stimuli, is used as a behavioral indicator of the binding process. Event-related changes in low-frequency EEG synchrony are expected to indicate crossmodal binding.

Grand-average EEG spectrograms are computed for frequencies of 3.5 to 14Hz with a wavelet-analysis. Activity is compared with those of the prestimulus phase because we are interested rather in changes in synchronization than in absolute values. Using this procedure, signal changes over time can be observed in a broad frequency range that is advantageous to spectral analysis. The prototypical pattern of binding is specified as the difference between the averaged spectrogram of the binding and no-binding trials. This pattern during crossmodal binding is similar for two groups of electrodes (LH and RH parietal positions), showing higher energy most prominently at about 4 and 7 Hz. Used as a filter, such a prototype pattern affords us to quantify single trials (not single events) in terms of total energy associated with binding. Thereby the process investigated can be searched for in any trial regardless of the condition or the behavioral response, e.g. the quantified energy can serve as a modulator in a combined fMRI-EEG-analysis. Congruence trials show higher energy than incongruity trials in the EEG in the parietal and centroparietal electrodes.

In this study, the synchronously presented stimuli successfully induce the ventriloquism effect despite noise and interference in the headphones due to the magnetic field and HF-pulse switching of the MR scanner. With simultaneous recording

of EEG and fMRI, dynamic aspects of brain activity can be related to metabolically associated local signals. The simultaneous recording demands sophisticated techniques, because the artifacts induced by the magnetic field and especially the reading pulses of fMRI exceed the EEG signal by more than two orders of magnitude (Allen et al., 2000). Interleaved fMRI recording does not solve this problem conclusively and discards the important EEG information during image acquisition. In addition, the hemodynamic information during EEG recording is lost.

Most studies on multisensory integration have focused on synchronization in the gamma band (for a review, see Senkowski et al., 2008). A reliable extraction of gamma EEG from signals recorded during EPI scanning is not possible with current methods of MR-artifact correction (Mandelkow et al., 2006). However, von Stein and Sarnthein (2000) have distinguished inter area interactions in the theta and alpha band from integration processes of local signal propagation that are reflected in higher frequencies. Assuming that percept-related crossmodal binding is a spatially distributed process, the spectrogram pattern found in the present study is consistent with such a relation of oscillation frequency and length of connections; that is, the interaction between areas is also reflected in lower frequencies. Whereas it is suggested that gamma frequencies (30 - 100 Hz) reflect intra-area binding processes, responses to multimodal stimulation are found in the theta band range (3.5 - 7.5 Hz) as well. In evoked audiovisual potentials, the most pronounced multisensory component has been found in the theta band (Mishra et al., 2007).

Analysis of fMRI data revealed that congruence trials relate to network activity including the insula, the superior temporal sulcus, and the parieto-occipital sulcus (Bischoff, Walter, Blecker, Morgen, Vaitl, & Sammer, 2007). In the current study, the crossmodal binding as indicated by lower frequencies in the EEG is very likely

generated by synchronized neuronal activity within such a network. The relation of the hemodynamic response of fMRI and single-neuron activity has been called "an area of open inquiry" (Stein & Stanford, 2008, p. 264), and this is also true for the relation between fMRI and EEG signals during multisensory integration. Future analyses should use the quantified EEG energy to compare the electrophysiological signal to the hemodynamic signal on a trial-by-trial basis. However, when doing this, it is crucial for the parameter to differentiate sufficiently between conditions. The quantification of single trials by parameter extraction shows that congruence trials evoke more energy in the prototypical pattern of binding than no-binding trials. This supports the notion that the correlate found here actually does reflect the specific processes, while taking into account that the energy sum is lowered by processes not matching the binding pattern. An overall increase in synchronization in one condition does not result in higher energy values. Therefore, the parameterized EEG can be interpreted as an estimator of the crossmodal binding process, and it may be used to guide EEG-constrained fMRI analyses.

The difference spectrogram and the quantified energy successfully identify a crossmodal binding correlate. For the prototype, trials in which binding is perceived are selected, the quantification is extended to incorrect trials as well. The changes in the theta band are related to the percept of a synchronous audiovisual event. This is in line with the proposed role of theta in the association of memory elements (Schürmann & Basar, 1999) and in the association of the time information relative to the situational context, i.e. the sequence of events (Klimesch, 1999). In conclusion, percept-related crossmodal binding involves low-frequency cortical network interactions, as indicated by enhanced theta synchronization in EEG spectrograms.

Chapter 4

EEG-Guided fMRI-Analysis Reveals Involvement of Low-Frequency EEG and the Superior Temporal Sulcus in Audiovisual Binding

The problem of crossmodal binding is how the feature information of different sensory streams is translated into the unified representation of an object in perception. The present study investigated correlates of audiovisual binding using an analysis of functional magnet resonance imaging (fMRI) guided by the electroencephalogram (EEG). EEG and fMRI were recorded simultaneously. This allowed the investigation of the relationship of classical multisensory integration areas and task related EEG. Binding conditions were introduced by synchronously presented audiovisual stimuli of low complexity. Participants were asked to report the temporal congruency and location of auditory stimuli while observing both visual and auditory stimuli that were switched on/off either synchronously or asynchronously. During temporally congruent stimulus presentation very frequently an auditory localization shift toward the synchronously presented visual stimuli results can be observed. This ventriloquism effect (VE) indicates the successful perceptual integration of the audio-visual event. Without this effect, it is not sure that crossmodal binding occurred in that act of perception.

In the EEG, time-frequency analysis of theta and alpha band activity at parietal electrodes was different for temporal congruency conditions. A weighted total energy measure of single trial EEG spectrograms was introduced as modulator in the EEG-guided fMRI analysis of binding. Thus, binding related variations in EEG single trial energy measures showed correlation with hemodynamic changes in the right superior temporal sulcus for audiovisual binding. This activation is not solely allegeable by the

temporal congruency of stimuli or the ventriloquist illusion as equivalent analyses showed. Concluding, activation in the superior temporal sulcus and low-frequency neuronal synchronization are correlated during the perceptual binding of audio-visual events.

Introduction

In neuroscience the binding problem implies two questions. The first is how the features of an object or event are integrated correctly in the multitude of events continuously perceived. The second question is how the features of this object or event are translated into a coherent conscious perception. Initially, the binding problem was discussed in the field of feature integration in visual object recognition (for a review, see Singer, 1999). The discovery of the dorsal and ventral pathway of visual processing (Mishkin & Ungerleider, 1982) raised the question how information of both streams is reunited. Similarly, the binding problem can be extended to feature integration in multisensory or multimodal perception. The ventral ('what') and dorsal ('where') pathways are known for the auditory, somatosensory, and visual modality (Schroeder & Foxe, 2002). However, not much is known about how multisensory binding works on a perceptual level. The term "crossmodal" binding describes an interaction of involved modalities being more than simply to the number of modalities, i.e. "bi-", "tri-" or "multimodal" stimulation. Binding is related to the conscious representation of a coherent entity (Revonsuo, 1999). In contrast, multisensory integration describes all biological and mental convergent processing of multimodal stimuli, which does not necessarily involve conscious perception (Stein & Stanford, 2008). In most studies crossmodality is assumed with respect to congruent multimodal stimulation. However, it appears to be advantageous to verify crossmodal interaction and perceptual binding by the use of illusions and perceptual shift effects. According to the Maximum Likelihood Estimation model of multisensory integration (Ernst & Banks, 2002), the spatial precision of stimuli in different modalities is estimated constantly for each modality, and each modality's contribution to the percept is weighted accordingly. For

example, a stimulus from one modality that offers better spatial information dominates the localization of a bimodal stimulus, which has weaker spatial resolution in the second modality. The present study uses a crossmodal spatial effect, the ventriloquism effect. The ventriloquism effect describes the mislocalization of an auditory stimulus towards a simultaneously presented visual stimulus (e.g., Howard & Templeton, 1966). Interestingly, if spatial precision of a visual stimulus is manipulated by blurring, the ventriloquism effect can be reversed (Alais & Burr, 2004). This finding strongly supports the model of Maximum Likelihood Estimation. The integration of sensory information leads to a percept shifted in space. Stimuli are not perceived as conflicting or separate, nor are the spatial information totally neglected in one modality. Therefore, the ventriloquism effect does not represent a absolute dominance based on one of the modalities, but rather an interaction while the percept is assembled.

Activity of brain during multisensory task performance found in neuroimaging studies depends on the modalities involved, the experimental parameter, the paradigm and the analytic approaches used (Calvert, 2001; Calvert & Thesen, 2004; Laurienti et al., 2005). Primary sensory areas have been assumed to be purely unisensory for a long time. However, recent studies show that multisensory information is processed in primary sensory areas as well (Amedi et al., 2005; Ghazanfar & Schroeder, 2006). The interaction of primary sensor areas with higher-level convergence areas are currently under discussion (Bauer, 2008; Senkowski et al., 2008). Although a number of heteromodal cortical integration areas were studied in several neuroimaging studies, their exact function for crossmodal integration (and binding) not yet clear. The superior temporal sulcus (STS) is involved in audiovisual integration of speech and in object identification (Calvert et al., 2000; Beauchamp et al., 2004; Amedi et al., 2005). The superior colliculus (SC) and the insula/claustrum are activated by simultaneous onsets

of stimuli across different modalities (Bushara et al., 2001; Calvert, et al., 2001). The intraparietal sulcus (IPS) and the parieto-occipital sulcus (POS) are involved in the detection of multisensory stimuli that correspond in spatial location or visuo-motor aspects (Bremmer et al., 2001; Bushara et al., 2001; Callan et al. 2001; Calvert et al,. 2001; Fort et al., 2002; Molholm et al., 2002; Macaluso et al., 2004; Amedi et al., 2005).

The main characteristic of multisensory integration on the neuronal level is a difference in the neuronal firing rate evoked by a bi- or multimodal stimulus-combination in comparison to the most effective of these stimuli individually (Stein & Stanford, 2008). This difference may result in a significant multisensory enhancement or suppression. It has been proposed, that multisensory integration is related to dynamical aspects of brain activity beside the firing rate of neurons (Singer, 1998, for a review, see Senkowski et al., 2008). The temporal correlation hypothesis states that the synchronization of distributed neural signals in networks of brain areas is a key mechanism for the assignment of information and multisensory integration (Singer & Gray, 1995; for a review, see Senkowski et al., 2008). Concurrent activity in areas not adjacent or close to one another is observed as oscillatory responses in the EEG (Milner, 1974; von der Malsburg, 1981). The various frequency bands of such synchronized oscillations were related to a variety of cognitive and sensorimotor processes; most studies on multisensory integration have focused on gamma band synchronization (for a review, see Senkowski et al., 2008). However, concurrent stimulation in two modalities also evokes a power increase in the theta frequency range (Basar, 1999; Basar et al., 2001). Schürmann and Basar (1999) report theta responses to audiovisual stimuli predominantly over the association areas of the frontal midline and the temporo-parietal locations. Sakowitz et al. (2005) have found multisensory-specific components in all

frequency bands. The most pronounced one was early appearing theta synchronization in frontocentral and parietal sites.

The hypothesis of crossmodal integration by hierarchically organized dedicated areas and the hypothesis of crossmodal integration by temporal coupling of areas forming temporarily a functional network do not exclude each other. For instance, sets of transmodal nodes acting as epicenters for convergence across areas have been proposed already by Mesulam (1998).

In a previous study, we investigated hemodynamic activation associated with binding using the ventriloquism effect. Comparing binding with asynchronous audiovisual stimulation with region of interest (ROI) analyses we found higher activation in the right insula, and bilaterally in the STS and POS (Bischoff et al., 2007). Similarly, analyzing the differences in ERD/ERS between binding and asynchronous stimulation, we found higher energy for crossmodal binding most prominently in the theta band. (Bischoff et al., submitted).

In the present work we study the association between activated areas as indicated by fMRI and EEG-synchronization during crossmodal binding. EEG was measured simultaneously during fMRI (EPI-acquisition). Simultaneous measurement allowed for direct comparison of the results of both methods, because they were recorded at the same time during identical experimental conditions. Event-related changes in low-frequency EEG synchrony (ERD/ERS) indicated crossmodal binding and were parameterized to be correlated with the fMRI.

Methods

Subjects

Twenty-five subjects were recruited for the study. Because of a complete absence of the ventriloquism effect, data of 5 subjects were excluded. Data of one subject was

excluded for technical reasons. Thus, data of 19 subjects remained for analysis (12 female; mean age =26,2; S.D.=5,0).

All subjects were naive as to the purpose of the experiment. None of them had a history of or a current neurological or psychiatric illness. All reported normal or corrected-to-normal vision (MR-suitable glasses) as well as normal hearing. All subjects provided written informed consent for their participation.

Visual stimuli

Visual stimuli were gray discs (RGB 200,200,200) which could appear on two vertically aligned positions on the edge of the projected screen (XGA, 1024*768). The background was black with a gray fixation cross at the center. Each disc had a diameter of 240 pixels. During a trial (duration = 8 s), the discs flashed repeatedly at one of three different rates for each trial; 6 times in 8 seconds, 8 times in 8 seconds, and 10 times in 8 seconds (0.75 Hz, 1.0 Hz, 1.25 Hz). For each time, the discs appeared for 200 ms. During the temporally congruent stimulation, the discs appeared alternating on the left and on the right with an onset lag between the left and the right disc (560, 400, or 320 ms). During the temporally incongruent stimulation, they appeared simultaneously on both sides. The visual stimuli were displayed on a back projection screen behind the scanner. Via a dual-mirror, which was attached to the head-coil, subjects could watch the screen. The angle of vision was approximately 9° vertically and 12° horizontally. The projector was located outside the scanner room; a special lense afforded the projection through a waveguide.

Auditory stimuli

Auditory stimuli were sinusoidal tones with a frequency of 440 Hz and a duration of 120 ms. The tones were presented at the same rates as the visual stimuli (0.75, 1.00, 1.25 Hz). During the presentation of temporally congruent stimuli, the auditory stimuli

were coupled with the visual stimuli on one side, e.g. the sounds always appeared simultaneously with the discs on the right side. The auditory stimuli started with an onset lag of 40 ms after the onset of the respective visual stimuli and ended 40ms before the visual stimuli did. During the presentation of temporally incongruent stimuli the visual stimuli were presented simultaneously, the auditory stimulus was given using a long onset lag (510, 440, 340 ms).

The auditory stimuli were presented using MR-compatible electrodynamic high quality sound stereo headphones. The permanent magnet in these headphones is substituted by the static B0-field of the MR-scanner. This allows the electrodynamic headphones to replace low fidelity standard air headphones. However a disadvantage is that the gain level of the channels of the electrodynamic headphones partly depends on the position of the earphone in the scanner. Therefore, the headphones were attached to the head fixation clamps which in turn are part of the patient table. This configuration assured both a fixed placement in the MR-scanner and a good fit of the headphones to the subject's head. For each subject, stereo-balance and gain were calibrated during presentation of a series of test sounds. The balance of the headphones was individually adjusted until the participants judged the sound as originating from the central point. To match the acoustic noise conditions in the subsequent task, these adjustments were made during an fMRI-sequence.

Experimental task
During each trial, the stimuli were presented repeatedly for 8 seconds. In trials with temporally congruent audiovisual stimuli, the visual stimuli on one side were synchronously presented with the auditory stimuli. Note that the disc-stimuli were presented asynchronously with regard to each other. In trials with temporally

incongruent audiovisual stimuli, both disc-stimuli were presented synchronously with regard to each other but asynchronously to the auditory stimuli.

After the stimulation period, participants indicated which of the visual stimuli had been presented synchronously to the auditory stimuli: those on the left, none, those on both sides, or those on the right side. Subjects had to check one within a line of four boxes (by moving a cursor with two buttons) and confirm the choice pressing a third button. The answer boxes not chosen were dimmed out to show that the response was registered. The interval for this response was fixed to 4 seconds. Second, they used a five-point scale to indicate whether the sound had originated from far left, left, center, right or far right. Again, the response had to be confirmed with an extra button press within a 4-s interval. All responses were performed with the index finger of the right hand. During the time between two trials (21+/- 1.5 s) only the fixation cross remained visible. Onsets were jittered in relation to the BOLD-measurement stimulus.

Every experimental run consisted of 48 trials. The stimulation was congruent in 24 trials (synchronous audiovisual stimuli) and incongruent (asynchronous audio and visual stimuli) in 24 trials, equaling 192 synchronous and 192 asynchronous stimuli pairs. The presentation order was pseudorandomized. Stimulus presentation and response acquisition were controlled by 'Presentation'-software (Neurobehavioral Systems, Albany). To become familiar with the keypad and to ensure that the instructions were understood correctly, subjects practiced up to 10 sample trials of the task prior to the session in the scanner.

Trial categorization / conditions

The temporal congruency of two streams of audiovisual stimuli was manipulated. If the stimuli were temporally congruent some of a stream of visual stimuli appeared coincidently with a stream of auditory stimuli. When this stimulation elicited the

ventriloquism effect, the trials were coded with "cong-1" (temporal congruency with ventriloquism effect). If no ventriloquism effect occurred or the congruency was not detected by the participant trials were coded "cong-0" (temporal congruency, without ventriloquism effect). If the stimuli were temporally incongruent no visual stimuli appeared coincidently with the auditory stimuli. Trials without localization shift were successful controls and coded "inc-0" (temporal incongruity control). When a mislocalization occurred during incongruent stimulation or the incongruity was not detected by the participant, the trial was coded "inc-1" (temporal incongruity, deviant).

Behavioral Data analysis

Main effects and interaction effects of the audiovisual synchrony and the stimulation frequency on the number of correct responses on performance were analyzed with a 2 (congruency) x 3 (repetition frequency) ANOVA.

EEG-Data recording

EEG data was recorded from 29 sintered Ag/AgCl ring electrodes (10-20 system, plus FC1, FC2, CP1, CP2, FC5, FC6, CP5, CP6, TP9, TP10 [frontocentral, centro-parietal and temporo-parietal positions]). The scalp electrodes were attached using the BrainCap electrode-cap (Falk-Minow Services, Herrsching-Breitbrunn, Germany), which is part of the MR-compatible EEG-recording system BrainAmp-MR (BrainProducts, Munich, Germany). An electrode located between Fz and Cz was used as the reference electrode during recording. Electrode input impedance was kept below 5 kOhm. Two channels of the 32-channel system were used for recording the electrocardiogram (ECG) to control for heartbeat artifacts in the EEG. One channel recorded vertical EOG. Signals were amplified with a BrainAmp-MR 32-channel amplifier. The sampling rate was 5 kHz; filters were set to 0.016-250 Hz. Trigger signals for stimulation, responses and onset of the gradient for the first slice-

measurement of a volume accompanied the recording of EEG data with the BrainVision Recorder software (BrainProducts).

EEG measurement was performed continuously inside the MR-scanner during functional scanning. To prevent artifacts in the EEG due to pump-induced movements of the subjects, the helium pump of the MR scanner was switched off. Within the scanner-tube, head movements were restricted by a head clamp, attached to the head coil. The subjects lay in a supine position with their head on a vacuum cushion, which avoided small electrode movements inside the magnetic field. Thereby such movements could not cause additional artifacts by inducing electromotive forces.

EEG preprocessing

MR-Gradient artifacts were corrected according to of Allen et al.'s template-method (Allen et al., 2000), with post correction downsampling to 250 Hz and low-pass filtering at 30 Hz. Cardioballistic artifact correction was carried out according to Allen et al. (1998). Both methods are implemented in BrainVision Analyzer software (BrainProducts). After correction of pulse artifacts, ECG channels were discarded.

Eye-blink artifacts and eye-movement artifacts were controlled by visual inspection after performing an independent component analysis (infomax algorithm, in BrainVision Analyzer, Brainproducts, Munich, Germany). Exclusion of corresponding artifact components was based on visual inspection of each component's scalp topography and spectral distribution as well as on visual comparison of each component's time course with the time course of the EOG channel. Per subject three components were excluded. The VEOG channel was discarded afterwards and all channels were re-referenced to their common average. The re-referenced data were sampled down further to 125 Hz. Digital pass band filters between 3.5 Hz / 24 dB and 14.0 Hz / 24 dB were applied.

Time-frequency representations of segmented epochs across frequencies were acquired in increments of 0.125 Hz through continuous wavelet transformation (CWT) using the Morlet mother wavelet. In order to detect single-trial spectral changes, event-related synchronization/synchronization (ERD/ERS) was computed for each wavelet transform, with the 500 ms before stimulation onset being used as the reference phase. The ERD/ERS representations of two groups of electrodes were pooled. The first group consisted of electrodes P3, P7, CP1, and CP5; the second, P4, P8, CP2, and CP6. Segments were cut off at 0 ms and 3000 ms equivalent to the time of one volume of the fMRI. Computations were performed with the signal-processing toolbox LetsWave software (http://www.md.ucl.ac.be/nefy/facecatlab/mouraux/letswave/).

EEG analysis

The parameter extraction was performed with MATLAB software (The MathWorks, Inc.). The activation in each condition is represented by the grand average of the ERD/ERS spectrograms. The differences between these grand average time-frequency representations define the three contrasts of interest in the present study.

The first contrast of interest was 'cong-1 – inc-0', representing the effect of the crossmodal binding process leading to the perception of one event.

The second contrast was 'cong-1 – cong-0', representing the illusion or localization shift of the ventriloquism effect.

The third contrast was 'cong-0 + cong-1 – inc-0 + inc-1', representing the effect of temporal congruency without regard of effects on the localization.

Analyses were conducterd separately for the electrodes, which were ipsilateral to the synchronous audiovisual stimuli, and the contralateral group of electrodes. Also congruent trials were separated into trials in which the visual stimuli presented in the left visual field were synchronous with the auditory stimuli – labeled with an additional

'left' (cong-0-left, comg-1-left), and trials with the synchronously presented visual stimuli in the right visual field – labeled with an additional 'right' (cong-0-right, comg-1- right).

The time-frequency spectrograms were smoothed in time with a Gaussian filter (FWHM = 1.248 s) and the maximum set to 1 to create prototypical difference spectrograms. The modulation-parameters for EEG-guided fMRI-analysis were calculated by multiplying single trial ERD/ERS with the time-frequency representations of the contrasts and summing up the energy (amplitudes over time) in the resulting spectrogram. For each single trial, the total amount of energy quantified the energy related to the processes defined by the difference spectrograms. For the three analyses this results in two sets of modulation-parameters, one set for the electrodes ipsilateral to the visual stimuli synchronous with the auditory stimuli and one set for the contralateral electrodes.

FMRI acquisition

Functional MRI was measured with a 1.5 Tesla Siemens Symphony whole-body scanner with Quantum-gradients. Functional images were obtained using a single shot T2*-weighted gradient-echo planar imaging (EPI) sequence. Thirty slices were acquired in descending fashion to cover all parts of the brain. Slice thickness was 4 mm with 1 mm gaps. Other MR-scanning parameters were: repetition time TR = 3000 ms, echo time TE = 59 ms, flip angle = 90°, field of view FOV = 192x192mm, voxel size = 3x3x5 mm, in-plane image resolution was 64 x 64 voxels. For control of the normalizing procedure and for the purpose of documentation, an anatomical MRI was acquired using a T1-weighted, three-dimensional MPR sequence.

FMRI analysis

FMRI data were analyzed with SPM5 (Wellcome Institute of Neurology at University College London, UK. http://www.fil.ion.ucl.ac.uk/spm). The first five images of each functional recording session were discarded to account for the magnetic field saturation time. The preprocessing of the EPI-data included movement correction and correction of field image distortion, normalization into the MNI-reference space and smoothing (FWHM = 8 mm). At the first level, general linear models were applied to the data of each subject for each set of modulation-parameters of the three analyses.

Regressors in the general linear models were 'congruent stimulation on the left', 'synchronous congruent on the right' and 'incongruent stimulation', 'response phase' and 'fixation phase'. Parametric modulations were conducted for each analysis with the parameters described above. The movement parameters computed during the realignment procedure were included in the models. Contrasts of interest were the positive correlations with the modulators

T-tests were calculated voxel-wise to obtain statistical parametric maps. Results were computed based on region of interest (ROI) analyses, limiting the correction for multiple testing to a-priori specified regions of the brain. These ROI were bilateral insula, superior temporal sulcus (STS) – without the temporal pole – superior colliculi (SC), intraparietal sulcus (IPS) and parieto occipital sulcus (POS). Masks with voxelsize of 2x2x2 mm were created using the MARINA-tool ("MAsks for Region of INterest Analysis", by Walter et al., 2002), which is based on the anatomical parcellation of the brain published by Tzourio-Mazoyer et al. (2002).

Results

Behavioral Responses

Table 1
Responses of the participants, means and standard deviation

	Detection of synchrony	Ventriloquism effect	Manipulation successfull cong-1
cong	22.05 (1.72)	17.11(2.87)	16.00 (2.87)
	Detection of asynchrony	Localization to center	Manipulation successful inc-0
inc	22.84 (1.71)	12.68 (4.80)	11.79 (4.70)

cong = temporally congruent audiovisual stimulation; inc = temporally incongruent audiovisual stimulation; 24 trials total for each condition. Manipulation successful = Detection of audiovisual synchrony and localization shift for congruent stimuli (cong-1) or detection of asynchrony and no localization shift for incongruent stimuli (inc-0) respectively.

Behavioral responses were analyzed for all 19 participants. The 2x3 (congruency x repetition frequencies) ANOVA resulted in no main effect, but a statistically significant interaction 'frequency * condition' was found ($F[2, 36] = 4.154$; $p = 0.024$; Mauchly-test for sphericity p=.468). Separated for repetition frequencies, the slowes frequency showed a difference for congruent and incongruent stimulation. Congruent stimuli presented with a frequency of 0.75Hz were answered correctly in 7.05 (S.D. = 1.08) trials (of total 8) and incongruent stimuli in 7.84 (S.D. = .0.50) trials, the difference was statistically significant ($p = 0.001$). Congruent stimuli presented with a frequency of 1.00Hz were processed correctly in 7.42 (S.D. = 0.69) trials and incongruent stimuli in 7.47 (S.D. = .84) trials. Stimuli presented with a frequency of 1.25Hz were answered correctly in 7.58 (S.D. = 0.69) trials and incongruent stimuli in 7.53 (S.D. = 1.02).

In the ANOVA of responds for sound-localization the main effect of condition showed statistical significance ($F[1,18] = 17.56$; $p = 0.001$). The main effect of repetition frequency and the interaction "frequency * condition" reached no level of significance. For the overall successful stimulus manipulation– both questions answered as intended –the ANOVA the showed a main effect of condition ($F[1,18] = 14.89$; $p = 0.001$) but no effect of repetition frequency and the interaction.

EEG time-frequency representations
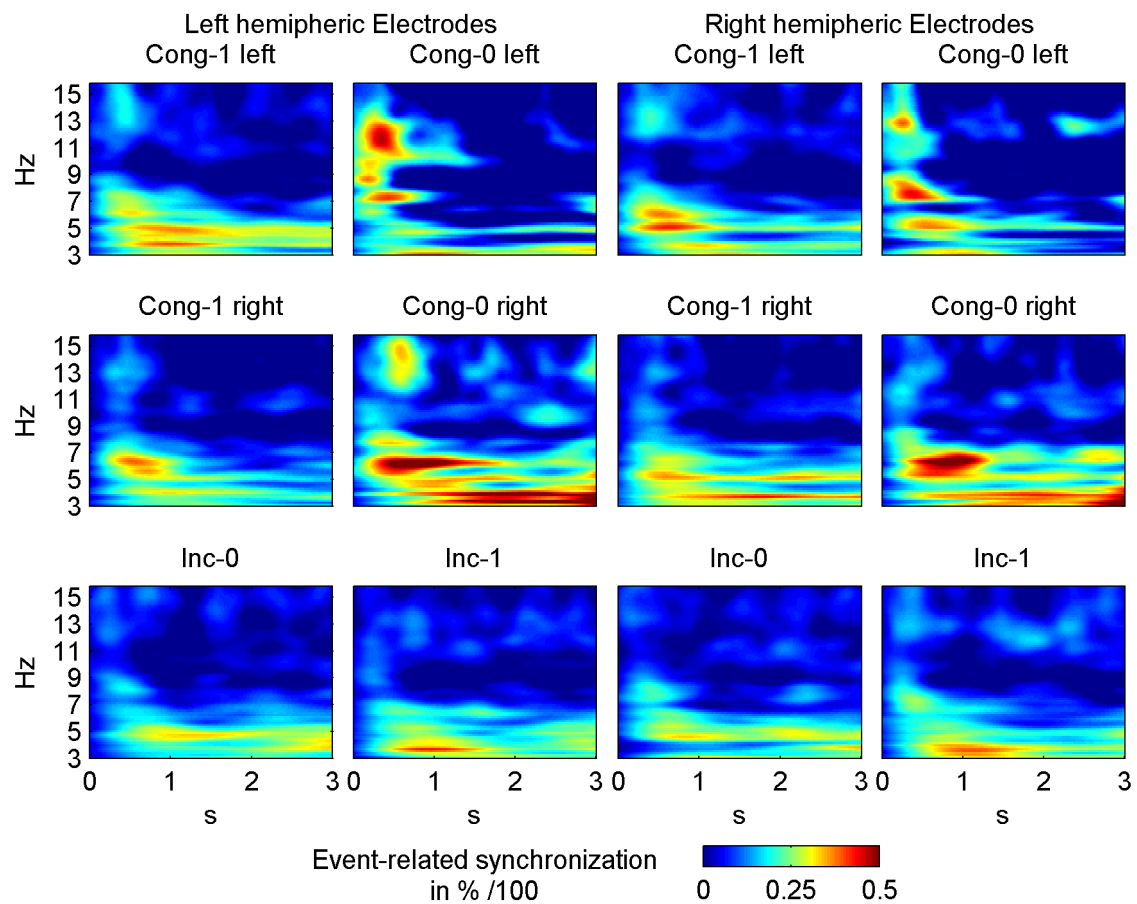
Grand averages

Figure 1. Time-frequency representations of event-related synchronization change (ERD/ERS) relating to the 500 ms pre-stimulus interval. The first two columns show data derived from the left hemispheric electrodes, the last two columns show data derived from the right hemispheric electrodes; cong-1 left = temporally congruent stimuli in the left visual field, congruency detected and ventriloquism effect elicited; cong-1 right = same as cong-1 left, but stimuli in the right visual field; cong-0 left = temporally congruent stimuli in the left visual field, congruency not detected or no ventriloquism effect elicited; cong-0 right = same as cong-1 left, but stimuli in the right visual field; inc-1 = temporally incongruent stimuli, incongruity detected and no localization shift; inc-0 = temporally incongruent stimuli, incongruity not detected or localization shift elicited.

Table 2
Maxima and minima of proportionate synchronization change (ERD/ERS)

	left hemispheric electrodes		right hemispheric electrodes	
	Maximum	Minimum	Maximum	Minimum
cong-1 left	0.40	-0.07	0.43	-0.12
cong-1 right	0.39	-0.10	0.43	-0.09
inc-0	0.33	-0.06	0.34	-0.06
cong-0 left	0.47	-0.20	0.44	-0.20
cong-0 right	1.18	-0.11	0.74	-0.09
inc-1	0.43	-0.09	0.40	-0.07

1 = double of pre-stimulus baseline mean

Spectrograms of contrasts

The bases for the parameter extraction from single trials for the EEG-guided fMRI-analysis are the time-frequency representations of differences between the conditions, the three contrasts of interest.
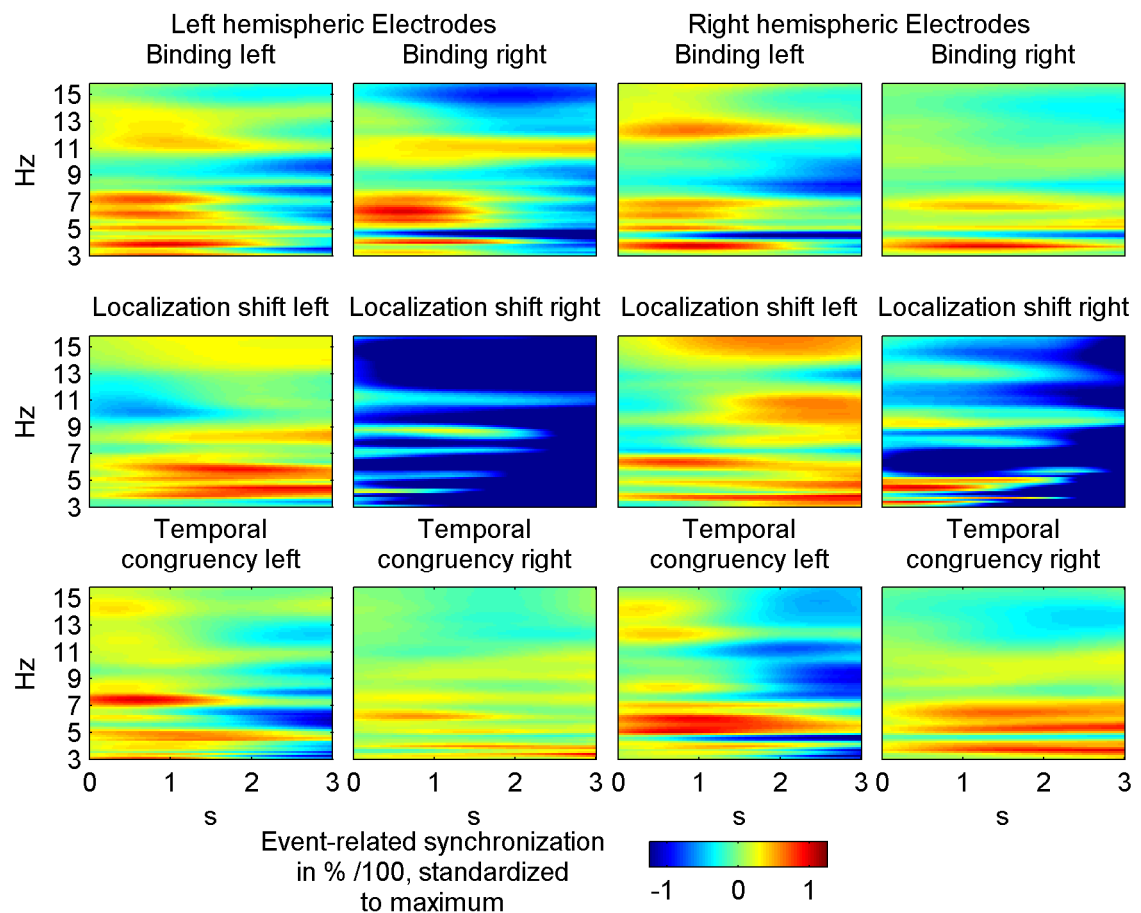

Figure 2. Time-frequency representations of the differences in synchronization (ERD/ERS) for the contrasts of interest (max set to 1). The first two columns

show data derived from the left hemispheric electrodes, the last two columns show data derived from the right hemispheric electrodes; binding = The difference between cong-1 and inc-0 reflected the difference in EEG-activation during audio-visual events in comparison to separated audio and visual events without localization shift, binding left = temporally congruent stimuli in the left visual field (cong-1 left – inc-0); localization shift = The difference between cong-1 and cong-0 reflected the effect of the illusion or localization shift of the ventriloquism effect; temporal congruency = The difference between all congruent stimuli (cong-1 and cong-0) and all incongruent stimuli (inc-0 and inc-1) reflected the overall congruency effect of synchronous audiovisual stimulation.

EEG-Guided fMRI group analyses

Binding

Guiding the fMRI-analysis with a parametric modulation of the regressors by the difference of 'cong-1 – inc-0' shows a single cluster in the right superior temporal sulcus for electrodes ipsilateral to the visual field of congruent audiovisual stimuli ('cong-1 left' represented by spectrograms of left hemispheric electrodes, 'cong-1 right' represented by spectrograms of right hemispheric electrodes etc.). The cluster size is 58 voxels, the maximum of activation in the ROI of the right hemispheric STS is at 45 -27 0 (MNI coordinates), with a t-value of 6.83 and a p_{FWE}-value of 0.001 (family-wise error corrected). For electrodes contralateral to the visual field of congruent audiovisual stimuli, binding is correlated with no statistically significant activations in the ROIs. No statistically significant negative correlations were found.
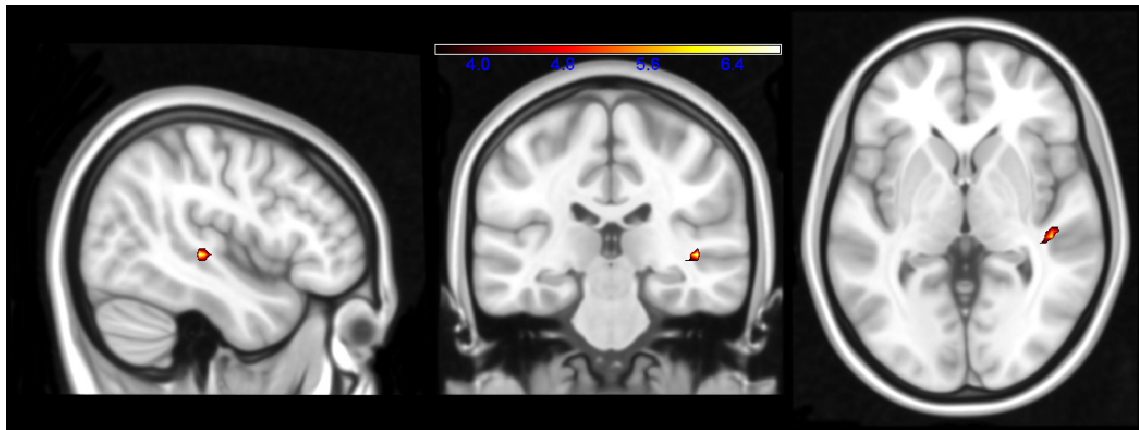
Figure 3. Effect of parametric BOLD modulation by binding (cong-1 – inc-0) at electrodes ipsilateral to the visual field of congruent audiovisual stimuli. Activation is projected on the standard MNI template (ICBM152). Slices are cut at the coordinates of the maximum (45 -27 0). Minimum t-value for illustration is set to 3.61, which equals the t-value for a uncorrected voxel-wise testing at p=0.001 whole-head, the maximum t-value is 6.83.

Localization shift

The parametric modulation of fMRI regressors by the difference of 'cong-1 – cong-0', showed activation in the right insula for electrodes ipsilateral to the visual field of congruent audiovisual stimuli. The cluster is 35 voxels in size, the maximum of activation is at 30 -21 15 (MNI coordinates), with a t-value of 6.44 and a p_{FWE}-value of 0.002 (family-wise error corrected). No statistically significant negative correlations were found.
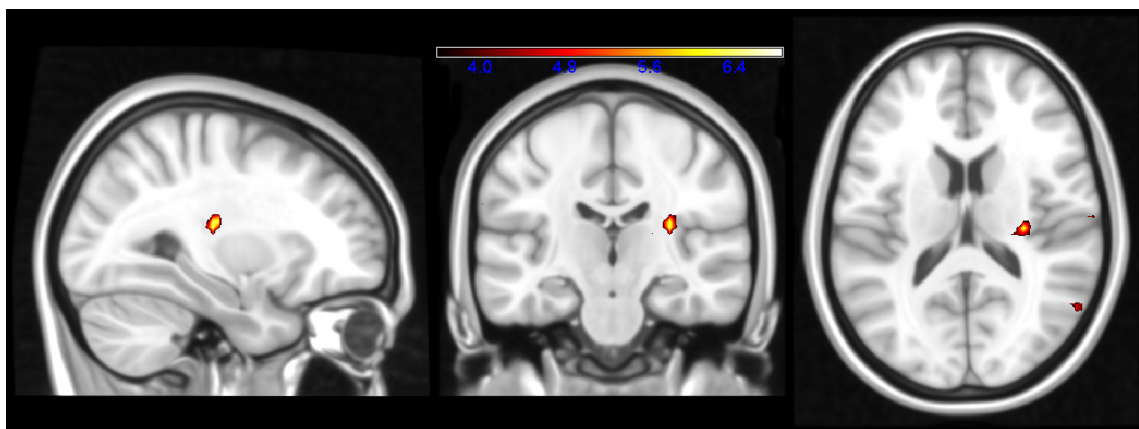

Figure 4. Effect of parametric BOLD modulation by the localization shift (cong-1 – cong-0) at ipsilateral electrodes. Activation is projected on the standard MNI

template (ICBM152). Slices are cut at the coordinates of the maximum (30 -21 15). Minimum t-value for illustration is set to 3.61, which equals the t-value for uncorrected voxel-wise testing at p=0.001 whole-head, the maximum t-value is set to 6.83.

For contralateral electrodes the modulation by 'cong-1 – cong-0' showed a positive correlation in the left IPS. The cluster is 54 voxels in size, the maximum of activation in the ROI of the right hemispheric STS is at -21 -66 39 (MNI coordinates), with a t-value of 4.88 and a p_{FWE}-value of 0.029 (family-wise error corrected). No statistically significant negative correlations were found.
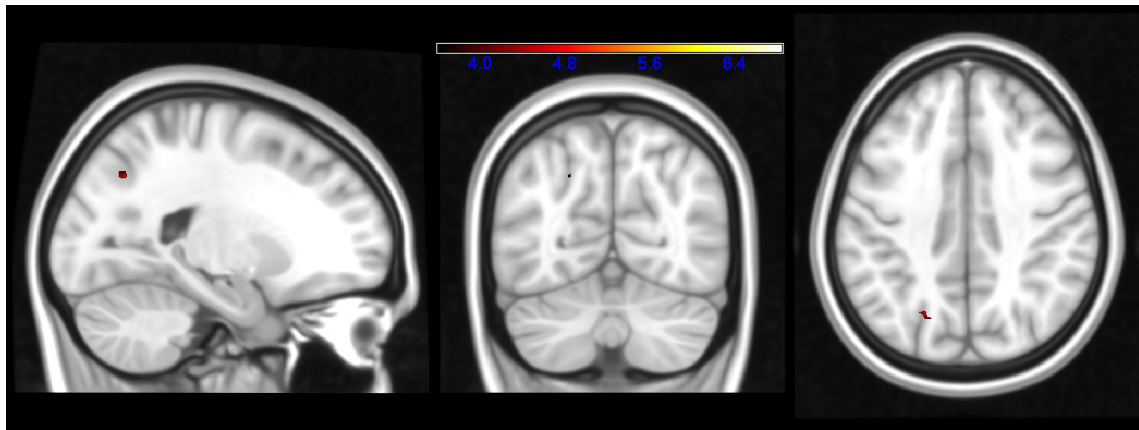

Figure 5. Effect of parametric BOLD modulation by the localization shift (cong-1 – cong-0) at contralateral electrodes. Activation is projected on the standard MNI template (ICBM152). Slices are cut at the coordinates of the maximum (-21 -66 39). Minimum t-value for illustration is set to 3.61, which equals the t-value for uncorrected voxel-wise testing at p=0.001 whole-head, the maximum t-value is set to 6.83.

Temporal congruency

The parametric modulation of fMRI regressors by the difference of congruent versus incongruent stimuli (cong-1 + cong-0 vs. inc-0 + inc-1), showed neither a positive nor a negative correlation in the ROI-analyses.

Discussion

In the present study, we investigated the relationship between multisensory convergence areas and low-frequency neuronal synchronization in the perceptual binding of crossmodal audiovisual stimuli. EEG and fMRI were recorded simultaneously; both data sets represented the identical task performance. Whenever temporally congruent audiovisual stimulation was associated with the ventriloquism effect, which is a localization shift of the auditory stimuli toward the visual stimuli, trials were classified as representing crossmodal binding. With EEG, crossmodal binding was reflected by the difference between synchronization during those crossmodal binding trials and incongruent audiovisual stimulation without localization shift. On a single trial basis, the synchronization differences were quantified to predict BOLD changes in an EEG-guided fMRI-analysis. The modulation of the BOLD-response by ERD/ERS gives information about the association of cortical areas and EEG-dynamic within the framework of crossmodal binding. As a result areas associated with increased synchronization are related to crossmodal binding. Both, the illusion of the ventriloquism effect and the difference between synchronous and asynchronous stimulation were analyzed in the same way.

Without indicating binding by the ventriloquism effect, one cannot determine whether sensory information was integrated across the modalities. When the ventriloquism effect occurred in a congruent trial, our criterion for crossmodal binding events was fulfilled. It may well be that congruent stimulation without the ventriloquism effect was also integrated, the adequate criterion to assure a clean 'no binding' condition is therefore given for incongruent trials without a localization shift (inc-0). The behavioral results show that the congruently presented stimuli successfully induced the ventriloquism effect despite noise and interference in the headphones due to the

magnetic field of the MR scanner. Approximately two thirds of congruent trials showed a localization shift as expected by the ventriloquism effect (cong-1), which is comparable with other studies with similar perceptual effects, e.g. the localizing task of Bonath et al. (2007) induced the ventriloquism effect in 72 percent of the trials.

Changes in the brain states during multisensory integration were reflected in the time-frequency representations. Equivalent to the fMRI analysis a correlate of crossmodal binding was derived from the EEG by contrasting cong-1 trials and inc-0 trials. The difference of grand averages showed increases in synchronization most prominently in the theta band frequency range. The theta activity found for bimodal stimulation (Basar 1999; Basar et al., 2001) seems to be sensitive to the crossmodal integration. The pattern for crossmodal binding was similar for both hemispheres of electrodes and hemispheres of the visual field, in which congruent stimulation was presented. The localization shift of the ventriloquism effect showed a pattern of synchronization changes in similar frequencies as the binding contrast but with a different time course. The comparison was based on a relation of about 2:1 for trials in which manipulation worked as expected versus false or missing responses in the congruent condition. The synchronization gain during binding was similarly found in the pattern of temporal congruency, but it was more heterogeneous between the electrodes and sides of stimulation. This showed that the physical characteristics of the stimuli did not account for all changes in the brain states.

The EEG-guided fMRI-analysis revealed hemodynamic activity corresponding to the EEG-correlate of binding in the right STS. The other ROIs including the areas found activated in the unguided fMRI-analysis (Bischoff et al., 2007) showed no association to the neurophysiological activity during binding. Therefore the STS is specifically involved in the binding of crossmodal stimuli, arguing for its modulating role in inter-

area synchronization when the present findings are related to the literature. Multisensory audiovisual integration areas proximate to unisensory audio and visually sensitive areas in the STS were found with fMRI (Beauchamp et al., 2004) and coherence in the gamma band range was related to the mediating function of the STS in synchronization of primary cortices (Calvert, 2001; Maier et al., 2008). The neural basis of the ventriloquism effect has been investigated by Noesselt et al. (2007) and Bonath et al. (2007), who used fMRI and source localization of separately measured ERP differences between audio-visual and unimodal stimulation at central and parietal electrodes. The combined analysis showed dipolar sources in the auditory cortex of the planum temporale as well as relatively greater activation in auditory cortex contralateral to the shifted sound position in fMRI (Bonath et al., 2007). Analyzing effective connectivity of fMRI activations, effects of temporal congruency (without regard to a behavioral indicator) were found in the STS and in primary visual and auditory areas (Noesselt et al., 2007). In a dipole modeling of short-latency ERP related to the sound-induced flash illusion, EEG sources are located in the auditory cortex and the superior temporal gyrus; this EEG activity is concurrent with gamma activity over the occipital lobe (Mishra et al., 2007). Acoustic deviants in a row of audiovisual stimuli show increases in gamma activity in the STS, whereas illusionary deviants induced by the visual incongruity of a McGurk effect do not (Kaiser, Hertrich, Ackermann & Lutzenberger, 2006).

Gamma band coherence between the lateral auditory belt region and the superior temporal sulcus in monkeys is reported by Maier et al. (2008) for a special form of congruent audiovisual stimulation. They use looming and receding stimuli beside unimodal controls. The congruent looming stimulation elicits increased gamma-band coherence compared to the unimodal stimuli, to the incongruent stimuli and to the

receding stimuli (Maier et al., 2008). Changes of the coherence of the local field potentials (LFPs) activity are attributed by the authors to enhanced phase-locking by feedback projections from the STS to the auditory cortex. However, the differences between bimodal stimulations in terms of saliency and time course– in the congruent looming condition, both modalities were highly salient, inducing stronger neuronal responses –may effect the coherence (Bauer, 2008). The incongruity is not given from the point of onset but develops by the dynamical induced movement. The results of Maier, Chandrasekaran, and Ghazanfar (2008) therefore are addressing the perceptual level of binding and reveal a role of gamma coherence in crossmodal binding of moving events presumably mediated by the STS. Nevertheless their findings can not clearly be interpreted as binding by synchronization (Bauer, 2008). In the present study synchrony changes in the (human) STS were likewise investigated, but related those directly with a behavioral effect, recorded synchronization in lower frequencies and correlated the correlates of audiovisual congruency in EEG and fMRI. The main comparison in this study also contains conditions with differing physical properties, but complementary analyses of the multisensory illusion and the effect of stimulation synchrony afford us to relate the results to the perceptual level.

Our findings add to the involvements of the STS described in the literature the relationship of neurophysiological low-frequency activity and hemodynamic activity during crossmodal binding. The results found in this study can not be reduced to an effect of the synchrony of stimuli, since the parametric modulation by the temporal congruency resulted in no correlation at all. The cluster in the STS is only observed when filtering out the trials without the ventriloquism illusion.

The localization shift of the ventriloquism effect was correlated with the right insula and the left IPS. Assuming that trials without effect did not initialize the task as

intended, e.g. were to difficult to locate or the task was not performed throughout the whole time of stimulation. The insula is sensitive to simultaneous onsets of stimuli, therefore involved in synchrony detection (Bushara et al., 2001; Calvert, et al., 2001), whereas the IPS is involved in multisensory spatial location processing (Bremmer et al., 2001; Bushara et al., 2001; Callan et al., 2001; Calvert et al., 2001; Fort et al., 2002; Molholm et al., 2002; Macaluso et al., 2004).

While convergence areas have been found by single-cell and neuroimaging studies and neuronal oscillation is discussed as an important mechanism of integration, the relation of both with the perceptual level of binding is seldom investigated. One reason for this is the focus on gamma band activities in multisensory integration research (for a review, see Senkowski et al., 2008). To our knowledge a reliable extraction of gamma EEG from signals recorded during EPI scanning is not possible with current methods of MR-artifact correction (Mandelkow et al., 2006, Gebhardt et al., 2008). We focused on the effects of audiovisual stimulation in the theta band like reported by Basar (1999).

Separately recorded fMRI can be used for determining EEG dipoles to model the neural activity (Menon & Crottaz-Herbette, 2005). However this offers no information of the relationship of the signal courses. Furthermore the finding of a vascular response partially dissociating with any local neuronal activity (Sirotin & Das, 2009) challenge the assumed association of neuronal activity measured in local field potentials or EEG and the hemodynamic response measured in fMRI (Logothetis et al., 2001; Logothetis & Pfeuffer, 2004). In the present study the simultaneously combined measuring afforded an extremely important advantage for the interpretation of found activations. If both signals correlate, the process localized with fMRI can be attributed more unambiguously. Therefore the quantified EEG correlate of crossmodal binding was used to model the simultaneously measured BOLD-response in fMRI. This modulation of the

hemodynamic signal gave information about the association of cortical areas and EEG-dynamic within the framework of crossmodal binding. As a result areas associated with increased synchronization were related to the perceptual binding of crossmodal audiovisual stimuli.

Chapter 5

General Discussion

In the present research, the neurofunctional mechanisms of crossmodal audiovisual binding in the human brain have been investigated. When crossmodal binding of congruent audiovisual stimuli occurred, the ventriloquism effect was used as an indicator for it. Incongruent audiovisual stimuli without localization shift served as a bimodal control. In a network of multisensory convergence areas, a hemodynamic correlate of crossmodal binding was found in the insula the parieto-occipital sulcus (POS) and the superior temporal sulcus (STS). An electrophysiological correlate in changes of neural synchronization was identified with time-frequency analyses of theta and alpha band activity at parietal electrodes. The time-frequency representations showed different patterns for binding and control conditions. On a single trial basis, the synchronization was weighted on the basis of these different patterns and quantified as total energy to predict BOLD modulations associated with binding in an EEG-guided fMRI-analysis. The main result revealed the STS to be associated with increased synchronization and hence related to audiovisual crossmodal binding.

The ventriloquism effect was successfully induced by synchronously presented audiovisual stimuli despite the noise of the MR scanner and the interference in the headphones caused by the magnetic field. The illusion occurred in two thirds of all experimental trials; the lowest rate in a subject was 46 percent. This effect rate is relatively robust compared to other studies with similar perceptual effects. In the localizing task of Bonath et al. (2007) the ventriloquism effect was induced in 72 percent of all trials and control stimulation was successful in 69 percent on average. In the sound-induced illusory flash paradigm of Bhattacharya, Shams and Shimojo (2002),

a sound-induced illusory flash was perceived in three quarter of all trials, when stimuli were presented in the peripheral field of view. In the fovea, the effect occurred only infrequently (Bhattacharya et al., 2002). In the similar illusory double-flash design of Mishra et al. (2007) the illusory flash was perceived in around a third of all trials when an auditory stimulus precedes and follows a single flash. However, the range of the inter-individual rate varied between single-digit and over eighty percent. In the McGurk effect paradigm of Kaiser, Hertrich, Ackermann, Mathiak and Lutzenberger (2005), subjective changes from the standard syllable were reported by participants in 80 percent of the physically different acoustic deviants and in 57 percent of the illusionary deviants induced by visual incongruity. Therefore, the rate of successful stimulus manipulation in the present projects was an adequate basis for analyzing differences of averaged signals. Furthermore in the combined analysis the weighted quantified EEG-energy was correlated with the BOLD-response in all experimental trials. The fraction of successful experimental manipulations was only relevant for the weighting, but it did not affect power per se.

In the first project of this thesis brain activations by crossmodal binding were identified in a network of classical multisensory convergence areas. Audiovisual binding indicated by the ventriloquism effect activated the right insula, the bilateral POS and the STS. The insula showed no activation when temporally congruent stimulation was compared with incongruent stimulation. Therefore, the activation in the insula may result from the binding of parts respectively shape. This is in line with findings in visual-tactile form integration by Amedi et al. (2005). Also, the detection of temporal congruency of stimulus onsets is a possible explanation (Bushara et al., 2001; Calvert, et al., 2001). The activation found in the POS can be related to the localization during the ventriloquism effect as the detection of spatial congruency has been

attributed to the POS (Bushara et al., 2001; Callan et al. 2001; Calvert et al., 2001; Fort et al., 2002; Molholm et al., 2002). The activation found in the STS may reflect the audiovisual integration of object features in crossmodal binding as the STS is shown to be involved in object identification (e.g. Beauchamp et al., 2004). Audiovisual temporal correspondence increases effective connectivity in the STS and in primary visual and auditory areas (Noesselt et al., 2007), supporting a modulating role of the STS in audiovisual processing. For speech stimuli, activation in the STS is reported to be insensitive to the relative location of stimuli (Macaluso et al., 2004), suggesting that the spatial aspect of the task in the present projects does not contribute to the activation. This is supported by the finding that a localization task activates the IPS in a stronger fashion than an object recognition task whereas in the STS, the object recognition task elicits more activation than the localization task (Sesteri et al., 2006). The combined analysis in the third project shed further light on the roles of the areas that were found to be involved (see below).

In the second project, the cortical changes that occurred during multisensory integration are reflected in time-frequency representations in EEG. In the EEG, time-frequency representations of crossmodal binding were created by contrasting congruent stimuli eliciting a ventriloquism effect (cong-1) and incongruent stimuli without localization shift (inc-0) equivalent to the fMRI analysis. The difference of grand averages most prominently showed synchronization increases in the theta band frequency range. Audiovisual crossmodal binding was reflected in power changes in the lower frequency range, suggesting that the theta activity found for bimodal stimulation is sensitive to crossmodal integration (Basar 1999; Basar et al., 2001). In order to test whether this pattern of averaged time-frequency representation is suitable to quantify the binding-related EEG activity, trial-by-trial energy was weighted with the averaged

time-frequency representation. The mean trial-by-trial energy of congruent stimulation was compared to the mean energy of incongruent stimulation. For both groups of electrodes (left and right hemispheric electrodes), the congruent stimulation showed higher energy when weighted with the averaged pattern, supporting the assumption that the synchronization pattern identified the binding-related power changes.

When the time-frequency representations of the difference between congruent stimuli eliciting a ventriloquism effect and incongruent stimuli without localization shift were analyzed separately for the visual field hemisphere in which the congruent stimuli were presented, a similar pattern was observed for both sides of congruent stimulation (left/right visual field) and both groups of electrodes. In similar frequencies but with different time course as this binding contrast (cong-1 – inc-0), the localization shift of the ventriloquism effect (cong-1 – cong-0) showed a specific pattern of changes in synchronization. The comparison of the localization shift was based on a relation of about 2:1 for trials in which manipulation occured versus false or missing responses in the congruent condition (cong-1 : cong-0). Unsuccessfully manipulated stimuli might have evoked unfavorable residual processes which distorted the time-frequency representation. As an example, the synchronization increase in the low theta range specific for inc-1 around the first second may reflect a distorting process whereas inc-0 did not show this increase. The synchronization gain during binding was similarly found in the pattern of temporal congruency, but it was more heterogeneous between the electrodes and sides of stimulation. This suggests that the physical characteristics of the stimuli did not account for all changes in the brain states.

Whereas other studies focus on the illusion of the effect to investigate the perceptual level of integration, the present projects distinguish the binding process from bimodal stimulation without binding. The electrophysiological correlate of the sound-induced

illusory flash effect (i.e., the illusionary second flash which is perceived when one flash is accompanied by two rapid sounds; Shams, Kamitani & Shimojo, 2000), has been analyzed in time-frequency decompositions by Bhattacharya et al. (2002). The average gamma band responses in the event-related potentials (ERPs) recorded at occipital electrodes were higher in illusion trials than in trials without illusion (Bhattacharya et al., 2002). The authors point out that the flash illusion only occurs when the modulation of visual processing by auditory stimuli exceeds a 'perception threshold' for the registration of conscious awareness. In the conception of the present thesis the comparison of near-threshold stimulation is disadvantageous, because the absence of the illusion does not ascertain the absence of a crossmodal interaction.

With simultaneous recording of EEG and fMRI, dynamic aspects of brain activity can be related to metabolic local signals. The relationship of multisensory convergence areas and neuronal oscillation has (yet) seldom been investigated. One reason for this is the focus on gamma band activities in multisensory integration research (for a review, see Senkowski et al., 2008). The simultaneous recording demands sophisticated techniques because the artifacts induced by the magnetic field and especially the reading pulses of fMRI exceed the EEG signal by more than two orders of magnitude (Allen et al., 2000). A reliable extraction of gamma EEG from signals recorded during EPI scanning seems impossible with current methods of MR-artifact correction (Mandelkow et al., 2006; Gebhardt et al., 2008). Separately recorded fMRI can be used for determining EEG dipoles to model the neural activity (Menon & Crottaz-Herbette, 2005). However, this provides no information on the relationship of the signal courses. Therefore, in the present thesis the quantified EEG correlate is used to model the simultaneously measured BOLD-response in fMRI. This modulation of the hemodynamic signal holds information on the association of cortical areas and EEG-

dynamic within the framework of crossmodal binding. As a result, areas associated with increased synchronization are related to audiovisual binding.

The assumed association of neuronal activity measured in local field potentials or EEG and the hemodynamic response measured in fMRI has recently been challenged in a study by Sirotin and Das (2009). In the primary visual area of macaque monkeys, they recorded LFPs, multi-unit spike trains, blood volume and blood oxygenation during a brief extra-foveal visual stimulus. The high-frequency LFPs and multi-unit spike trains showed a clear correlation. Two components of the hemodynamic signal were found. Both were related to the task, but only one of the hemodynamic components could be predicted by neuronal signals. The second component showed no correlation with the LFP or multi-unit activity and occurred even in "blank" trials without a visual stimulus (Sirotin & Das, 2009). Tests with trials of different duration showed that the second component in the hemodynamic signal was trial-related and correlated with the arterial contraction-dilation cycle. The trial-related component did not occur when auditory stimuli were used instead of visual stimuli. The authors attributed this component to anticipatory processes. The finding of a vascular response which is (partially) dissociated from local neuronal activity raises concerns in the interpretation of the former. However, it also adds an extremely important advantage for simultaneously combined measuring. If both signals correlate, the activation localized with fMRI can be attributed more precisely.

The simultaneously combined analysis in the third project revealed hemodynamic activity corresponding to the EEG-correlate of binding in the right STS. This association clarifies the findings of the fMRI-analysis of the first project. The EEG-correlate of the ventriloquism effect (illusion greater than no illusion) correlated with activity in the right insula and the left IPS. These results fit nicely into the roles

proposed to the areas. The absence of the illusion can either be attributed to a failure of processing stimulus synchrony or to a high level of difficulty in localization. The insula is sensitive to simultaneous onsets of stimuli (Bushara et al., 2001; Calvert, et al., 2001), whereas the IPS is involved in multisensory spatial location processing (Bremmer et al., 2001; Bushara et al., 2001; Callan et al., 2001; Calvert et al., 2001; Fort et al., 2002; Molholm et al., 2002; Macaluso et al., 2004). The association with crossmodal binding found in this project cannot be reduced to an effect of the synchrony of stimuli as the parametric modulation through the temporal congruency resulted in no correlation at all. The cluster in the STS is only observed when the trials without ventriloquism illusion are filtered out.

The main finding of the EEG-guided fMRI analysis showed the STS to be specifically involved in perceptual binding of crossmodal stimuli. This finding strongly suggests the STS as an area of multisensory convergence and a mediator of neural synchronization. With fMRI multisensory integration areas in the STS have been found proximate to unisensory auditory and visual areas in the STS (Beauchamp et al., 2004), and coherence in the gamma band range has been related to the mediating function of the STS in synchronization of primary cortices (Calvert, 2001; Maier et al., 2008). The neural basis of the ventriloquism effect has been investigated by Bonath et al. (2007) and by Noesselt et al. (2007), who used fMRI and source localization of separately measured ERP differences between audio-visual and unimodal stimulation at central and parietal electrodes. The combined analysis shows dipolar sources in the auditory cortex of the planum temporale as well as relatively greater activation in auditory cortex contralateral to the shifted sound position in fMRI (Bonath et al., 2007). When effective connectivity of fMRI activations is analyzed using audiovisual stimuli, effects of audiovisual temporal congruency can be found in the STS and in primary sensory areas

(Noesselt et al., 2007). In a dipole modeling of short-latency ERP related to the sound-induced flash illusion, EEG sources were located in the auditory cortex and the superior temporal gyrus. This EEG activity was concurrent with gamma activity over the occipital lobe (Mishra et al., 2007). These findings can be explained by a mediating role of the STS in inter-area synchronization.

When the perceptual level of crossmodal binding is examined with illusions, findings are not always clear. Mishra et al. (2007) separated data of a sound-induced illusory flash design for participants who saw the illusory flash more frequently than other participants. They analyzed ERP, source localization, and wavelet analyses at central and occipital electrodes. In participants with high illusion rates they found higher amplitudes of an early modulation of visual cortex activity after the second sound. In addition, an illusion-related activity in the auditory cortex and the STS was found in a single trial analysis of short-latency ERP (Mishra et al., 2007).

Another crossmodal effect utilized to investigate perception is the McGurk effect (McGurk & MacDonald, 1976). Here, the simultaneous presentation of an auditory phoneme and the visual lip movements of a different phoneme are perceived as a third phoneme, e.g., auditory "ba" and a visual "ga" creates the percept "da". In a stream of repeating auditory phonemes, the ones which are accompanied by an incongruent lip movement are perceived as mismatching. With MEG, increases in gamma band activity have been found to be correlated with those subjective deviants (Kaiser et al., 2005). Comparing these perceptions of illusory acoustic to real acoustic changes, the illusory changes elicit fast gamma activity over the midline occipital cortex while real deviants were positively correlated with earlier lower gamma activity over the left superior temporal cortex and negatively with lower gamma activity over the occipital cortex (Kaiser et al., 2006). Hence, the McGurk effect does not seem to involve the

convergence areas that have been found in the present projects. The correlation of real deviants and gamma activity over the superior temporal cortex could result from the disordered crossmodal binding. The illusionary deviant McGurk stimuli may pass through the binding processing in the convergence areas as the congruent audiovisual stimuli do, while the more difficult real deviants may require more processing or higher attention.

Gamma band coherence between the lateral auditory belt region and the superior temporal sulcus in monkeys is reported by Maier et al. (2008) for a special form of congruent audiovisual stimulation. They used looming and receding stimuli as well as unimodal controls. The congruent looming stimulation elicited increased gamma-band coherence compared to the unimodal stimuli, to the incongruent stimuli, and to the receding stimuli (Maier et al., 2008). Changes of the coherence of the local field potentials (LFPs) were independent of power changes and attributed to enhanced phase-locking by feedback projections from the STS to the auditory cortex. The looming stimuli are similar to approaching signals, with a clear dynamic over the 1600ms of stimulation. They are considered to be ecologically more relevant and are perceived with greater saliency (Maier & Ghazanfar, 2007). However, differences between the bimodal stimulations may affect the coherence. In terms of saliency and time course, both modalities were highly salient in the congruent looming condition and induced stronger neuronal responses (Bauer, 2008). The incongruity is not given from the point of onset but develops dynamically. The results of Maier et al. (2008) therefore reveal a specific role of gamma coherence in crossmodal binding of moving events presumably mediated by the STS. Nevertheless their findings cannot explicitly be interpreted as binding by synchronization (Bauer, 2008). In the present thesis synchronization changes in the (human) STS were likewise investigated. However, by examining the association

of an EEG-correlate with an fMRI-correlate, the synchronization could be related to a behavioral effect.

The low-frequency EEG holds information about crossmodal binding. The examination of fMRI associated with high-frequency neuronal coherence could afford to compare this with the involvement of intra-area synchronization of neuronal assemblies, but is technically even more challenging. Also for the further exploration of the found association different modalities and different types of binding seem promising. If associations in convergence areas can be found for different combinations of modalities, they may indeed reflect a principle mechanism. In this thesis, the audiovisual location binding is influenced by the temporal binding inducing the ventriloquism effect. Reciprocally, location binding can induce a shift in temporal perception (e.g., sound-induced illusory flash, Shams et al., 2000). The involvement of the STS on the perceptual level should be observable for location binding as well as for conditional binding. For example, in a paradigm similar to the one used in this thesis a learned association of the vertical position of temporally congruent visual stimuli with the sound pitch should induce a ventriloquism effect whereas temporally congruent visual stimuli at vertical positions not associated with auditory stimuli should not be bound the same way. Likewise, effects of crossmodal binding may be examined without differences in temporal congruency when the first modality is disambiguated by the second modality and this interaction is compared with ambiguous bimodal stimulation.

Concluding, the perceptual binding of crossmodal audiovisual stimuli was reflected in hemodynamic and electrophysiological brain activity. The quantified EEG correlate was used for a parametric modulation of the simultaneously measured BOLD-response in fMRI. This modulation of the hemodynamic signal gave information about the association of cortical areas and EEG-dynamic within the framework of crossmodal

binding. The association is of special interest in the context of the two hypotheses of binding, binding by hierarchically organized convergence areas and binding by temporal correlation. As a result a very distinct area associated with increased synchronization was related to the perceptual binding. When this activation in the STS is related to projections from or to unisensory areas as discussed in the literature, the two hypotheses could be combined for audiovisual binding. The association revealed in the STS provides evidence for the STS to be the mediating convergence area of audiovisual binding.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257-262.

Allen, J. P., Josephs, O., & Turner, R. A. (2000). Method for removing imaging artifact from continuous EEG recorded during functional MRI.. *NeuroImage, 12*(2), 230-239.

Allen, J. P., Polizzi, G., Krakow, K., Fish, D. R., & Lemieux, L. (1998). Identification of EEG events in the MR scanner: The problem of pulse artifact and a method for its subtraction. *NeuroImage, 8*(3), 229-239.

Amedi, A., von Kriegstein, K., van Atteveldt, N., Beauchamp, M.S., & Naumer, M.J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research, 166*(3/4), 559–571.

Basar, E. (1999). *Integrative brain function: Neurophysiology and cognitive processes* (Vol.2). Berlin: Springer.

Basar, E., Basar-Eroglu, C., Karakas, S., & Schürmann, M. (2001/1). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology, 39*(2-3), 241-248.

Bauer, M. (2008). Multisensory Integration: A Functional Role for Inter-Area Synchronization? *Current Biology, 18*(16), 709-710.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience, 7*(11)**,** 1190–1192.

Beauchamp, M. S. (2005). Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics, 3*, 93–113.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review, 5*(3), 482–489.

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*(2), 321–332.

Bhattacharya, J., Shams, L., & Shimojo, S. (2002). Sound-induced illusory flash perception: role of gamma band responses. *Neuroreport, 13*(14), 1727–1730.

Bischoff, M., Blecker, C. R., Gebhardt, H., Morgen, K., Zentgraf, K., Vaitl, D., & Sammer, G. (submitted). Indicators of audiovisual binding in low-frequency EEG.

Bischoff, M., Walter, B., Blecker, C. R., Morgen, K., Vaitl, D., & Sammer, G. (2007). Utilizing the ventriloquism effect to investigate audio-visual binding. *Neuropsychologia, 45*, 578-586.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H. J., & Hillyard, S. A. (2007). Neural Basis of the Ventriloquist Illusion. *Current Biology, 17*(19), 1697-1703.

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., & Hoffmann, K. P., Zilles, K., & Fink, G. R. (2001/1). Polymodal motion processing in posterior parietal and premotor cortex: A human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron, 29*(1), 287–296.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *Journal of Neuroscience, 21*(1), 300-304.

Callan, D. E., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. *Brain Research. Cognitive Brain Research, 10*, 349–353.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11*, 1110–1123.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in human heteromodal cortex. *Current Biology, 10*, 649–657.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage, 14*(2), 427-438.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology, 98*(1/3), 191-205.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*, 37–46.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation, 1*(1), 123–132.

David, O., Kilner, J. M., & Friston, K. J. (2006). Mechanisms of evoked and induced responses in MEG/EEG . *NeuroImage, 31*(4), 1580-1591.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron, 57*(1), 11-23.

Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology, 10*(20), 731-735.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences, 5*(1), 16-25.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic Predictions: Oscillations and Synchrony in Top–Down Processing. *Nature Review Neuroscience, 2*, 704–716.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433.

Fort, A., Defpuech, C., Pernier, J., & Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex, 12*, 1031–1039.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. J. (1991). Comparing functional (PET) images: The assessment of significant change. *Journal of Cerebral Blood Flow & Metabolism, 11*, 690-699.

Gebhardt, H., Blecker, C. R., Bischoff, M., Morgen, K., Oschmann, P., Vaitl, D., & Sammer, G. (2008). Synchronized measurement of simultaneous EEG-fMRI: A simulation study. *Clinical Neurophysiology, 119*(12), 2703–2711.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences, 10*(6), 278-285.

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature, 338*, 334-337.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences, 15*(1), 20–25.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. London: Wiley.

Kaiser, J., Hertrich, I., Ackermann, H., Mathiak, K., & Lutzenberger, W. (2005). Hearing lips: gamma-band activity during audiovisual speech perception. *Cerebral Cortex, 15*(5), 646-653.

Kaiser, J., Hertrich, I., Ackermann, H., & Lutzenberger, W. (2006). Gamma-band activity over early sensory areas predicts detection of changes in audiovisual speech stimuli. *Neuroimage, 30*(4), 1376-1382.

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

Kant, I., & Guyer, P. (1998). *Critique of pure reason* (1. paperback ed.). Cambridge [u.a.]: Cambridge Univ. Press.

King, A. J. (2005). Multisensory Integration: Strategies for Synchronization. *Current Biology, 15*(15), 339-341.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29*(2-3), 169-195.

Klimesch, W. (2006). Binding principles in the theta frequency range. In H. D. Zimmer (Eds.), *Handbook of binding and memory: Perspectives from cognitive neuroscience.* Oxford: Oxford University Press.

Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., & Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research, 166*(3/4), 289 –297.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 12*, 869–878.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*, 150–157.

Logothetis, N. K., & Pfeuffer, J. (2004). On the nature of the BOLD fMRI contrast mechanism. *Magnetic Resonance Imaging, 22*(10), 1517–1531.

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology, 66*, 735–69.

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: A PET study. *Neuroimage, 21*, 725-732.

Maier, J. X., & Ghazanfar, A. A. (2007). Looming biases in monkey auditory cortex. *Journal Neurosciences, 27*(15), 4093–4100.

Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology, 18*(13), 963-968.

Matsuda, T., Matsuura, M., Ohkubo, T., Ohkubo, H., Atsumi, Y., Tamaki, M., Takahashi, K., Matsushima, E., & Kojima, T., (2002). Influence of arousal level for functional magnetic resonance imaging (fMRI) study: Simultaneous recording of

fMRI and electroencephalogram. *Psychiatry and Clinical Neurosciences, 56*(3), 289– 290.

Mandelkow, H., Halder, P., Boesiger, P., & Brandeis, D. (2006). Synchronization facilitates removal of MRI artifacts from concurrent EEG recordings and increases usable bandwidth. *NeuroImage, 32*(3), 1120-1126.

McDonald, J. J., Teder-Sälejärvi, W. A., Di Russo, F., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. *Journal of Cognitive Neurosciences, 15*, 10–19.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Menon, V., & Crottaz-Herbette, S. (2005). Combined EEG and fMRI studies of human brain function. *International Review of Neurobiology, 66*, 291–321.

Mesulam, M. M. (1998). From sensation to cognition. *Brain, 121*(6), 1013-1052.

Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review, 81*(6), 521–35.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research, 6*(1), 57-77.

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early crossmodal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *The Journal of Neuroscience, 27*(15), 4120-4131.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J., (2002). Multisensory auditory–visual interactions during early sensory processing in humans; a highdensity electrical mapping study. *Cognitive Brain Research, 14*, 115–128.

Nielsen, F. A., & Hansen, L. K. (2002). Automatic anatomical labeling of Talairach coordinates and generation of volumes of interest via the BrainMap database Neuroimage, Poster No.: 10497 at HBM'2002. ???

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., & Driver, J. (2007). Audiovisual Temporal Correspondence Modulates Human Multisensory Superior Temporal Sulcus Plus Primary Sensory Cortices. *The Journal of Neuroscience, 27*(42), 11431-11441.

O'Hare, J. J. (1991). Perceptual integration. *Journal of the Washington Academy of Sciences, 81*(1), 44–59.

Pfurtscheller, G., & Aranibar, A. (1979). Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. *Electroencephalography and Clinical Neurophysiology, 46*(2), 138–146

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Journal of Clinical Neurophysioogyl, 110*(11), 1842–1857.

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron, 28*(2), 617–625.

Revonsuo, A. (1999). Binding and the phenomenal unity of consciousness. *Consciousness and Cognition, 8*(2), 173-185.

Revonsuo, A., & Newman, J. (1999). Binding and consciousness. *Consciousness and Cognition, 8*(2), 123–127.

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Cognitive and Behavioral Neurology, 12*, 191– 200.

Sakowitz, O. W., Quirogaa, R. Q., Schürmann, M., & Basar, E. (2005). Spatio-temporal frequency characteristics of intersensory components in audiovisually evoked potentials. *Cognitive Brain Research, 23*(2-3), 316-326.

Schack, B., Vath, N., Petsche, H., Geissler, H.G., & Moller, E. (2002). Phase-coupling of theta-gamma EEG rhythms during short-term memory processing. *International Journal of Psychophysiology, 44*(2), 143–163.

Schürmann, M., & Basar, E. (1999). Alpha oscillations shed new light on relation between EEG and single neurons. *Neuroscience Research, 33*(2), 79-80.

Schroeder, C. E., & Foxe, J. J. (2002). Timing and laminar profile of converging inputs in multisensory areas of the macaque neocortex. *Brain Research. Cognitive Brain Research, 14*(1), 195–207.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends in Neurosciences, 31*(8), 401-409.

Sesteri, C., Di Matteo, R., Ferretti, A., Del Gratta, C., Caulo, M., Tartaro, A., Olivetti Belardinelli, M., and romani, G.L. (2006). "What" versus "Where" in the audiovisual domain: An fMRI study. *Neuroimage, 33*, 672-680.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions - What you see is what you hear. *Nature, 408*(6814), 788-788.

Singer, W. (1998). Consciousness and the structure of neuronal representations. *Philosophical Transactions: Biological Sciences (The Royal Society), 353,* 1829-1840.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron, 24*, 49–65.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18*, 555–586.

Sirotin, Y. B., & Das, A. (2009). Anticipatory haemodynamic signals in sensory cortex not
predicted by local neuronal activity. *Nature, 457*, 475–479.

Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology, 13*, 519-521.

Stein, B. E., & Meredith, A. M. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9,* 255-266.

Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters, 357*, 163–166.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences, 3*(4), 151–162.

Teder-Salejarvi, W. A., Di Russo, F., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience, 17*(9)*,* 1396-1409.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6*, 171–178.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15*, 273-289.

Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception and Psychophysics, 63*(4), 651–659.

Von der Malsburg, C. (1981). *The correlation theory of brain function.* Internal Report 81-2. Reprinted in Models of Neural Networks II (1994), E. Domany, J. L. van Hemmen, and K. Schulten, eds. (Berlin: Springer). 1994: 95-119.

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*(3), 301-313.

Von Stein, A., Rappelsberger, P., Sarnthein, J., & Petsche, H. (1999). Synchronization between temporal and parietal cortex during multimodal object processing in man. *Cerebral Cortex, 9*(2), 137-150.

Walter, B., Blecker, C., Kirsch, P., Sammer, G., Schienle, A., & Stark, R. (2003). MARINA: An easy to use tool for the creation of MAsks for Region of INterest Analyses [abstract]. In Presented at the ninth international conference on functional mapping of the human brain, vol. 19, no. 2 (available on CD-Rom in NeuroImage).

Worsley, K., Evans, A., Marrett, S., & Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism, 12*, 900–918.

Zampini, M., & Spence, C. (2004). The role of auditory cues in modulating the perceived crispness and staleness of potato chips. *Journal of Sensory Studies, 19*, 347–363.

Erklärung

Der Autor versichert, die vorgelegte Dissertation selbständig angefertigt zu haben.

Weiterhin versichert er, nur die aufgeführten Hilfsmittel verwendet zu haben.

Gießen, April 2009

Matthias Bischoff