Fachbereich Erziehungswissenschaften und Psychologie
der Freien Universität Berlin

Neural correlates of multisensory processing

Dissertation
zur Erlangung des akademischen Grades
Doktor der Psychologie
(Dr. phil.)

vorgelegt von
Diplom Psychologin
Yadira Roa Romero

Berlin, 2016

Erstgutachter: Prof. Dr. Michael Niedeggen

Zweitgutachter: Prof. Dr. Daniel Senkowski

Tag der Disputation: 19. April 2016

Danksagung

An erster Stelle möchte ich meinen beiden Betreuern Daniel Senkowski und Michael Niedeggen danken, die diese Arbeit ermöglicht und betreut haben. Michael Niedeggen danke ich insbesondere für die kritischen Fragen, die Möglichkeit meine Forschungsergebnisse regelmäßig im Forschungskolloquium zu diskutieren und die Bereitschaft diese Arbeit zu korrigieren. Mein besonderer Dank gilt Julian Keil, der stets für Fragen und Diskussionen offen war, mir methodische Inspiration bot und mich bei allen auftretenden Problemen stets unterstütz hat. Meiner Kollegin Johanna Balz, sowie Paulina Schulz, Ivan Prikhodko und Markus Koch möchte ich insbesondere für ihre tatkräftige Unterstützung bei den EEG- und Patienten Messungen danken, ohne ihre kontinuierliche Mithilfe wäre diese Arbeit nicht möglich gewesen. Meinen Kollegen Mathis Kaiser und Martin Krebber möchte ich danken für ihre wunderbare Art den Büroalltag aufzulockern.

Meiner Familie und meinen Freunden danke ich, dass sie mich auf meinem Weg durch die Promotion stets interessiert und aufbauend begleitet haben.

Peter, dir danke ich für die Möglichkeit das Unwesentliche in meinem Leben ausblenden zu können und stets neuen Mut schöpfen.

Table of Contents

Abbreviations

ANOVA = Analysis of Variance

DICS = Dynamic Imaging of Coherent Sources

e.g. = for example (Latin exempli gratia)

EEG = Electroencephalography

ERP = Event related potential

et al. = and others (Latin et alii)

fMRI = Functional Magnetic Resonance Imaging

GFP = Global field power

Hz = Hertz

IFG = Inferior frontal gyrus

LAURA =Local autoregressive average

MEG = Magnetencephalography

MNI =Montreal Neurological Institute

MRI = Magnetic Resonance Imaging

ms = milliseconds

s = seconds

STG = Superior temporal gyrus

STS = Superior temporal sulcus

TMS = Transcranial Magnetic Stimulation

Zusammenfassung

Die multisensorische Verarbeitung und Integration von Reizen aus unterschiedlichen Sinneskanälen ist essentiell für eine kohärente Wahrnehmung unserer Umwelt. Dabei stellt audiovisuelle Sprache ein faszinierendes Phänomen multisensorischer Integration dar, welches die zeitlich koordinierte Extraktion von auditorischen und visuellen Informationen aus verschiedenen Sinneseindrücken erfordert, um später eine kohärente Wahrnehmung audiovisueller Sprache zu ermöglichen. In diesem Zusammenhang wird synchronen neuronalen Oszillationen eine besondere funktionelle Bedeutung zugeschrieben, da diese als potentzieller Kommunikations- und Integrationsmechanismus zwischen entfernten Hirnarealen fungieren und so eine wichtige Rolle innerhalb der multisensorischen Integration einnehmen (Fries, 2005; Senkowski, Schneider, Foxe, & Engel, 2008). Gegenwärtig ist das Interesse an Oszillationen aus den niedrigem Frequenzbereich (3-40 Hz), welche in Prozesse der auditorischen Sprachverarbeitung, Aufmerksamkeit und Wahrnehmung involviert sind, stetig gestiegen. Neben den bisherigen Untersuchungen zu oszillatorischen Prozessen an gesunden Probanden, haben Studien zur unisensorischen Verarbeitung bei Patienten mit psychiatrischen Erkrankungen, wie beispielsweise der Schizophrenie ein gesteigertes Interesse erfahren. So konnten bisherige Studien zur unisensorischen Verarbeitung bei Patienten mit Schizophrenie Defizite bei der visuellen Perzeptbildung aufzeigen (Grützner, Wibral, Sun, et al., 2013) . Interessanterweise, konnten Studien zur unisensorischen Verarbeitung erste Evidenz für veränderte oszillatorische Aktivität bei Patienten mit Schizophrenie liefern (Popov, Rockstroh, Popova, Carolus, & Miller, 2014; Uhlhaas & Singer, 2010) . Trotzdem sind die spezifischen Hinweise für ein Vorliegen eines multisensorischen Integrationsdefizits bei Patienten mit Schizophrenie unzureichend. Daher soll die vorliegende Arbeit diese Lücke in der Befundlage schießen und die funktionelle Bedeutung von Oszillationen in mehreren Experimenten zur multisensorischen

Sprachverarbeitung, mittels der nicht invasiven Elektroenzephalographie (EEG) bei gesunden Probanden und Patienten mit Schizophrenie untersuchen.

Die erste Studie wurde mit dem Ziel konzipiert die oszillatorischen Mechanismen, die der Entstehung der McGurk Illusion (McGurk & MacDonald, 1976) bei gesunden Probanden zugrunde liegen, aufzudecken. Die McGurk Illusion ist ein Phänomen der multisensorischen Integration, welches auf der Präsentation von inkongruenten audiovisuellen Silben beruht. Die McGurk Illusion entsteht, wenn die visuelle Lippenbewegungen einer Silbe (/ga/) mit einer inkongruenten akustischen Silbe (/pa/)kombiniert werden. Diese Kombination von inkongruenter visueller und akustischer Information induziert oftmals ein verschmolzenes sprachliches Perzept. Beispielsweise wird die Kombination der Silben /ga/ und /pa/ von den meisten Versuchspersonen als neuartige Silbe /ka/ widergegeben. Anhand dieses Paradigmas untersuchte die erste Studie die Bedeutung der Ereignis-korrelierten Potenziale und der langsamen Oszillationen während der Wahrnehmung von kongruenter und inkongruenter audiovisueller Sprache bei gesunden Versuchspersonen. In dieser Studie zeigte sich, dass die frühe (0-500 ms) und späte (500 – 800 ms) Beta Band Power (13- 30 Hz) eine wichtige Rolle für die Wahrnehmung der McGurk Illusion bei gesunden Probanden spielt. Die frühe Reduktion der Beta Band Power während der McGurk Illusion spiegelt eine stärkere frühe Reizverarbeitung wider, wohingegen die Reduktion der späten Beta Band Power mit der Auflösung der audiovisuellen Inkongruenz und einer stärkeren Integration assoziiert ist.

In der zweiten Studie, sollten die Ergebnisse der ersten Studie erweitert und innerhalb der Population mit Schizophrenie Patienten untersucht werden. Gegenwärtig gibt es wenig Befunde bezogen auf die audiovisuelle Integration und deren zugrunde liegenden neuronalen und oszillatorischen Mechanismen bei Patienten mit Schizophrenie. In der zweiten Studie wurde das identische Paradigma wie in der ersten Studie verwendet und so die neuronalen Korrelate der McGurk Illusion, als spezifisches Phänomen der audiovisuellen Integration von Sprache bei Patienten mit Schizophrenie untersucht. Interessanterweise zeigten sich keine Verhaltensunterschiede in Bezug auf die Illusionswahrnehmung zwischen Patienten mit Schizophrenie und Kontrollprobanden. Jedoch zeigte sich eine stärkere Reduktion der späten (550-750 ms) medio-zentralen Alpha Band Power (8-10 Hz) während der Wahrnehmung der McGurk Illusion bei Kontrollen im Vergleich zu Patienten mit Schizophrenie. Da die Reduktion von Alpha Band Power innerhalb der Inhibition Timing Hypothese (Klimesch, Sauseng, & Hanslmayr, 2007) mit aktiven Verarbeitungsprozessen in Verbindung gebracht wurde, suggeriert das Fehlen der Alpha Band Reduktion bei Patienten mit Schizophrenie Defizite während der multisensorischen Integration.

Abschließend wurde in der dritten Studie die audiovisueller Sprachverarbeitung und die Rolle des crossmodalen Prediction Error bei Patienten mit Schizophrenie und Kontrollprobanden untersuchen. Das Konzept des Prediction Error basiert auf der predictive coding Theorie (Friston, 2002). Diese Theorie postuliert, dass das Gehirn Vorhersagen in Bezug auf bevorstehende sensorische Ereignisse generiert und diese Vorhersagen mit den tatsächlich auftretenden Reizen abgleicht. Die Differenz zwischen den erwarteten Reizen und den tatsächlich auftretenden Reizen wird als Prediction Error bezeichnet (Friston, 2002; Rao & Ballard, 1999). In bisherigen Studien konnte gezeigt werden, dass neuronale Oszillationen in die perzeptuelle Verarbeitung involviert sind und möglicherweise zu der Generierung des Prediction Error Signals beitragen (Arnal & Giraud, 2012; Arnal, Wyart, & Giraud, 2011). In der dritten Studie wurde daher ein audiovisuelles Sprachparadigma genutzt, in welchem den Probanden kongruente und inkongruente audiovisuelle Silben präsentiert wurden. Die audiovisuelle Sprache variierte hinsichtlich der Kongruenz und dem Grad der visuellen Vorhersagbarkeit. Mit Hilfe dieses Paradigmas wurde in dieser Studie untersucht wie das Auftreten des crossmodalen Prediction Errors innerhalb der Sprachverarbeitung durch neuronale Oszillationen reflektiert wird. Interessanterweise, zeigte sich in den Resultaten, dass die frontale Theta Band Aktivität (7-8 Hz) stärker ausgeprägt war, wenn die Vorhersagen nicht bestätigt wurden und ihnen die Präsentation einer hoch prediktiven visuellen Silbe voranging. Allerdings war diese frontale Theta Band Aktivität bei den Kontrollprobanden stärker ausgeprägt als bei den Patienten mit Schizophrenie. Dieses Ergebnis spiegelt die verstärkte Detektion eines multisensorischen Prediction Error bei Kontrollprobanden wider, welche jedoch nicht bei den Patienten auftritt.

Zusammenfassend stehen die Ergebnisse der drei hier vorgestellten Studien in Einklang mit den bisherigen Studien und unterstreichen die Bedeutung langsamer Oszillationen für multi- und unisensorische Verarbeitungsprozesse. Wesentlich ist, dass die Studien der vorliegenden Arbeit die bisherigen Studien stützen und die wichtige Rolle der langsamen Oszillationen während multisensorischer Prozesse, wie der Wahrnehmung der McGurk illusion bei gesunden Probanden (**Studie 1**), sowie bei Patienten mit Schizophrenie (**Studie 2**) aufzeigen konnten. Außerdem wurde gezeigt, dass langsame oszillatorische Aktivität der veränderten Ausbildung des multisensorischen Prediction Errors bei Patienten mit Schizophrenie zugrundeliegt (**Studie 3**). Die Resultate dieser Studien unterstreichen die Wichtigkeit weiterer Studien in diesem Forschungsfeld. Diese können dazu beitragen die Relevanz von Oszillationen bezogen auf die multisensorischer Integration, perzeptueller Verarbeitung und predictive coding für gesunde Probanden und innerhalb psychiatrischer Störungen, wie der Schizophrenie aufzudecken.

Aufbauend auf den hier präsentierten Ergebnisse, lassen sich einige interessante Forschungsfragen für weitere Studien ableiten. Eine diese potenziellen Nachfolgestudien könnte sich mit der Anwendung kognitiv-perzeptueller Trainings, als besonders vielversprechender Ansatz zur Veränderung defizitären multisensorischen Verarbeitens beschäftigen. Verschiedenste psychiatrische Störungen, wie sich bei Patienten mit Schizophrenie oder Autismus zeigt, haben Defizite bei der multisensorischen Emotions- und Sprachverarbeitung. Insbesondere diese Patientengruppen könnten von solchen Training profitieren.

Abschließend lässt sich sagen, dass die vorliegenden Ergebnisse einen ersten Ansatzpunkt für die Untersuchung von oszillatorischen Hirnmechanismen darstellen, welche der komplexen multisensorischen Sprachwahrnehmung bei gesunden Probanden und psychiatrischen Störungen, wie der Schizophrenie zugrundeliegen.

Abstract

Multisensory processing and the integration of stimuli from various senses are essential for the formation of coherent percepts within our environment. Audiovisual speech processing is an intriguing multisensory phenomenon, which requires the timely coordinated extraction of auditory and visual information from separate input streams to later form a coherent audiovisual speech perception. In this context, a special functional significance has been attributed to synchronized oscillations, which might serve as potential mechanism for communication and integration between distinct cortical areas and play a crucial role for multisensory integration (Fries, 2005; Senkowski et al., 2008). Recently, there has been an increasing interest in oscillations of the lower frequency bands (3-30 Hz), which are involved in various processes such as auditory speech processing, attention and perception. Beside research in healthy participants also studies on sensory processing in patients with psychiatric disorder, such as schizophrenia gained increasing interest. Hence, previous studies indicated perceptual deficits in unisensory processing in patients with schizophrenia (Butler & Javitt, 2007; Grützner, Wibral, Sun, et al., 2013). Interestingly, these unisensory studies have also evidenced abnormal oscillatory activity in schizophrenia (Popov et al., 2014; Uhlhaas & Singer, 2010). However, specific support for deficient multisensory processing in schizophrenia is still incomplete. Hence, to bridge this gap the present thesis scrutinized the functional significance of slow oscillations in a broad range of multisensory speech tasks, using non-invasive electroencephalography (EEG) in healthy participants and patients with schizophrenia.

The first study investigated the underlying oscillatory mechanisms contributing to the subjective perception of the McGurk illusion (McGurk & MacDonald, 1976) as a phenomenon of multisensory speech processing. The McGurk illusion occurs when presenting a visual lip movement pronouncing a syllable (e.g. /ga/) is paired with an incongruent auditory syllable (e.g. /pa/). This pairing of incongruent visual and acoustic information usually induces a

fused speech percept. For instance, the combination of /ga/ and /pa/ the syllable (/ka/) is reported by most of the participants. Using this paradigm, the first study examined the role of event-related potentials and slow oscillations during the perception of congruent and incongruent audiovisual speech in healthy controls. The study demonstrated that early (0-500 ms) and late (500 – 800 ms) beta-band power (13- 30 Hz) play an important role in the perception of the McGurk illusion in healthy subjects. The early beta power suppression during the McGurk illusion reflects stronger early stimulus processing, whereas the later beta power suppression relates to the resolution of audiovisual incongruence and stronger integration.

In the second study, the findings from the first study were extended within the population of patients with schizophrenia. To date, evidence of audiovisual integration and its underlying neuronal and oscillatory mechanisms in patients with schizophrenia is very sparse. Hence, using the same design as the first study, the second study examined the neural correlates of the McGurk illusion, as a specific phenomenon of audiovisual speech integration in this patient group. Interestingly, no behavioral differences in illusion rates were found. However, the study revealed stronger suppression of late (550-750 ms) medio-central alpha-band activity (8-10 Hz) during the perception of McGurk illusion in matched controls compared to patients. As alpha band suppression has been linked to active processing within the inhibition timing hypothesis (Klimesch et al., 2007), the lack of late alpha-band suppression in patients suggests a deficient integrative multisensory processing.

Finally, the third study investigated audiovisual speech processing and the role of crossmodal prediction error in patients with schizophrenia and matched controls. The concept of prediction error refers to predictive coding theory (Friston, 2002). This theory postulates that the brain generates predictions about upcoming sensory stimuli and compares incoming sensory input with these predictions. The difference between the expected stimuli and sensory input is termed prediction error (Friston, 2002; Rao & Ballard, 1999). It has been shown that neural oscillations are involved in perceptual processing and might also contribute to the computation of a prediction error signal (Arnal & Giraud, 2012; Arnal et al., 2011). The third study used an audiovisual speech paradigm in which participants passively watched congruent and incongruent audiovisual syllables. Audiovisual speech was varied in terms of congruency and predictability of presented syllables. Using this paradigm the third study examined how crossmodal predictions errors in speech are reflected in oscillatory activity pattern. Interestingly, the results revealed that frontal theta-band activity (7-8 Hz) was more pronounced when predictions were invalid and preceded by a highly predictive visual syllable. However, this frontal theta activity was stronger in controls compared to patients, reflecting enhanced detection of multisensory prediction error in controls, but not in patients.

Taken together, the results from the three studies are in line with previous research and extend the knowledge about the role of slow oscillations during multi- and uni-sensory processing. Importantly, the studies of this thesis bolster previous research by demonstrating the crucial role of slow oscillations in multisensory processing during the perception of the McGurk illusion in healthy participants (**Study 1**) and patients with schizophrenia (**Study 2**). Furthermore, slow oscillatory activity is involved in altered formation of multisensory prediction error in schizophrenia (**Study 3**). The results from these studies emphasize the relevance for future research in this field, which will help to uncover the significance of oscillations in multisensory integration, perceptual processing and predictive coding in healthy participants and psychiatric disorders, such as schizophrenia.

Several interesting research questions might be raised based on the presented data. One of these potential follow-up studies could involve the application of cognitive-perceptual trainings as a highly valuable approach to modulate deficient multisensory processing. Various psychiatric disorders, such as patients with schizophrenia and autism spectrum disorders have deficits in multisensory emotion and speech processing and could benefit from respective trainings.

In sum, the data presented here provide a first starting point for the investigation of oscillatory brain mechanisms underlying complex multisensory speech perception in healthy participants and psychiatric disorders, such as schizophrenia.

1 Introduction

1.1 Multisensory processing and integration of speech

Research in multisensory processing focuses on the processing of stimuli from at least two different sensory modalities (Stein et al., 2010). Furthermore, multisensory processing subsumes processes such as multisensory integration and many other forms, but does not necessarily specify the nature of interactions between the stimuli from the different modalities. With regard to multisensory integration, this term refers to a process where stimuli from at least two different sensory modalities are combined and bound together into a new percept. Hence, following this definition I will use the term multisensory integration in this thesis to refer to a fusion process, where two sensory inputs are bound together into a unified perception. For instance, such a fusion process is reflected by the McGurk illusion or within the perception of flavor (Stein et al., 2010).

The process of multisensory integration involves perceptual as well as cognitive processes and occurs along temporal and spatial domains (Tseng et al., 2015). During the recording of single neuron activity in the superior colliculus of cats, three basic principles have been found. These principles describe the effect of relative timing of sensory events, spatial stimulus configuration, and stimulus intensity on the outcome of multisensory integration (Stein, Meredith, & Wallace, 1993). However, multisensory integration in speech can also reflect a visual stimulus affecting or enhancing the processing within the auditory modality. Consequently, natural audiovisual speech provides a suitable setting to investigate the influence of visual upon auditory information, due to the natural temporal delay between onset of face-movements and vocal chord vibrations.

With regard to visual speech input, this refers to information based on observing a speakers' mouth and lip movements, as well as tongue and teeth. Articulatory movements of the mouth and the lips naturally precede the onset of the auditory speech by few tens of milliseconds.

Hence, these visual cues provide temporal information about acoustic properties such as sound onset, and help to focus the attention of the observer towards the speaker (Peelle & Sommers, 2015). Beside temporal information, visual speech also carries information about the place of articulation in the vocal tract. This helps to disambiguate similar sounding speech and to identify auditory phonemes (Arnal et al., 2011; Obleser, Lahiri, & Eulitz, 2003). Accordingly, the relevance of visual speech information during speech processing becomes especially evident in noisy environments, when the auditory input is degraded. This has been effectively proven by several behavioral studies (Grant & Seitz, 2000; Ross et al., 2007) using visual speech in combination with degraded auditory speech signals. These studies revealed benefit in participants' speech recognition performance during noise when visual speech information was presented simultaneously compared to the presentation of auditory speech in noise alone. However, the benefits of audiovisual speech are not restricted to situations when the acoustic speech signal is degraded. Likewise it has been evidenced that visual speech sped up auditory processing as reflected by reduced latencies of early auditory ERPs during the presentation of non-degraded congruent audiovisual speech compared to auditory speech alone (Van Wassenhove 2005). A different line of evidence specifically questioned how visual information induces predictions about the upcoming auditory speech signal (Arnal et al., 2011; Desantis, Mamassian, Lisi, & Waszak, 2014). These predictions can be either valid or invalid regarding the subsequent auditory information. In case of an invalid prediction, the expected information does not match the actual outcome and the result of this comparison is the so called prediction error (Friston, 2002; Rao & Ballard, 1999). In speech processing, however, stimuli occur in visual and auditory sensory modalities and predictions need to be generated across the senses. Hence, the concept of speech related crossmodal prediction error (Arnal & Giraud, 2012; Arnal et al., 2011), might be a suitable framework to shed light to the question of how visual speech influences upcoming auditory information and how multisensory integration in speech processing during mismatching stimuli is instantiated. However, there is also the case that predictions are invalid but this audiovisual mismatch is resolved and results in the integration and the formation of a coherent speech perception. The analysis by synthesis theory provides a framework to answer the question why mismatching audiovisual stimuli can result in a coherent percept (Stevens and Halle 1967). According to this theory the saliency of visual information, the attentional weigths towards visual and auditory modality and the redundancy of visual and auditory input modulate the audiovisual speech outcomes (Van Wassenhove 2005). For conflicting visual and auditory information, which might be reconciled into a coherent speech percept, the shift of attentional weighting towards visual modality is essential.

With reference to a specific form of multisensory integration in the domain of speech, the McGurk illusion is an intriguing and well-characterized paradigm to study the impact of visual stimuli on auditory perception. The McGurk effect (McGurk & MacDonald, 1976), depends on the dynamic interplay between vision and audition. It is a powerful demonstration of how auditory information can be shaped by visual information. The McGurk illusion can occur when presenting a lip movement pronouncing a syllable (e.g., /ga/) paired with an incongruent auditory syllable (e.g., /pa/). This pairing of incongruent visual and acoustic information often induces a novel, fused illusory speech percept (e.g., /ka/ in case of a visual /ga/ and an auditory /pa/). Although the McGurk effect occurs quite frequently there is also evidence that the susceptibility of the McGurk effect fluctuates between trials and is different across participants (60-80% illusory trials) (Keil, Müller, Ihssen, & Weisz, 2012; Magnotti & Beauchamp, 2014). Considering the neural substrates of the McGurk illusion most studies arising from EEG and fMRI (Keil et al., 2012; Nath & Beauchamp, 2013) suggest the involvement of superior temporal sulcus (STS/STG) and inferior frontal gyrus (IFG). This is in line with various studies in EEG (Arnal et al., 2009; Keil et al., 2012; Schepers, Schneider, Hipp, Engel, & Senkowski, 2013) and fMRI (Calvert, Campbell, & Brammer, 2000; Sekiyama, Kanno, Miura, & Sugita, 2003; Stevenson & James, 2009), which examined the neuronal substrates of multisensory integration processes in general speech perception. Beside the specific role of the left STS and the IFG these studies also revealed activations in the visual and auditory cortex, during multisensory speech integration. Due to the temporal constraints that influence multisensory processing in speech, methods with high temporal resolution, such as EEG are especially suited to examine the underlying neuronal process reflected by event-related potentials (ERPs) and synchronous neural oscillatory activity. The following section will provided a brief overview about neural oscillations in the human brain and how they are related to different perceptual and cognitive processes.

1.2 Neural oscillations in multisensory speech processing

Recent EEG/MEG studies highlight the crucial role of ongoing neural oscillations that affect our perception. Different brain states of consciousness and attention have been associated with different oscillatory activity pattern (Buzsaki, 2006; Klimesch et al., 2007). Neural oscillations have been measured with MEG/EEG recordings and reflect rhythmic fluctuations in neural membrane potential caused by summed post-synaptic potentials (Lopes da Silva, 1991). As such, neural oscillatory activity is a phenomenon that can be measured at different scales in the human brain, such as microscopic (postsynaptic single unit recordings) and macroscopic (postsynaptic multiunit recordings, i.e. local field potentials). The measured brain signals contain various frequencies that can be analyzed by application of frequency-domain methods, such as Fourier Transformation, which decomposes the frequency content of a complex signal. Thus, neural oscillations can be described by measures as frequency, amplitude (power) and phase. Traditionally, neural oscillations have been characterized by their frequency band and were labeled accordingly. With respect to their frequencies oscillations were differentiated in slow (frequency < 30 Hz) and fast (frequency > 30 Hz) oscillatory activity. Slow oscillations comprise various different frequency bands such as delta (0.1-3 Hz), theta (3-7 Hz), alpha (8-12 Hz) and beta (13-30 Hz)-band activity. Whereas fast oscillations comprise gamma (30 -250 Hz), that can be differentiated in low (30-60 Hz) and high gamma (60-120 Hz) and ultrahigh frequencies above 120 Hz. However, this clearcut taxonomy in the lower frequency bands has been challenged (Buzsaki, 2006).

The role of synchronous neural oscillations in sensory processing has been emphasized for the first time by the “binding by synchrony” hypothesis (Gray & Singer, 1989). In uni-modal processing the “binding by synchrony hypothesis” represents a theoretical approach to explain perceptual feature binding and highlights the critical role of high frequency oscillations. Following this hypothesis, other theories emerged, which emphasized the importance of synchronized oscillations for cortical communication between different brain areas (Fries, 2005). The “communication through coherence” hypothesis assumes the coordinated interaction between oscillations and provides a crucial approach to elucidate how the brain works. Further this framework proposes how oscillatory activity can be exploited as a flexible mechanism for the interaction and communication between different neuronal groups. Based on the properties of neural oscillations, which presumably reflect up and downshifts of postsynaptic potentials, neural oscillations serve as a mechanism inducing variations in neural excitability (Bishop 1933). These variations in excitability of neurons represent rhythmically occurring windows for communication, which are limited to coherently oscillating neuronal groups.

1.2.1 Evidence for different functions across neural oscillations

Previous research has demonstrated that synchronized neural oscillations in humans play a major role in basic sensory processing in various modalities such as vision (Klimesch, Fellinger, & Freunberger, 2011), audition (Weisz, Hartmann, Müller, Lorenz, & Obleser, 2011), somato-sensation (Baumgarten, Schnitzler, & Lange, 2014; Schulz et al., 2015) and the chemical senses (Jung et al., 2006). Furthermore, neural oscillatory activity has been linked to various higher cognitive functions, such as feature binding (Tallon-baudry & Bertrand, 1999), working memory (Sauseng, Griesmayr, Freunberger, & Klimesch, 2010), multisensory integration (Senkowski et al., 2008) and conscious awarness (Engel & Singer, 2001). In this thesis, all three studies focus on macroscopic level measuring local field potentials, as measured in EEG, which reflect synchronous activity of thousands of neurons. Further, the focus of this thesis is on slow oscillatory activity comprising the frequencies from 4-30 Hz. Previous evidence revealed that narrow-banded slow oscillations serve different functions. For instance, delta oscillations were related to different sleep stages (Dang-vu et al., 2008) and are important for memory consolidation during sleep (Piantoni, Werf, Jensen,

& Someren, 2015). Next, theta oscillations have been found in demanding tasks (working memory, stroop task), associated with cognitive control functions (Cavanagh & Frank, 2014) and top down processing. Alpha oscillations have been mostly studied in the visual system (Bareither, Chaumon, Bernasconi, Villringer, & Busch, 2014; Freunberger, Klimesch, Griesmayr, Sauseng, & Gruber, 2008; Klimesch et al., 2011). However more recent studies affirm the role of alpha oscillation as an inhibitory mechanism which gates perception (Klimesch et al., 2007). Further evidence points out the role of alpha oscillations also in different modalities such as audition (Leske et al., 2014; Müller et al., 2013; Strauß, Wöstmann, & Obleser, 2014; Weisz et al., 2011). Traditionally beta oscillations have been associated with motor functions (Pfurtscheller, Zalaudek, & Neuper, 1998). However, more recently evidence suggest that beta oscillations are also involved in top down processing (Engel & Fries, 2010) and relevant for language and speech processing (Arnal et al., 2011; Bastiaansen & Hagoort, 2006). Finally, high frequency oscillations such as gamma have been classically investigated in active stimulus encoding, selective attention (Jensen, Kaiser, & Lachaux, 2007), multisensory (Mishra, Martinez, Sejnowski, & Hillyard, 2007; Schneider, Lorenz, Senkowski, & Engel, 2011) and speech processing (Schepers, Yoshor, & Beauchamp, 2014).

1.2.2 Oscillatory activity in multisensory speech processing:

Neural synchrony serves as a potential mechanism to segregate and integrate the neuronal activity of different brain areas. This is essential for the processing of sensory stimuli from different sensory modalities. For this reason, it is likely that neural synchrony also plays a crucial role for multisensory integration (Fries, 2005). Within the broad research field of multisensory integration there is a growing interest in elucidating the underlying neuronal mechanisms of multisensory processing. With respect to multisensory speech processing, recent evidence points to the relevance of slow oscillations, which might be suited to establish a communication mechanism between brain areas, such as auditory and visual cortex (Fingelkurts, Fingelkurts, & Krause, 2003; Palva & Palva, 2012; Von Stein &

Sarnthein, 2000). In a seminal work, Von Stein and Sarnthein (2000) showed that beta-band oscillations might serve as a communication mechanism between distant cortical areas during multimodal semantic processing. Similarly, Keil et al. (2012) found that beta-band phase synchrony (coupling) between left STG and fronto-parietal areas is increased during the perception of the McGurk illusion. Furthermore, other studies demonstrated that beside beta-band, also alpha (Driel, Knapen, Es, & Cohen, 2014; Fingelkurts et al., 2003) and theta-band (Phillips, Vinck, Everling, & Womelsdorf, 2014) oscillations contribute to communication between cortical areas. For instance, Fingelkurts (2003) reported a denser functional coupling of alpha-band oscillations during the presentation of incongruent stimuli compared to congruent speech stimuli. Together these strands of evidence underline the functional role of slow oscillations, such as beta and alpha oscillations for the inter-areal communication between distant cortical networks during multisensory speech processing and integration. The importance of network connections becomes most evident by the assumptions of the dysconnection hypothesis (Stephan, Friston, & Frith, 2009), which state that psychiatric disorders, such as schizophrenia and autism, presumably arise from disrupted brain connectivity. Hence, the next section will focus on basic evidence for neural markers in schizophrenia, which emphasize the role of synchronous neural oscillations in this patient group.

1.3 Altered sensory processing and oscillations in schizophrenia

Schizophrenia is severe, chronic disorder with present point prevalence of 0.4-0.5 % worldwide (Saha, Chant, Welham, & Mcgrath, 2005). The typical onset is at adulthood, however schizophrenia is recognized as a lifetime disorder, which starts to develop even in infancy and manifests during adulthood (Nasrallah, Tandon, & Keshavan, 2015). Schizophrenia is characterized by prominent symptoms such as delusions, hallucinations, incoherent thinking, disorganized speech and blunted affect. Traditionally, these symptoms have been classified as positive and negative syndrome clusters (Kay, Fiszbein, & Opler, 1987). However, more recently the role of cognitive and social dimensions within this disorder have been highlighted (Bowie et al., 2009). Since symptoms such as delusions, hallucinations and incoherent thinking vastly derogate the quality of life from patients with schizophrenia, clinical and psychiatric research has targeted the underlying biological and neuronal mechanisms, which contribute to this disorder. Within the past ten years several potential biomarkers, such as changes in fMRI (Hazlett, Goldstein, & Kolaitis, 2012; Horga, Schatz, Abi-dargham, & Peterson, 2014), EEG activity (Senkowski & Gallinat, 2015a; Uhlhaas & Singer, 2015) and neurotransmitter (Howes et al., 2012; Taylor & Tso, 2014) pattern have been investigated. Notably, neural oscillations have been proposed as a very promising biomarker, which seem to be disturbed in patients with schizophrenia (Uhlhaas & Singer, 2006). More recently, the crucial role of altered neural oscillations in higher cognitive processing within various tasks, such as working memory (Senkowski & Gallinat, 2015a) and interference control (Popov, Wienbruch, Meissner, Miller, & Rockstroh, 2015) have been demonstrated. To substantiate the role of EEG correlates as potential biomarker in schizophrenia I will present a brief overview of findings in ERPs and neural oscillations in uni- and multisensory processing in schizophrenia in the following section.

1.3.1 Neural markers for unisensory processing deficits in schizophrenia

A significant body of evidence obtained from EEG studies reported that patients with schizophrenia have various deficits unisensory stimulus processing, specifically investigated in ERPs (Butler et al., 2007; Rosburg, Boutros, & Ford, 2008). In particular, early ERPs in patients with schizophrenia show significantly attenuated amplitudes during basic auditory (Javitt, 2009) and visual processing (Knebel, Javitt, & Murray, 2012). Dysfunctions in the auditory modality of patients with schizophrenia are well-documented and comprise auditory ERP components such as auditory P50, N1 and Mismatch Negativity (MMN). Specifically, the absence of modulation in auditory P50 during the paired-click paradigm have been associated with impaired sensory gating and inhibitory processing (Brockhaus-Dumke, Mueller, Faigle, & Klosterkoetter, 2008). Further, a widely replicated finding is the reduction of the auditory N1 amplitude, which has been interpreted as reduced auditory attention in schizophrenia (Boutros et al., 2009; Rosburg et al., 2008; Turetsky, Bilker, Siegel, Kohler, & Gur, 2009). Together, these alternations in auditory P50 and N1 might reflect a breakdown in early auditory processing in patients with schizophrenia. In turn, evidence for auditory dysfunctions reflected by altered neural oscillations in schizophrenia is sparse. Popov et al. (2011) found that sensory gating deficits in schizophrenia are associated with reduced fronto-temporal alpha-band and frontal gamma-band power after presentation of the second tone during a paired click-paradigm. Using the same paradigm similar findings of reduced frontal gamma-power have been demonstrated in patients with schizophrenia (Leicht et al., 2010). Those findings have been interpreted as reduced sensory gating and highlight the role of neural oscillations in basic auditory processing. Moreover, also visual dysfunctions, such as impairments in object recognition (Doniger, Foxe, Murray, Higgins, & Javitt, 2002), face processing (Earls, Curran, & Mittal, 2015), backward masking (Butler & Javitt, 2007) and perceptual grouping (Sehatpour, Molholm, Javitt, & Foxe, 2006) in schizophrenia have been characterized by alternations in ERPs components such as the visual P1, N1 and N170. A growing body of evidence highlights the role of altered high frequency oscillation contributing to visual dysfunctions in schizophrenia. For instance, studies investigating perceptual grouping during the detection of Mooney faces showed reduced high (60-140 Hz) gamma-band power in schizophrenia (Grützner, Wibral, Sun, et al., 2013; Uhlhaas, Phillips, Mitchell, & Silverstein, 2006). These findings were similar to those by Spencer, Niznikiewicz, Shenton & McCarley (2008), who found reduced gamma phase-locking during visual oddball processing in this patient group. Further, studies investigating facial affect recognition (Popov et al., 2014) and processing of faces in schizophrenia (Lynn & Salisburg, 2009) revealed impairments in oscillatory activity. During visual processing of dynamically evolving emotional face stimuli, patients showed a reduced alpha-band power and pre-stimulus alpha-band connectivity, which reflects an abnormal pattern of alpha-band activity compared to matched controls (Popov et al., 2014). These findings have been linked to abnormal dynamics during facial affect processing in schizophrenia. Taken together, given the early unisensory processing deficits indicated by ERP and neural oscillation studies it is conceivable that multisensory processing, that relies on the early auditory and visual sensory processing inputs might be also impaired.

1.3.2 Neural markers for multisensory processing deficits in schizophrenia

With regard to behavioral experiments investigating multisensory processing in schizophrenia, the majority of studies used audiovisual stimuli, such as simple beep and flashes (Boer-schellekens, Stekelenburg, Pieter, Gool, & Vroomen, 2014; Gelder, Vroomen, Annen, Masthof, & Hodiamont, 2002a; Williams, Light, Braff, & Ramachandran, 2010) or audiovisual speech stimuli (Pearl et al., 2009; Ross et al., 2007). In the following I will specifically focus on studies in multisensory speech processing schizophrenia.

In a behavioral study Ross et al. (2007) demonstrated a diminished benefit of viewing lip movements for the recognition of auditory speech that was most obvious at an intermediate noise level in patients with schizophrenia. A further prominent audiovisual speech paradigm, that offers the opportunity to study multisensory integration in schizophrenia is the McGurk effect. Previous studies have reported reduced McGurk-illusion rates in schizophrenia (Gelder, Vroomen, Annen, Masthof, & Hodiamont, 2002b; White et al., 2014). However, other studies did not find group differences (Martin, Giersch, Huron, & Wassenhove, 2013; Myslobodsky, Golderberg, Johnson, Hicks, & Weinberger, 1992). These inconsistencies in findings could be due to differences in defining criteria for the illusion perception, group ages (Pearl et al., 2009), and heterogeneity of the clinical samples. Evidence for multisensory speech processing in schizophrenia on neural level (ERPs and oscillations) is very sparse. To date, only one study examined multisensory speech processing on a neural level (ERPs) in patients with schizophrenia (Stekelenburg, Maes, Van Gool, Sitskoorn, & Vroomen, 2013). In this study alternations in multisensory speech processing in schizophrenia have been evidenced, as reflected by ERPs at the P2 latencies, using audiovisual speech stimuli. The authors interpreted their ERP findings in the light of multisensory processing deficit in this patient group.

Together these strands of evidence, point towards the potential significance of synchronous neural oscillations in multisensory speech processing and integration in schizophrenia. To date, however only little research has been conducted on the relevance of neural synchrony for multisensory integration in schizophrenia. Thus far, the precise mechanism how oscillatory activity at different frequencies across uni- and multisensory processing area is involved in multisensory speech processing in patients with schizophrenia remains elusive.

1.4 Overview of studies

The current thesis addressed some recent research questions that are derived from the literature and findings outline above:

1. What are the oscillatory mechanisms contributing to the perception of the McGurk illusion in healthy participants?
2. Are the oscillatory mechanism contributing to the perception of the McGurk illusion altered in patients with schizophrenia?
3. What are the spectral signatures of crossmodal audiovisual prediction error during multisensory speech processing in schizophrenia?

The first study aimed to investigate whether incongruent visual speech information can influence the upcoming auditory perception. I used the McGurk illusion, as an example of multisensory integration in speech, where visually induced predictions are violated but still result in fused coherent new percept .The experiment was designed to compare the subjective illusory perception during the incongruent McGurk stimuli with congruent syllables. Thereby, we explicitly investigated neural signatures of the subjective perception during the McGurk illusion and congruent condition by keeping the physical input of the auditory signal identical. The design of the McGurk paradigm was kept identical for the first and the second study. In first as well as in second study the participants were asked to rate their subjective perception. Further, the participants were instructed to explicitly not solely focus on the visual or auditory modality, but rather to consider both for their perception. It was hypothesized that slow oscillatory activity in auditory, visual and multisensory processing areas contributes to the formation of the McGurk illusion. In the comparison between congruent and illusory trials, we found a stronger suppression in beta-band power during the McGurk illusion compared to congruent control stimuli in early and late time intervals at fronto-central and left temporal electrode sites.

Early beta-band suppression was found to be stronger during the McGurk illusion compared to congruent condition. This is in line with previous evidence (Engel & Fries, 2010) and might indicate increased processing demand in the McGurk illusion at this interval

Most importantly, it could be shown, that stronger beta-band suppression during the McGurk illusion contributed to the illusory percept formation. This finding in the late beta-band is consistent with previous research indicating the importance of beta-band oscillations in top-down processing (Arnal & Giraud, 2012) and multisensory integration (Hipp, Engel, & Siegel, 2011). The results from this study furnish first evidence that low oscillatory activity, namely beta-band oscillations, operate at distinct processes at early (incongruence monitoring) and late (integration and illusory percept formation) stages during the perception of McGurk illusion.

Hence, if low oscillatory activity, as reflected by beta-band oscillations, was involved in the formation of an illusory percept and if this had an influence on the perception of the McGurk illusion in healthy participants, this would lead to the next question: Are slow oscillations altered in patients with schizophrenia during the perception of the McGurk illusion? For this purpose, the second study used the identical McGurk paradigm to investigate differences in oscillatory activity during the McGurk illusion in patients with schizophrenia and matched controls. Surprisingly, we found no behavioral differences in McGurk illusion rate between patients and matched controls. However, this study added evidence that differences in slow oscillatory activity, especially in the late medio-central alpha-band power, reflect altered auditory processing during the McGurk illusion in patients. In particular, alpha-band suppression was enhanced in matched controls during the processing of congruent stimuli compared to the McGurk illusion, which is in line with the notion that reduced alpha-band power reflects (Klimesch, 2012) active sensory processing. In contrast, this pattern was reversed in patients with schizophrenia and the differences in alpha-band power were most prominent in McGurk illusion condition. Hence, this finding underlines altered audiovisual processing in the McGurk illusion in schizophrenia, namely an altered influence of the visual context on auditory processing. Even though, this was not manifested in behavioral differences. In conjunction, the results from the first and second study suggest that slow oscillatory activity is involved in the formation of the McGurk illusion. Additionally, alpha-band power reflects a specific pattern of oscillatory activity, which differentiates between patients with schizophrenia and matched controls.

The third experiment was designed to extend the findings from the first and the second study and thereby investigate the influence of differences in predictability of visual speech on auditory perception in schizophrenia. To this end, patients and matched controls were presented with high and low visual syllables paired with congruent and incongruent auditory syllables. Further, to warrant the attention of the participants, auditory and visual catch trials were included. It was hypothesized based on previous evidence (Arnal et al., 2011) that highly predictive visual information induces a stronger crossmodal prediction error compared to low predictive visual information when paired with an incongruent auditory syllable. Therefore, the influence of high and low predictive visual speech on congruent and incongruent auditory syllables on oscillatory activity in patients with schizophrenia and matched controls was compared. In line with the notion that an increase of frontal theta-band power reflects increased top-down processing in healthy participants (Cavanagh & Frank, 2014; Cohen & Donner, 2013), it could be shown that frontal theta-band power was strongly increased in matched controls during the processing of high compared to low predictive visual syllables. Importantly, this increase in frontal theta-band power was diminished in patients with schizophrenia. These findings provide first evidence that patients with schizophrenia show a crossmodal top-down processing deficit, reflected by an altered crossmodal prediction error processing. By revealing the differential influence of visual prediction strength on auditory processing in patients and matched controls, the third study extended the previous findings and demonstrated for the first time a crossmodal prediction error processing deficit in schizophrenia patients.

Based on these three studies I conclude that multisensory speech processing indeed depends on low oscillatory activity. Moreover, low oscillatory activity might reflect a neural marker that differentiates patients with schizophrenia and matched control during multisensory speech processing. Depending on the task and experimental manipulation it could be shown that invalid visually induced predictions upon auditory speech information can either result in the perception of a fused perception (McGurk illusion) or induces a crossmodal prediction error. Both tasks have in common that visual predictions are invalidated. However the resulting processes induced by the invalid predictions vary between the two tasks. In case of the McGurk illusion a perceptual recalibration is necessary to form a coherent perception and resolve the audiovisual mismatch. In contrast, during the crossmodal prediction error the resolution of erroneous predictions and their update is mandatory. Interestingly, these both tasks involve different types of low oscillations, such as alpha and beta-band oscillations (McGurk illusion) and theta-band (crossmodal prediction error) oscillatory activity. Together, the present findings extend the role of slow oscillations not only in the context of multisensory speech processing but also point towards their potential role as a mechanism for establishing inter areal connection. This is especially necessary for the interplay between auditory, visual, multisensory processing and frontal processing areas during multisensory processing.

2 First study: Early and late beta-band power reflects audiovisual perception in the McGurk illusion.

Roa Romero, Y.; Senkowski, D. and Keil, J., (2015). Early and Late Beta Band Power reflects Audiovisual Perception in the McGurk Illusion. *Journal of Neurophysiology,13* (7)*:*2342-2350. http://dx.doi.org/10.1152/jn.00783.2014

Introduction

Multisensory perception, as exemplified in the McGurk illusion (McGurk & MacDonald, 1976), depends on the dynamic interplay between vision and audition. It is a powerful demonstration of how auditory information can be shaped by visual information. The McGurk illusion can occur when presenting a lip movement pronouncing a syllable (e.g., /ga/) paired with an incongruent auditory syllable (e.g., /ba/). This pairing of incongruent visual and acoustic information often induces a novel, fused speech percept (e.g., /da/ in case of a visual /ga/ and an auditory /ba/). Recent studies on multisensory perception focused on neural synchrony as a potential mechanism to segregate and integrate the neural activity of different brain areas (Senkowski et al., 2008; Van Atteveldt, Murray, Thut, & Schroeder, 2014). How exactly neural synchrony may contribute to the McGurk illusion is not well understood.

Thus far, human electrophysiological (EEG) studies have primarily focused on the temporal characteristics of multisensory processing, as reflected in event-related potentials (ERPs) (Besle, Fort, Delpuech, & Girad, 2004; Stekelenburg, Maes, et al., 2013; Stekelenburg & Vroomen, 2007) and magnetoencephalography (MEG; Arnal et al., 2009). For instance, Besle et al. (2004) found that the faster identification of audiovisual syllables was paralleled by a reduction of the N1 amplitude in the supra-temporal auditory cortex. In a similar vein, Stekelenburg and Vroomen (2007) showed that the amplitude of the N1 is suppressed under audiovisual stimulation when the visual stimulus (e.g. lip movement) predicts the onset of the auditory stimulus (e.g. acoustic syllable). Notably, this suppression was found irrespective of the audiovisual congruence. The authors suggested that this N1 suppression effect reflects a reduction of signal uncertainty. This effect was replicated by further MEG (Arnal et al., 2009) and EEG (Wassenhove, Grant, & Poeppel, 2004) studies. Taken together, these findings demonstrate that visual information impacts auditory processing at early processing stages and that a suppression of short-latency ERPs could reflect a facilitation of auditory processing.

Aside from early multisensory interactions, as reflected in time-locked ERPs, neural synchrony might serve as a potential mechanism to segregate and integrate neural activity across the different sensory modalities and might therefore be critically involved in the McGurk illusion (Keil et al., 2012). Previous evidence based on multimodal semantic processing showed that neural oscillations in the beta band (13-30Hz) might serve as a communication mechanism between distant cortical areas (Von Stein & Sarnthein, 2000). Moreover, Fingelkurts et al. (2003) investigated functional coupling between EEG electrodes in the alpha and beta band during an oddball paradigm composed of standard congruent and deviant McGurk type stimuli. A main finding of this study was a denser coupling in the beta band during the presentation of the McGurk type stimuli compared to the congruent standards. In another study, Lange, Christian, & Schnitzler (2013) showed that post-stimulus beta band power is suppressed during the processing of incongruent compared to congruent audiovisual speech stimuli. Hence, these studies suggest that oscillatory responses, especially in the beta band, play an important role for integrative audiovisual speech processing, including the McGurk illusion.

In the current high-density EEG study, we investigated the McGurk illusion as a phenomenon of speech-specific multisensory integration. We explicitly investigated the neural signatures of the subjective percept in the McGurk illusion. This goes beyond previous studies that have examined the neural signatures of the McGurk illusion in terms of stimulus incongruence and mismatch processing (Saint-Amour, De Sanctis, Molholm, Ritter, & Foxe, 2007). To this end, we compared the neural responses to objectively congruent syllables with incongruent audiovisual stimuli that elicited a subjectively congruent percept (i.e. the McGurk illusion). Importantly, the auditory syllables were identical in these stimuli. This enabled us to directly compare the neural responses to congruent audiovisual syllables and syllables that induced the McGurk illusion. We focused our analysis on ERPs and oscillatory responses, especially in the beta band. Our study showed that the post-stimulus suppression of beta band power was stronger during McGurk illusion compared to congruent stimuli.

Methods

Participants

Twenty-five subjects (mean age 31.04 years, range 18-51 years; 12 females) participated in the study. All participants gave written informed consent, were right-handed, and had normal or corrected to normal vision. Moreover, all participants had normal hearing and no record of neurological disorders. Six participants were excluded from the analysis because they did not show the McGurk illusion (i.e. the number of perceived illusion trials was less than 15 %).The data of the remaining nineteen participants (mean age 31.84 years, range 18-51 years; 10 females) were used for the statistical analyses. The study was approved by the Medical Ethics Committee of the Charité- Universitätsmedizin Berlin and was conducted in accordance with the Declaration of Helsinki.

Stimuli

Video clips of a female actress uttering the syllables /pa/, /ga/, and /ka/ were recorded using a digital camera (Canon 60D, 50 frames per second, 1280x720px, 44.1 Khz stereo audio) and exported at 30 frames per second (Apple Quicktime Player, Version 7). Video sequences were taken in frontal view displaying the face (visual angle = 5.95° x7.36°), in front of a black background. The clips of all syllables were equalized with respect to luminance using the SHINE toolbox (Willenbockel et al., 2010). In order to minimize eye movements during the experiment a small white fixation cross above the mouth at the philtrum was added to all clips. The original sound files, which were recorded at 44.1 Khz, were off-line downsampled to 11.025 Khz and band pass filtered (300 – 3400 Hz, 4th order two-pass Butterworth filter). The syllables were presented with the real audiovisual onset delay. Specifically, the auditory syllables started 337 ms (/pa/), 336 ms (/ga/), and 630 ms (/ka/) after lip movement onset. The duration of the visual motion was 858 ms (/pa/), 891 ms (/ga/), and 1221 ms (/ka/). Auditory syllables had a duration of 329 ms (/pa/), 360 ms (/ga/), and 394 ms (/ka/) and were presented at 30dB (SPL).

Experimental design

The experiment consisted of 750 trials that were presented in 10 blocks, composed of 75 trials each. Each block had a duration of about 5 minutes and all stimuli were presented via PsychToolbox (https://www.psychtoolbox.org/). Visual stimuli were displayed on a 21-inch CRT screen at a distance of 1.2 m and auditory stimuli were presented via a single centrally positioned speaker (Bose Companion® 2). During the experiment different types of congruent and incongruent audiovisual syllable trials were presented. Congruent syllable trials contained matching audiovisual syllables (N = 300) (e.g., visual /pa/and auditory /pa/, visual /ga/ and auditory /ga/, visual /ka/ and auditory /ka/), whereas incongruent syllable trials contained non-matching audiovisual syllables (N = 450) (visual /pa/ and auditory /ka/, visual /ka/ and auditory /pa/, visual /pa/ and auditory /ga/, visual /ka/ and auditory /ga/, visual /ga/ and auditory /ka/). Our pilot data showed that the combination of a visual /ga/ and an auditory /pa/ is often perceived as an illusory syllable /ka/, i.e. the so-called McGurk illusion. In the following, we will refer to this syllable combination when the resulting perception is either "/ka/" or "something else" as 'McGurk illusion trials'. In total 300 McGurk trials were presented. In addition to the congruent and McGurk trials, 150 incongruent syllable combination trials were presented. These trials served as control stimuli to ensure that the McGurk illusion was specific to the McGurk trials and not a result of an arbitrary audiovisual mismatch. In each trial the first static frame of the video clips was presented for a random interval ranging from 1000 to 1500ms (mean = 1250 ms), in order to minimize expectancy effects and to control for the influence of visual ERPs due to picture onset. Following the video clip, the last frame of each clip, where the mouth of the actress was closed, was presented for 1000 ms. During this time the fixation cross turned into a question mark for 500 ms at a random time point and participants were required to indicate by a button press with the index, middle, ring or small finger of their right hand whether they had perceived the syllable /pa/, /ga/, /ka/, or "something else", respectively. We would like to emphasize that the "something else" category was not reflecting the perceived incongruence per se, but also possible other percepts such as /ta/ or /bga/, as verbally reported by the participants after the experiment. All trials had a total duration of 3700-4200 ms (see Figure 1).

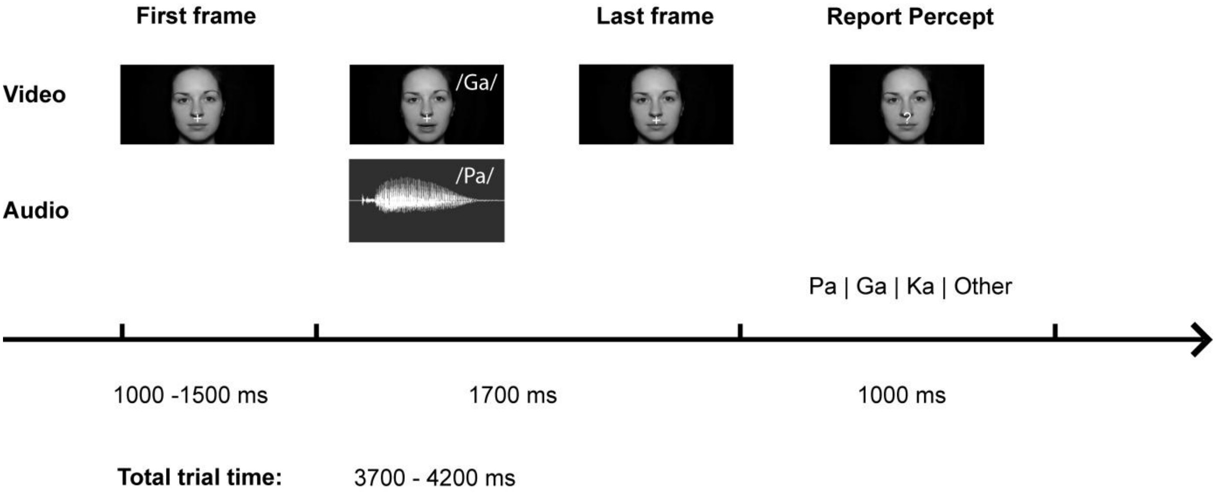

Figure 1

Trial and timing overview: Example McGurk trial with video frames of the syllable /ga/ (top row) and audio trace of the syllable /pa/ (middle row) used in this experiment. After the last frame, subjects were asked to indicate the perceived percept via button press. This response time window was indicated by the presentation of a white question mark with a randomly jittered onset.

Acquisition and preprocessing of EEG data

Data were recorded using a 128 channel active EEG system (EasyCap, Herrsching, Germany), with two eye electrodes to monitor eye movements. Recordings were made against nose reference with a pass band (0.016-250 Hz) and digitized at a sampling rate of 1000 Hz. Preprocessing and offline data analysis were performed using EEGlab (Delorme & Makeig, 2004), Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011), and custom-made Matlab scripts (MathWorks, Natick, MA). Continuous data were high-pass (1Hz, FIR), low-pass (125 Hz, FIR), and notch (49.1-50.2Hz, 4th order two-pass Butterworth filter) filtered. In addition, data were down-sampled to 500 Hz. For the data analysis epochs of 4 s (-1 to 3 s) around sound onset were extracted. Firstly, epochs containing muscular artefacts were rejected by visual inspection. Subsequently, trials containing remaining artifacts of amplitudes ± 100 µV were rejected automatically. In order to correct for EOG and ECG artifacts, independent component analyses were conducted (extended runica; (Lee, Girolami, & Sejnowski, 1999). On average 11.84 ± 2.69 independent components were rejected after visual inspection. Finally, the remaining noisy channels were interpolated using spherical interpolation (on average 17.15 ± 6.21 channels) and the epoched data were re-referenced to common average.

Prior to the calculation of the ERPs all epochs were high-pass filtered at 2 Hz (2nd order two-pass Butterworth filter), low-pass filtered at 35 Hz (12th order two-pass Butterworth filter), and baseline corrected using an interval from -500 ms to -100 ms before the onset of the sound. For the time-frequency analysis of lower frequency oscillatory responses (i.e. 4-40 Hz) multi-taper convolution transformation with frequency depending Hanning window was computed in 2 Hz steps (time window $\Delta t = 5/f$, spectral smoothing: f = $1/\Delta t$). For the analysis of higher frequency oscillatory responses (i.e. 40-100 Hz) Slepian tapers (fixed time window t = 0.2 s, fixed spectral smoothing: f = 10 Hz) were applied. Averaged oscillatory activity was baseline corrected (relative change) from -500 to -100 ms prior to sound onset.

Statistical analysis

In the analysis, McGurk trials (visual /ga/, auditory /pa) were directly compared with congruent control trials (visual /pa/, auditory /pa/). Importantly, the auditory syllable (i.e. /pa/) was identical in McGurk and congruent control trials, whereas there were slight differences between the visual inputs of McGurk and congruent control trials. We also compared McGurk trials with congruent syllable combinations (i.e. visual /ga/, auditory /ga/, and visual /ka/, auditory /ka/). The comparison between congruent /ga/ /ga/ and McGurk illusion trials was done to elucidate the effect of varying auditory stimulation (i.e. auditory /ga/ compared to auditory /pa/) together with identical visual stimulation (i.e. visual /ga/). The comparison between congruent /ka/ /ka/ and McGurk illusion trials was done to elucidate the effect of the same percept (i.e. percept /ka/ following congruent and incongruent audiovisual stimulation). In order to elucidate the effect of audiovisual mismatch we also compared the McGurk illusion trials to all incongruent audiovisual trials. The analysis of behavioral data focused on the relative proportion of McGurk illusions in the McGurk trials and on the correct identification of congruent control trials. Reaction tendencies were calculated as a relative proportion of illusion, audio percept, and visual percept responses in all McGurk trials (Keil et al., 2012). To account for the violation of normal distribution in reaction tendencies and the possibility of skewed distributions, a non-parametric Friedmann ANOVA with dependent variable reaction tendency (three-levels: rate of illusion, audio and visual percept) was calculated. Additionally, we performed follow-up Wilcoxon tests. The EEG data analysis focused on the comparison of ERPs and oscillatory responses to McGurk illusion trials and congruent control trials. The number of trials was equalized according to the lower trial number of both stimulus categories (i.e. of McGurk illusion trials or congruent control trials). On average, for each condition 71 trials were entered into the analysis. The differences in amplitudes and power between McGurk illusion and congruent control trials were statistically compared by means of a cluster-based permutation test (Maris & Oostenveld, 2007). Statistical analyses were calculated separately for lower (4-40 Hz) and higher (40-100 Hz) frequency ranges. In order to examine whether any possible effects are primarily driven by the incongruent audio-visual stimulation (i.e. visual /ga/ and auditory /pa/) and not due to the multisensory fusion process that leads to the McGurk illusion, similar analysis for ERPs and oscillatory power were calculated for the 6 subjects that were excluded because they did not show the McGurk illusion (i.e. non-perceivers with an illusion rate smaller than 15 %).

Results

Behavioral results

The recognition rate of congruent trials (visual /pa/, auditory /pa/) was 95.11 % showing reliable recognition of congruent control syllables. In the McGurk trials (visual /ga/, auditory /pa/) participants reported an illusory percept in 84.2 % of all trials. In 49.5 % this syllable combination was perceived as a /ka/ and in 34.5 % participants indicated that they perceived "something else". Perception of unimodal dominance of either auditory (/pa/) or visual (/ga/) stimulus components was reported in 12.3 % and 3.5 % of the McGurk trials respectively. The non-parametric Friedmann ANOVA revealed a significant main effect ($\chi^2(2) = 27.70$, $p < .001$). Follow-up Wilcoxon tests showed that participants reported the illusory percept (i.e. /ka/ or "other syllable") more frequently than the auditory (i.e. /pa/; $Z = -3.70$, $p < .001$) or the visual percept (i.e. /ga/; $Z = -3.85$, $p < .001$). In addition, the difference in reaction tendencies between auditory and visual percepts reached significance, with a higher rate in auditory compared to visual percepts ($Z = -2.34$, $p = .015$, see Figure 2).

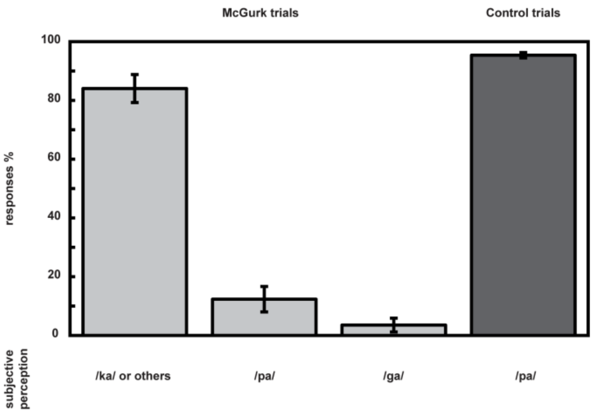

Figure 2

Behavioral results. Gray shaded columns (1-3) with number of responses of the different subjective percepts relative to total number of McGurk trials (visual /ga/ and auditory /pa/). A McGurk illusion (/ka/ or /others/) was perceived in 84.2 %, whereas auditory (/pa/) and visual (/ga/) perceptions were reported in 12.3 % and 3.5 % of the McGurk trials. Column 4 shows percentage of correctly identified /Pa/-syllables during control trials (congruent /pa/- /pa/). On average, control trials were correctly identified in 95.11%.

Event-related activity

In order to examine the processing of identical auditory stimuli within congruent and incongruent videos evoking illusory percepts, stimulus-evoked activity between McGurk illusion trials (i.e. McGurk trials in which participants reported an illusion) and congruent control trials was compared. The cluster-based permutation test revealed one significant positive and one negative cluster. The negative cluster (p = .008) reflected a larger negative deflection at central and parietal electrodes for congruent control trials compared with McGurk illusion trials. The cluster was found at the interval of the auditory evoked N1 component (i.e. 78-171 ms). The positive cluster (p = .02) reflected a larger positive deflection at central and parietal electrodes for the congruent control trials. It was found at the interval of auditory evoked P2 component (i.e. 172-208 ms) (Fig. 3).

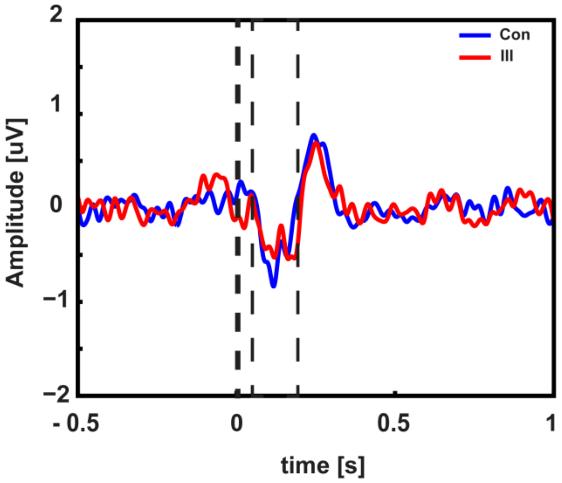

Figure 3

Time course of event-related activity: Event-related activity relative to auditory onset at central electrodes for trials in which a McGurk illusion was reported (red) and congruent trials (blue). Significant differences in time course are marked by dashed line. Trials in which a McGurk illusion was perceived elicited a reduced N1 component compared to congruent trials in which the syllable was correctly identified.

Additionally, we compared congruent /ga/ /ga/ and congruent /ka/ /ka/ trials with McGurk illusion trials to account for the effect of the identical visual stimulation (visual /ga/, auditory /ga/ vs. visual /ga/, auditory /pa/) and the effect of the same percept (visual /ka/, auditory /ka/ vs. illusory /ka/ perception for the auditory /ga/ visual /pa/ combination). For congruent /ga/ /ga/ and McGurk illusion ERPs we found significant differences in the N1 (p < .001, 112-172 ms) and P2 ($p < .001$, 172-234 ms) component at central electrodes. Congruent /ga/ /ga/ trials showed a stronger negative deflection than McGurk illusion trials. Similar differences in N1 ($p = .005$, 86-149 ms) and P2 ($p = .003$, 167-278 ms) were found for the comparison of congruent /ka/ /ka/and McGurk illusion ERPs at central electrodes.

Further, in order to control for the effect of incongruence we compared ERPs for incongruent and McGurk illusion trials. The analysis of ERPs revealed 2 significant positive clusters and 2 significant negative clusters. The first positive ($p < .001$, 85-176 ms) and negative cluster ($p < .001$, 85-173 ms) reflected a stronger N1 for incongruent trials compared with McGurk illusion trials. In addition, the second positive cluster ($p = .002$, 175-230 ms) reflected a stronger P2 component at fronto-central and parietal electrodes for incongruent trials compared with McGurk illusion trials. The second negative cluster ($p = .001$, 265-386 ms) reflected a stronger late negative deflection at central and parietal electrodes for incongruent trials compared with McGurk illusion trials.

Moreover, we calculated ERPs for all incongruent trials and the congruent control (/pa/-/pa/) trials. The analysis of ERPs revealed 2 significant positive clusters and 1 significant negative cluster. The first positive cluster ($p = .004$, 330-410 ms) reflected a stronger late positive deflection at occipital electrodes for congruent control trials compared with incongruent trials. The second positive cluster ($p = .005$, 260-320 ms) reflected a stronger late positive deflection at fronto-central and parietal electrodes for congruent control trials compared with incongruent trials. The first negative cluster ($p = .011$, 260-310 ms) reflected a stronger late negative deflection at central and occipital electrodes for congruent control trials compared with incongruent trials.

In order to ascertain, that the effects we found were not merely due to the incongruent audiovisual stimulation, we calculated similar ERP for the subjects which did not perceive the illusion. We found no significant effects ($p <.05$) reflecting differences in ERP traces between congruent control (/pa/-/pa/) trials and McGurk trials perceived as /pa/.

Power of oscillatory activity

Aside from strictly time-locked event-related processes, the time-varying signatures of audiovisual processing of congruent and illusory percepts were of special interest. The focus of interest was on signatures that differentiate between the varying percepts, although in both conditions (congruent and McGurk) identical auditory stimuli were presented. Therefore, baseline corrected time frequency representations of congruent control trials and McGurk illusion trials were compared. In order to differentiate early and late effects the analysis interval was split into two time intervals (0-500 ms and 500-800 ms). This approach is similar to the one applied by Lange et al. (2013), who found distinct effects in low frequency responses (4-12 and 20-30 Hz) between -50 – 400 ms and 425 – 750ms between congruent and incongruent audiovisual syllable combinations. Moreover, Lange et al. (2013) reported effects of syllable congruency in high frequency responses (120-140 Hz) between 675-850 ms. To account for multiple comparisons in the present study, Bonferroni correction was applied. Since two tests were conducted (i.e. for a shorter and longer latency window) the alpha significance level was set to 0.025. The cluster-based permutation tests for comparison of illusory McGurk and congruent control trials revealed two positive clusters: one in the early time interval ranging from 50 to 250 ms after sound onset for the frequency band of 14-30 Hz and one in the late time interval ranging from 500 to 800 ms after sound onset for the frequency band of 14-30 Hz. For the early positive cluster beta band power was larger in congruent control than in McGurk illusion trials at left temporal and central sensors ($p = .01$) (Fig. 4). The late positive cluster showed increased beta band power in congruent control compared to McGurk illusion trials at fronto-central sensors ($p = .012$) (Fig. 5). Figure 6 shows the temporal development of beta power between congruent control and McGurk illusion trials in baseline and post-stimulus interval (Fig. 6). In summary, audiovisual stimulation resulted in a decrease in beta band power relative to baseline in congruent and McGurk illusion trials. However, the beta band reduction was stronger in McGurk illusion trials indicating a stronger activation of sensory processing areas (Pfurtscheller & Lopes, 1999).

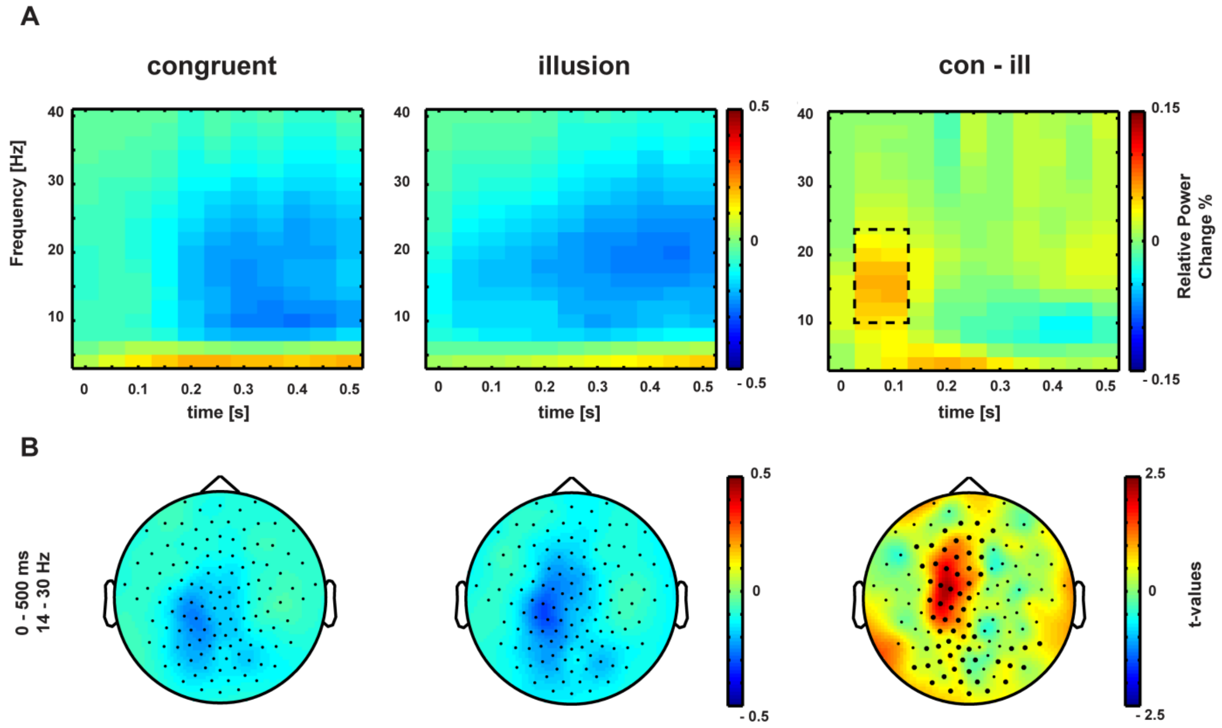

Figure 4

A) Time frequency representation of oscillatory power at the early post stimulus interval (time point 0 = onset of auditory stimulus). Beta band power is reduced following the perception of congruent (left column) audiovisual stimuli and the McGurk illusion (middle column) relative to baseline. The right column indicates the difference in relative power change between congruent and McGurk illusion trials. The dashed line marks the cluster in which this difference was statistically significant.

B) Topographies of early beta band power reduction (14-30 Hz, 0 – 500ms). Power relative to baseline following the perception of congruent (left column) and the McGurk illusion (middle column) is reduced at a left-lateralized central electrode group. The right column displays the topography of t-values at positive early beta band cluster for the comparison of congruent and McGurk illusion trials. The perception of the McGurk illusion was accompanied by a stronger beta band power reduction compared to the perception of congruent trials.

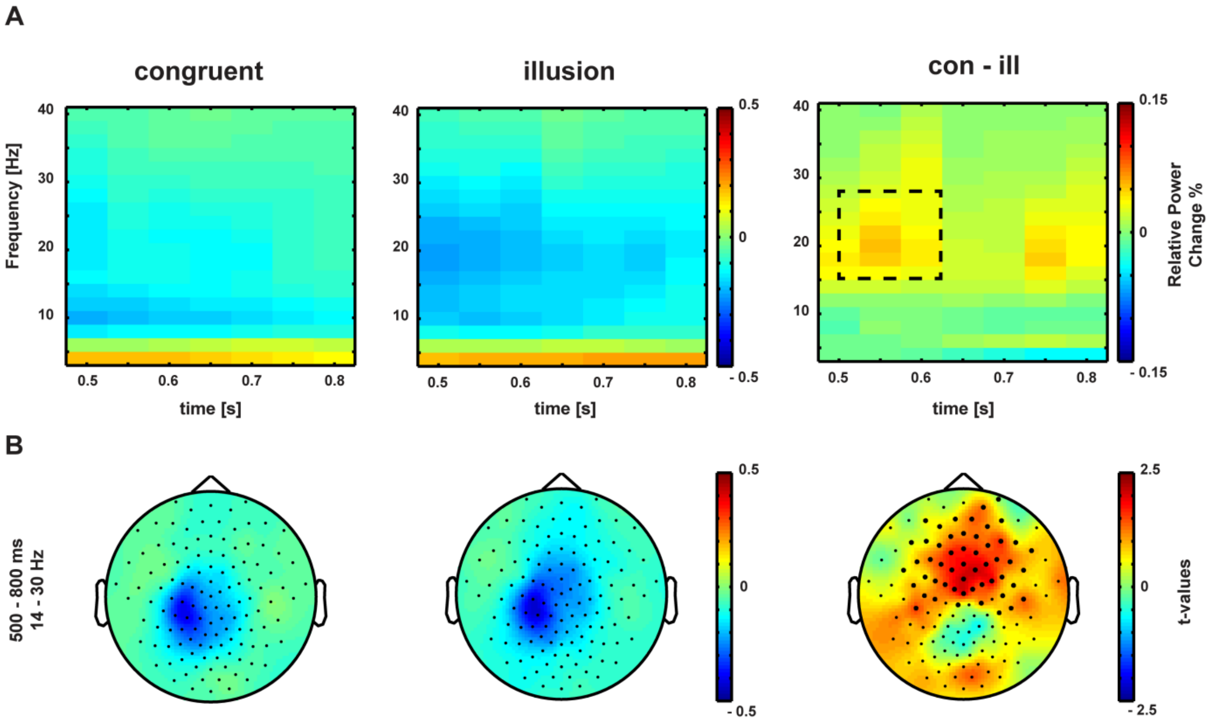

Figure 5

A) Time frequency representation of oscillatory power at the late post stimulus interval (time point 0 = onset of auditory stimulus). Beta band power is reduced following the perception of congruent (left column) audiovisual stimuli and the McGurk illusion (middle column) relative to baseline. The right column indicates the difference in relative power change between congruent and McGurk illusion trials. The dashed line marks the cluster in which this difference was statistically significant.

B) Topographies of late beta band power reduction (14-30 Hz, 500 – 800ms). Power relative to baseline following the perception of congruent (left column) and the McGurk illusion (middle column) is reduced at a left-lateralized central electrode group. The right column displays the topography of t-values at positive late beta band cluster for the comparison of congruent and McGurk illusion trials. The perception of the McGurk illusion was accompanied by a stronger beta band power reduction compared to the perception of congruent trials.

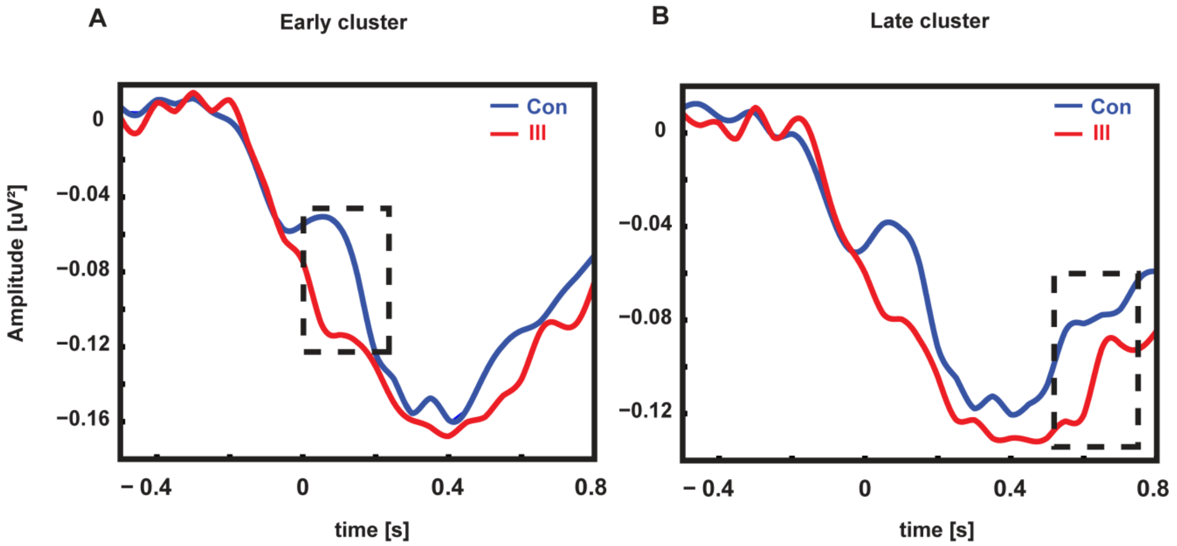

Figure 6

Time course of beta band power: A) Beta band power relative to auditory onset at early cluster for trials in which a McGurk illusion was reported (red) and congruent trials (blue). Significant differences in time course are marked by dashed line. Trials in which a McGurk illusion was perceived elicited a reduced beta band power in the early cluster compared to congruent trials in which the syllable was correctly identified.

B) Beta band power relative to auditory onset at late cluster for trials in which a McGurk illusion was reported (red) and congruent trials (blue). Significant differences in time course are marked by dashed line. Trials in which a McGurk illusion was perceived elicited a reduced beta band power in the late cluster compared to congruent trials in which the syllable was correctly identified.

Similar to ERPs we compared incongruent and McGurk illusion trials, we found two significant positive clusters in the early beta band (0-500ms). The first positive cluster (p < .001, 30-177 ms) reflected a stronger beta band power for incongruent compared with McGurk illusion trials at central and occipital electrodes. For the second positive cluster beta band power was larger in incongruent than in McGurk illusion trials at left temporal electrodes (p = .002, 213-377ms). In the late beta band we found no significant differences between incongruent and McGurk illusion trials (p >.05).

In order to account for the specific role of the late beta band effect, we calculated oscillatory power for all incongruent trials and compared to the congruent control (/pa/-/pa/) trials. We did not find significant differences between the congruent control trials and the other incongruent syllable combinations that did not produce the McGurk illusion.

Additionally, we calculated oscillatory power for the McGurk no perceivers and found no significant effects reflecting differences ($p > .05$) in ERP traces between congruent control (/pa/-/pa/) trials and McGurk trials perceived as /pa/.

In a final analysis step, we examined whether there was a relationship between early and late beta band effects across participants. To this end, a Pearson correlation coefficient was computed. However, this analysis did not reveal a significant relationship ($r (19) = -0.217$, $p = .37$), supporting the notion that the early and late beta band effects presumably reflect distinct integrative processes.

Discussion

We examined the neurophysiologic mechanisms underlying the McGurk illusion, as a phenomenon of multisensory perception. We explicitly focused on the neural signatures of the subjective percept in the McGurk illusion. To this end, we compared ERPs and oscillatory responses to congruent audiovisual syllables with an incongruent syllable combination that elicited a subjectively congruent percept (i.e. the McGurk illusion). We found that early ERPs to auditory onset were reduced during the perception of the McGurk illusion compared to congruent stimulation. Another central finding was stronger post-stimulus suppression of beta band power at short and long latencies during the perception of the McGurk illusion.

Incongruent audiovisual stimuli can reliably induce the McGurk illusion

In the McGurk trials (visual /ga/, auditory /pa/) participants reported an illusory percept in 84.2 % of all trials. This is similar to the results of the original study of McGurk and MacDonald, who reported a similarly high illusion rate (60 – 80 %) (McGurk & MacDonald, 1976). Although the McGurk illusion occurs quite frequently, there is also evidence that the occurrence of the McGurk illusion fluctuates across participants and depends on specific brain states (Keil et al., 2012), as well as stimulus properties (Martin et al., 2013). In our study we specifically aimed to investigate the McGurk illusion in trials with identical physical auditory stimuli but different visual context in order to distinguish the neural signatures of the resulting subjectively congruent percept (McGurk illusion, audio or visual percept). Moreover, Magnotti and Beauchamp (2014) recently proposed a model (noise encoding of disparity model) which provides an approach to compare multisensory integration across individuals and stimuli and accounts for inter-individual and inter-stimulus differences in the McGurk illusion (Magnotti & Beauchamp, 2014). Importantly, we optimized our experimental setup in order to get a very high McGurk illusion rate rather than a bistable percept. To this end, we presented a faint acoustic input (30dB SPL) in order to create equally salient visual and auditory stimuli. Thus, there was no dominance induced for either visual or acoustic input. Future studies should focus on the specific stimulus properties, such as sound intensity and temporal segregation of both stimuli in order to reliably induce this audiovisual illusion.

Early event-related potentials suggest an enhanced integrative processing in the McGurk illusion

The main finding in ERPs was a smaller N1 component, induced by the auditory input, in McGurk illusion compared to congruent speech trials. Previous studies, which compared ERPs to audiovisual speech stimuli with the linear combination of the constituent unisensory stimuli, have shown a reduction of the N1 for multisensory stimuli (Besle et al., 2004; Stekelenburg & Vroomen, 2007). Hence, a reduction of the N1 may be a marker of speech-specific audiovisual integration. Based on this interpretation, the smaller N1 component for McGurk illusion compared to congruent speech trials in our study could reflect a stronger audiovisual integration mechanism. Interestingly, Stekelenburg and Vroomen (2007) as well as Arnal et al. (2009) did not find differences on the N1 reduction effect between congruent and incongruent syllables. This finding somewhat contradicts our present observation.

However, they did not use speech stimuli that induced an illusory fused percept. In contrast to Stekelenburg and Vroomen (2007), the incongruent syllable combination in our study did evoke a McGurk illusion. Thus, our study shows that the auditory N1 amplitude reduction is not only found in the comparison of multisensory and combined unisensory speech stimuli but may also reflect stronger integration of audiovisual speech in the McGurk illusion.

Of particular note is the finding from a recent MEG study which directly compared congruent and incongruent audiovisual vowels (Lange et al., 2013). In this study participants were instructed to indicate whether the auditory and visual vowels matched or not. The authors found an increased N1m amplitude in response to incongruent compared to congruent vowels. Since the N1m response to incongruent audiovisual vowels did not differ from the N1m to unisensory auditory vowels, the authors suggested that the visual information in incongruent vowels might have been disregarded and rendered the incongruent stimuli similar to the unimodal auditory stimuli. Given the differences in findings obtained by Lange et al. (2013) and the present study, we propose that in McGurk illusion trials the visual stimulus impacts the processing of audiovisual speech differentially than it does in incongruent speech stimuli that do not induce an illusory percept. The smaller N1 in McGurk trials in our study indicates that in the physically incongruent audiovisual stimuli the visual context serves as an informative cue, which has a stronger impact on multisensory speech processing than it does in congruent trials.

Early suppression of beta band power reflects enhanced processing demands in the McGurk illusion

The suppression of early beta band power (0-500ms) was stronger in McGurk illusion compared to congruent trials. The effect was observed at left fronto-central to occipital scalp regions. Beta band power (13-30 Hz) has been associated with different cognitive aspects such as top-down control of attention and maintenance of cognitive sets (Engel & Fries, 2010). Specifically, Engel and Fries (2010) hypothesized that suppression of beta band power forecasts the probability of new processing demands. Notably, beta band power has been also linked to multisensory integration (Hipp et al., 2011; Senkowski et al., 2008) and several studies have demonstrated beta band power changes in language processing, such as semantic expectancy violations (Bastiaansen, Magyari, & Hagoort, 2010; Luo, Zhang, Feng, & Zhou, 2010). A decrease in beta band power has been interpreted as indexing the occurrence of unexpected stimuli and expectancy violations during speech processing (Weiss & Mueller, 2012). For instance, Bastiaansen et al. (2010) found that beta band power was more strongly suppressed at frontal regions during the presentation of word category violations compared to correct sentences. In addition, Luo et al. (2010) observed an early (0-200ms) and late (400-657ms) beta power suppression for unexpected semantic violations in response to unisensory visual stimuli. Accordingly, the suppression of early beta band power in the McGurk illusion in our study might reflect the occurrence of an unexpected stimulus that requires enhanced processing demands. We tested the specific role of the early beta band effect by comparing all incongruent trials, which did not produce a McGurk effect, with the McGurk illusion trials. The results showed that McGurk illusion trials were marked by an attenuated early, evoked responses but stronger beta band reduction compared to the other incongruent illusion trials. Based on these findings we conclude that the early beta band effects are not due to the incongruence of audio-visual stimulation but are related to the early process of fusion percept formation.

While there is no direct mapping between EEG scalp topography and the underlying neuronal sources, our observation that the beta band effect was more prominently found at left hemispheric scalp areas might indicate an involvement of speech related areas. For instance, the left-frontal topography could be indicative of a contribution of the left temporal speech areas. A previous fMRI study (Sekiyama et al., 2003) demonstrated the activation of left temporal cortex during the presentation of audiovisual syllables under low and high intelligibility conditions.

Late beta band power suppression in the McGurk illusion reflects speech-specific audiovisual integration

The late (500-800ms) suppression of beta band power at left and frontal scalp regions was stronger for McGurk illusion compared to congruent trials. This finding is in line with a previous MEG study that showed stronger late beta band suppression in left supramarginal gyrus for incongruent compared to congruent audiovisual stimuli in a match/mismatch task (Lange et al., 2013). While we compared the neural mechanisms underlying the subjectively congruent percepts of physically congruent and incongruent audiovisual stimuli, Lange et al. (2013) examined the effect of perceived match or mismatch of audiovisual stimuli. The authors reasoned that the suppression in beta band power might reflect error monitoring during incongruent stimuli. Going beyond this, we propose that the stronger suppression in beta band power in McGurk illusion trials in the present study mirrors stronger integrative multisensory processing. More specifically, the long latency effect on beta band power might reflect the formation of an illusory percept that follows the mismatch or incongruence detection at earlier processing stages. In line with the predictive coding framework (Arnal & Giraud, 2012; Rao & Ballard, 1999) , a previous study showed that invalid predictions (i.e. incongruent audiovisual speech stimuli) are accompanied by increased phase locking in beta band between 400 – 600 ms in the superior temporal sulcus (STS) (Arnal et al., 2011). The authors suggested that this beta band effect might be related to the resolution of the erroneous prediction based on the incongruent visual stimulus as well as top-down feedback processing. The STS has been considered as an essential convergence zone for audiovisual speech input and also plays a critical role in the McGurk illusion and audiovisual integration (Beauchamp, Lee, Argall, & Martin, 2004; Beauchamp, Nath, & Pasalar, 2010; Nath & Beauchamp, 2013).

Of particular interest was that the longer latency beta effect involved frontal scalp regions. This could indicate the involvement of top-down mechanisms that contributed to the formation of the McGurk illusion. Previous evidence (Keil et al., 2012) demonstrated that audiovisual integration in the McGurk illusion is reflected by increased pre-stimulus beta band power in left superior temporal gyrus (STG), as well as by an enhanced functional coupling in the beta band between frontal, temporal and parietal regions. Top-down processes at later processing stages may be also critically involved in the resolution of incongruent physical audiovisual input in order to form a coherent, illusory multisensory percept. Thereby, beta band power could play a major role. Taken together, the late beta band power effect could reflect a process of top-down induced incongruence resolution and formation of an illusory, yet subjectively congruent percept.

When interpreting our results, one should consider the possible contribution of pre-movement beta band modulations. Participants responded with different fingers to the different syllables, which could have contributed to the present findings. Previously, Pfurtscheller et al. (1999) showed that event-related beta band synchronization after finger and hand movement is somatotopically organized. However, Pfurtscheller et al. (1998) also demonstrated that pre-movement event-related desynchronization (10-12 Hz) and synchronization (16-20 Hz) are independent of movement type (index finger vs. thumb movement)(Pfurtscheller et al., 1998). Hence, we consider it unlikely that the use of different response fingers substantially contributed to the present results.

A possible three-step process of audiovisual integration

The formation of a subjectively congruent, coherent percept following low-salience audiovisual stimulation might require enhanced integrative processing. Our findings shed light on different event-related and induced oscillatory processes that take place in a temporal succession necessary for the formation of the McGurk illusion. At the first stage, the reduction of event-related N1 in McGurk trials might indexes the impact of visual context on multisensory speech processing, i.e. a stronger early audiovisual integration mechanism in the case of incongruent stimuli. The second stage, indexed by the early beta power suppression in McGurk trials, could reflect the detection of incongruent (audiovisual) stimuli and allocation of upcoming processing demands following the violation of the prediction based on the visual context. Finally at the third stage, the detected incongruence might be resolved by the integration of audiovisual stimuli and the formation of a coherent, subjectively congruent percept, namely the McGurk illusion. This last step is reflected in modulations of late beta power suppression. Importantly, we found no correlation between the effects on early and late beta band power. This indicates the two beta power suppression effects reflect different stages of audiovisual integration in the McGurk illusion.

Conclusion

In this study we investigated the underlying oscillatory processes that contribute to the formation of the McGurk illusion. Our results indicate that the processing of congruent and incongruent audiovisual stimuli, resulting in a subjectively coherent, illusory percept, is marked by altered evoked auditory responses and neuronal oscillatory activity in the beta band. Furthermore, our findings show that late oscillatory processes in the beta band contribute to the formation of the McGurk illusion. In particular, an early incongruence monitoring and integration process, as well as a late incongruence resolution and integration process, both indexed by stronger beta power suppression, seem to be crucial for the formation of the McGurk illusion.

3 Second study: Alpha-band oscillations reflect altered multisensory processing of the McGurk illusion in schizophrenia

Roa Romero, Y.; Keil, J.; Balz, J.; Niedeggen, M.; Gallinat, J. and Senkowski, D. (2016). Alpha-band oscillations reflect altered multisensory processing of the McGurk illusion in schizophrenia. *Frontiers in Human Neuroscience,*10:41*.* doi:10.3389/fnhum.20016.00041

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 12 February 2016
doi: 10.3389/fnhum.2016.00041

Alpha-Band Oscillations Reflect Altered Multisensory Processing of the McGurk Illusion in Schizophrenia

Yadira Roa Romero[1], Julian Keil[1], Johanna Balz[1], Michael Niedeggen[2], Jürgen Gallinat[3] and Daniel Senkowski[1]*

[1] Department of Psychiatry and Psychotherapy, Charité–Universitätsmedizin Berlin – St. Hedwig Hospital, Berlin, Germany, [2] Department of Education and Psychology, Free University Berlin, Berlin, Germany, [3] Department for Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

The formation of coherent multisensory percepts requires integration of stimuli across the multiple senses. Patients with schizophrenia (ScZ) often experience a loss of coherent perception and hence, they might also show dysfunctional multisensory processing. In this high-density electroencephalography study, we investigated the neural signatures of the McGurk illusion, as a phenomenon of speech-specific multisensory processing. In the McGurk illusion lip movements are paired with incongruent auditory syllables, which can induce a fused percept. In ScZ patients and healthy controls we compared neural oscillations and event-related potentials (ERPs) to congruent audiovisual speech stimuli and McGurk illusion trials, where a visual /ga/ and an auditory /pa/ was often perceived as /ka/. There were no significant group differences in illusion rates. The EEG data analysis revealed larger short latency ERPs to McGurk illusion compared with congruent trials in controls. The reversed effect pattern was found in ScZ patients, indicating an early audiovisual processing deficit. Moreover, we observed stronger suppression of medio-central alpha-band power (8–10 Hz, 550–700 ms) in response to McGurk illusion compared with control trials in the control group. Again, the reversed pattern was found in SCZ patients. Moreover, within groups, alpha-band suppression was negatively correlated with the McGurk illusion rate in ScZ patients, while the correlation tended to be positive in controls. The topography of alpha-band effects indicated an involvement of auditory and/or frontal structures. Our study suggests that short latency ERPs and long latency alpha-band oscillations reflect abnormal multisensory processing of the McGurk illusion in ScZ.

Keywords: schizophrenia, neural oscillations, multisensory integration, audiovisual, speech

Edited by:
Tetsuo Kida,
National Institute for Physiological Sciences, Japan

Reviewed by:
Pierre Mégevand,
Geneva University Hospitals, Switzerland
Anne Giersch,
Institut National de la Santé et de la Recherche Médicale , France

****Correspondence:***
Yadira Roa Romero
yadira.roa-romero@charite.de

Received: *26 October 2015*
Accepted: *25 January 2016*
Published: *12 February 2016*

Citation:
Roa Romero Y, Keil J, Balz J, Niedeggen M, Gallinat J and Senkowski D (2016) Alpha-Band Oscillations Reflect Altered Multisensory Processing of the McGurk Illusion in Schizophrenia. Front. Hum. Neurosci. 10:41. doi: 10.3389/fnhum.2016.00041

INTRODUCTION

Numerous studies using auditory (Leavitt et al., 2007; Rosburg et al., 2008; Popov et al., 2011) or visual stimuli (Butler et al., 2007; Tan et al., 2013) have shown perceptual deficits in schizophrenia (ScZ). Recently, perceptual processing in ScZ has also been investigated in multisensory setups, (Ross et al., 2007; Williams et al., 2010; Stone et al., 2011; Stekelenburg et al., 2013), but findings were less consistent than in unisensory studies. Multisensory processing requires the coordinated integration of information across widespread cortical areas, which is presumably impaired in

ScZ (Stephan et al., 2006; Uhlhaas and Singer, 2006). The coordination of information across brain areas likely involves neural synchronization, expressed in oscillatory activity (Fries, 2005; Senkowski et al., 2008). While previous research using unisensory stimuli has provided strong evidence for abnormal oscillatory activity in ScZ (Spencer et al., 2008; Uhlhaas and Singer, 2010; Grützner et al., 2013; Popov et al., 2014), currently only one patient study has examined oscillatory activity in multisensory processing (Stone et al., 2014). In a multisensory detection task, the authors observed altered gamma-band activity (i.e., 30–50 Hz) in ScZ patients.

An experimental paradigm that is well suited to examine multisensory processing is the McGurk illusion (McGurk and MacDonald, 1976). This illusion is found when lip movements pronouncing a syllable (e.g., /ga/) are paired with incongruent auditory syllables (e.g., /ba/). The pairing of specific incongruent visual and auditory syllables can induce a fused percept (e.g., a visual /ga/ and an auditory /ba/ is often perceived as /da/). Thus far, only few studies have examined the McGurk illusion in ScZ. Some studies reported a reduced McGurk illusion rate in ScZ patients compared with controls (Gelder et al., 2002; Pearl et al., 2009; White et al., 2014). However, a recent study found no group differences in illusion rates (Martin et al., 2013). Notably, previous studies using non-McGurk type audiovisual speech stimuli in ScZ also revealed inconsistent results (Surguladze et al., 2001; Szycik et al., 2009; Stekelenburg et al., 2013). For instance, Surguladze et al. (2001) used word stimuli to examine audiovisual speech perception and found no differences in susceptibility for the fusion perception between ScZ patients and healthy controls. In contrast, Stekelenburg et al. (2013) found ERP differences in the processing of congruent and incongruent audiovisual speech. Hence, further research is required to examine multisensory processing in ScZ.

In this high-density electroencephalography (EEG) study, we investigated the McGurk illusion in ScZ patients and matched control participants. Recently, we observed that neural oscillatory activity play a role in the McGurk illusion in healthy participants (Roa Romero et al., 2015). Hence, we hypothesized that oscillatory activity, reflecting the processing of the McGurk illusion is altered in ScZ patients. Here, we investigated effects across a broad frequency range of 4–40 Hz. Moreover, we examined possible interactions in event-related potentials (ERPs).

MATERIALS AND METHODS

Participants

Twenty-one patients with the DSM-IV diagnosis ScZ were recruited from outpatient units of the Charité–Universitätsmedizin Berlin. In addition, 21 age, education, and handedness matched healthy control participants, who were screened for mental disorders with the German version of the Structured Clinical Interview for DSM-IV-R Non-Patient Edition (SCID), participated in the study. Due to a lack of McGurk illusion perception (i.e., illusion rate $< 15\%$, ScZ patients = 5; matched controls = 5) and insufficient EEG data quality (ScZ patients = 2; matched controls = 2), data from seven ScZ patients and seven matched control participants were excluded. The illusion rates of excluded subjects did not significantly differ between groups (Mann–Whitney U test = 24, $p = 0.95$). All patients fulfilled the DSM-IV-TR and ICD 10 criteria for ScZ and no other axis I disorder. The psychiatric diagnosis was assessed by a senior psychiatrist at the recruiting institution. All participants had normal hearing, normal or corrected to normal vision, and no neurological disorders, alcohol or substance abuse. A random sample of 45% of all participants underwent a multi drug screening test. None of the tested participants had a positive test outcome. Severity of symptoms in ScZ patients was assessed with the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987). To test cognitive performance, the Brief Assessment of Cognition in Schizophrenia (BACS) was assessed (Keefe et al., 2004). **Table 1** provides an overview on demographic data, cognitive performance, and clinical scores. All participants gave written informed consent in accordance with the Declaration of Helsinki. The local ethics commission of the Charité–Universitätsmedizin Berlin approved the study.

Experimental Design

The setup was identical to our study in healthy participants (Roa Romero et al., 2015). During the experiment different types of congruent and incongruent audiovisual syllable trials were presented (**Table 2** and Supplementary Table S2). Congruent syllable trials contained matching audiovisual syllables (e.g., visual /pa/ and auditory /pa/), whereas incongruent syllable trials contained non-matching audiovisual syllables (e.g., visual /pa/ and auditory /ka/). The congruent syllable combination visual /pa/ and auditory /pa/ served as control condition in the EEG data analysis. To induce the McGurk illusion, we presented the combination of a visual /ga/ and an auditory /pa/, which frequently led to the illusory perception /ka/ or "something else." When the resulting perception of McGurk trials was /ka/ or "something else," we will refer to these trials as 'McGurk illusion trials.' Importantly, the auditory syllable (i.e., /pa/) in congruent control trials and in McGurk trials was identical. In total 300 McGurk trials were presented. In addition, 150 incongruent syllable trials were presented (**Table 2**). These other incongruent syllables served as distractor stimuli to ensure that the McGurk illusion was specific to McGurk trials and not merely the result of the audiovisual mismatch. In each trial, the first frame of the video clip was presented for a random interval ranging from 1000 to 1500 ms (mean = 1250 ms). After the video clip, which had on average a duration of 990 ms (Supplementary Table S2), the last frame of the clip was presented on average for 710 ms. The total video sequence was presented for 1700 ms. Following the video clip, the last frame of each clip, in which the mouth of the actress was closed, was presented for 1000 ms. During this time the fixation cross turned into a question mark for 500 ms at a random time point and participants were required to indicate by a button press with the index, middle, ring or small finger of their right hand whether they had perceived the syllable /pa/, /ga/, /ka/, or "something else," respectively. Each trial had a duration of 3700–4200 ms (**Figure 1**).

TABLE 1 | Overview of demographic data.

	Patients		Controls		Statistics	
	Mean	***SD***	**Mean**	***SD***	***t*-values**	***p*-values**
Age (years)	35.57	6.55	36.79	7.18	−0.468	0.664
Education (years)	10.64	1.21	10.42	1.28	0.453	0.654
Illness duration (years)	9.50	5.63	–	–	–	–
Chlorpromazine Eq. (daily dosage/mg)	375.79	155.10	–	–	–	–
	N		***N***			
Gender (m/f)	10/4		10/4		–	–
Handedness (r/l)	12/2		12/2		–	–
Antipsychotic Med.	14		–		–	–
Co-medication*	5		–		–	–
BACS						
Verbal memory	41.07	13.38	45.29	9.54	−0.959	0.346
Digit	19.21	4.30	20.00	4.67	−0.463	0.674
Motor	68.71	11.40	74.57	10.77	−1.397	0.147
Fluency	47.28	16.34	52.14	15.55	−0.806	0.428
Symbol coding	55.00	13.60	57.21	16.04	−0.394	0.697
ToL	17.43	2.50	17.29	2.33	−0.156	0.877
Total score	248.71	43.76	266.50	39.70	−1.126	0.270
PANSS						
Negative	19.07	3.56	–	–	–	–
Positive	17.14	2.77	–	–	–	–
General	38.29	3.69	–	–	–	–
Total score	74.50	7.43	–	–	–	–

**Co-medication of antipsychotics and mood stabilizers.*

TABLE 2 | Overview of presented congruent and McGurk-illusion trials.

Condition	Visual	Audio	*N*
Congruent*	Pa	Pa	100
Congruent	Ga	Ga	100
Congruent	Ka	Ka	100
McGurk	Ga	Pa	300
Incongruent	Pa	Ga	30
Incongruent	Pa	Ka	30
Incongruent	Ga	Ka	30
Incongruent	Ka	Pa	30
Incongruent	Ka	Ga	30

*The middle panels show the visual and the auditory syllables. The remaining five incongruent syllables combinations comprised 150 trials. In total 750 trials were presented. *Congruent Pa–Pa trials served as control condition.*

EEG Recording and Data Analysis

Electroencephalography data were recorded using a 128 channel active EEG system (EasyCap, Herrsching, Germany), which included two EOG electrodes (online: 1000 Hz sampling rate with a 0.016–250 Hz bandpass filter; offline: downsampling to 500 Hz, 1–125 Hz FIR bandpass filtering and 49.1–50.2 Hz, fourth order Butterworth notch filtering). To correct for EOG and ECG artifacts, independent component (IC) analyses were conducted (extended runica; Lee et al., 1999). On average 14.64 ± 0.82 (standard error of mean) ICs for ScZ patients and 16.71 ± 0.93 ICs for matched controls were rejected. Remaining noisy channels were interpolated using spherical interpolation (ScZ patients = 13.43 ± 0.88 channels; matched controls = 15.64 ± 1.03 channels). Epoched data were re-referenced to common average. For ERP analysis, data were filtered (2 Hz, second order and 35 Hz, 12th order two-pass Butterworth filter) and baseline corrected (−500 ms to −100 ms prior to sound onset). For the time-frequency analysis of lower frequency responses (i.e., 4–40 Hz) wavelet transformation with frequency depending Hanning window was computed in 2 Hz steps (time window $\Delta t = 5/f$, spectral smoothing: $f = 1/\Delta t$). For the analysis of higher frequency responses (i.e., 40–100 Hz) Slepian tapers (fixed time window $t = 0.2$ s, fixed spectral smoothing: $f = 10$ Hz) were applied. However, since we did not find robust modulations of high frequency (i.e., >40 Hz) responses in the current data, we focused the analysis to the frequency range from 4 to 40 Hz. Averaged oscillatory activity was baseline corrected (relative change, from −500 to −100 ms prior to sound onset).

The analysis of behavioral data focused on McGurk trials and congruent trials (**Table 3**). Reaction tendencies in McGurk trials were calculated as the relative proportion of illusion, audio, and visual percept responses (Keil et al., 2012). Independent t-tests between groups were conducted separately for rate of illusion, audio and visual percept (Bonferroni corrected α-level = 0.017).

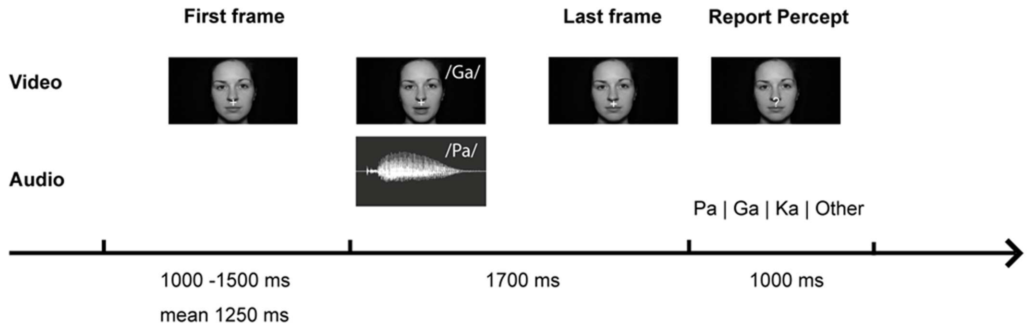

FIGURE 1 | McGurk trial with video frames of the syllable /ga/ and audio trace of the syllable /pa/. Participants were presented with congruent (auditory /pa/ and visual /pa/) or incongruent (auditory /pa/ and visual /ga/) audiovisual syllables. They were instructed to reported which syllable they perceived.

The analysis of EEG data focused on the comparison of ERPs and oscillatory responses to McGurk illusion and congruent trials. The number of trials was equalized according to the lowest number of trials in either condition. On average, for each condition 62 ± 3.71 trials for ScZ patients and 69 ± 3.47 trials for matched controls were used. In order to examine whether any possible effects are driven by the incongruent audio–visual stimulation in McGurk trials (i.e., visual /ga/ and auditory /pa/) and not due to the multisensory fusion process that leads to the McGurk illusion, the same analyses for ERPs and oscillatory power were calculated for 19 ScZ patients and 19 matched controls, irrespective of McGurk illusion perception (see Supplementary Material and Supplementary Table S1). Hence, all McGurk trials, irrespective of perception were included in this analysis. Note that in this analysis two participants were excluded from each group due to insufficient quality of EEG data. Similar to our previous study (Roa Romero et al., 2015), we examined ERP amplitudes and oscillatory power at a medio-central region of interest (ROI), comprising 16 channels. The activity of the channels was averaged and served as dependent variable in the statistical analyses. In addition, we calculated the Global field power (GFP) for each Condition and Group as a measure of location-independent cortical activity integrating all channels (Esser et al., 2006). Due to a more complex factorial design we applied a different statistical analysis approach compared to our previous study, in which we computed non-parametric cluster statistics (Roa Romero et al., 2015). Specifically, for ERPs, GFP, as well as oscillatory power running 2 × 2 ANOVAs with the factors Group (ScZ patients vs. matched controls) and Condition (congruent vs. illusion) were conducted for each sample point (Schurger et al., 2008; Kissler and Koessler, 2011; Kissler and Herbert, 2013). In accordance with our previous study (Roa Romero et al., 2015), ERPs were analyzed in a time window from 0 to 500 ms and oscillatory responses from 0 to 850 ms following auditory syllable onset. The above-described 2 × 2-factorial ANOVA was conducted for each sample point in these intervals. To account for multiple testing, a time stability criterion of at least 10 consecutive significant sample points (i.e., 20 ms) was applied (Guthrie and Buchwald, 1991; Picton et al., 2000). For oscillatory responses the running 2 × 2-factorial ANOVA was computed for each sample point and frequency (4–40 Hz) in the 0 to 850 ms interval. Due to the low temporal resolution of the time-frequency transformation, a time stability criterion of at least 100 ms was applied. Significant main effects or interactions were followed-up by *t*-tests. Finally, Pearson correlations were computed between psychopathology scores (PANSS), McGurk illusion rate, and EEG data. To statistically control for the influence of antipsychotic medication, medication dosage was converted to chlorpromazine equivalent level (Gardner et al., 2010) and entered as covariate to partial correlation analyses in the patient group.

RESULTS

Behavior

The recognition rate of congruent trials was at ceiling level (ScZ patients = 98.49%; matched controls = 97.92%). In McGurk trials ScZ patients and matched controls reported an illusory percept in 78.30 and 65.07% of trials, respectively, which was not significantly different [$t(26) = 1.395$, $p = 0.175$]. Moreover, the comparisons of the different percepts that could be evoked by McGurk trials did not reveal significant group differences (**Table 3**). We also examined whether there were behavioral differences between the samples of 19 participants per group for whom McGurk trials, irrespective of perception were analyzed. This comparison did also not reveal significant differences between ScZ patients (66.33%) and matched controls [53.14%; $t(36) = 1.191$, $p = 0.242$].

TABLE 3 | Means, standard deviation, and mean difference for behavioral performance (perceptual ratings in McGurk trials) in patients and controls.

	Patients		Controls		Statistics	
	Mean	***SD***	**Mean**	***SD***	***t*-values**	***p*-values**
Illusion percept %	78.30	20.36	65.07	29.09	1.395	0.175
Audio percept %	21.16	19.93	33.02	29.10	−1.259	0.219
Visual percept %	0.54	0.74	1.91	1.48	−1.516	0.141

Each group comprised of 14 participants.

Event-Related Potentials and Global Field Power

Stimulus-evoked activity between McGurk illusion trials (i.e., McGurk trials in which participants reported an illusion) and congruent control trials was compared between 14 ScZ patients and 14 matched controls. The running 2 × 2 ANOVA revealed a significant main effect of Group between 190 and 250 ms [$F(2,26) = 16.89$, $p = 0.0035$], due to larger amplitudes in ScZ patients (0.275 μV) compared with matched controls (–0.160 μV). Furthermore, a main effect of Condition was found between 175 and 195 ms [$F(2,26) = 8.06$, $p = 0.009$], indicating larger negative amplitudes in illusion trials (–0.550 μV) compared with congruent trials (–0.370 μV). Notably, a significant Group by Condition interaction was observed between 60 and 80 ms [$F(2,26) = 6.79$, $p = 0.015$], indicating that the amplitude differences between illusion and congruent trials were significantly larger for matched controls than for ScZ patients (**Figures 2** and **3**). Follow-up *t*-tests for the 60 to 80 ms interval, which were conducted separately for each condition, revealed significant amplitude differences between groups in McGurk illusion trials [$t(27) = 2.27$, $p = 0.04$] but not in congruent trials [$t(27) = 0.33$, $p = 0.74$]. Furthermore, to examine the general effect of incongruence we compared ERPs for all McGurk trials, irrespective of subjective percept with the ERPs to congruent control trials. The running 2 × 2 ANOVA, which included 19 participants in each group, revealed no significant interactions or main effects of Group. However, main effects of Condition were found between 246 and 266 ms [$F(2,36) = 8.12$, $p = 0.007$] and between 338 and 368 ms [$F(2,36) = 5.57$, $p = 0.023$]. The first condition effect indicates larger positive amplitudes in congruent trials compared with McGurk trials. The latter condition effect showed the reversed pattern, indicated larger positive amplitudes in McGurk trials compared with congruent trials (Supplementary Figure S1).

The analysis of GFP for 14 ScZ patient and 14 matched controls revealed a significant main effect of Condition between 140 and 160 ms [$F(2,26) = 10.16$, $p = 0.0037$, Supplementary Figure S4], due to larger amplitudes in congruent compared with illusion trials. No other main effects or interactions were observed. In a final analysis step, we explored whether

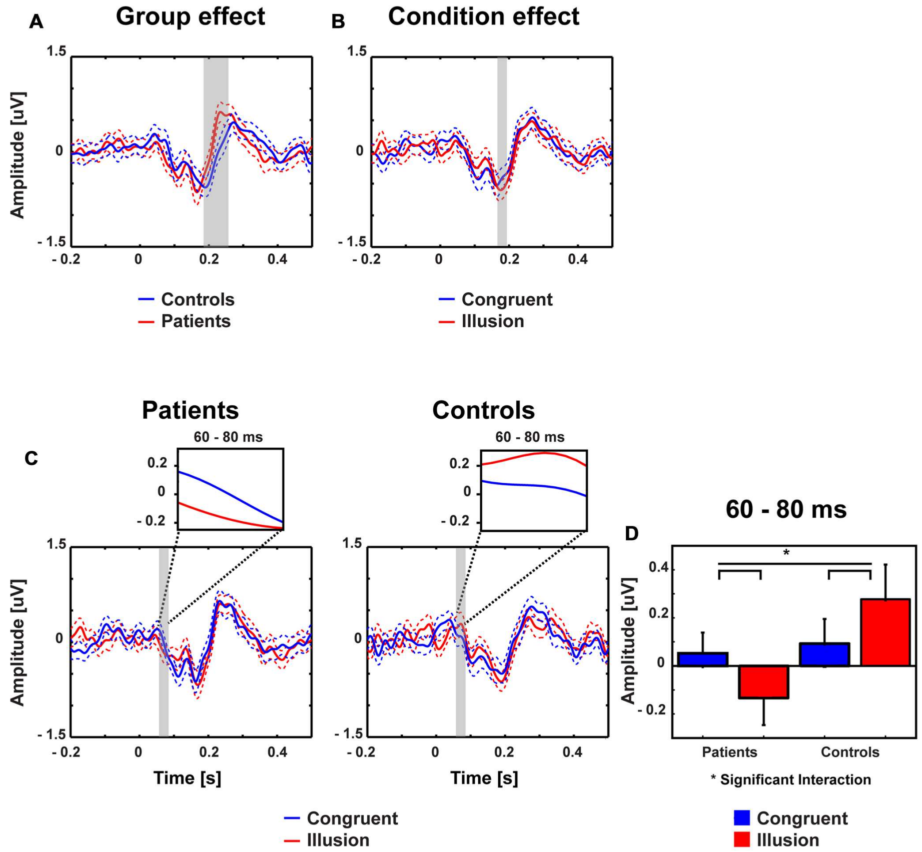

FIGURE 2 | Traces and amplitudes of early medio-central event-related potentials. Traces of ERPs in patients (left) and controls (right) for illusion (red line) and congruent (blue line) trials. Time zero denotes the onset of the auditory syllable. Dashed lines represent standard error of mean. The significant time intervals are highlighted in gray. **(A)** Group effects were found between 190 and 250 ms. **(B)** Condition effects were found between 175 and 195 ms. **(C)** Interactions between group and condition were found after 60–80 ms. **(D)** Mean ERP amplitudes of the 60–80 ms time interval with error bars (standard error of mean). In patients amplitudes were more positive in congruent compared with illusory trials. By contrast, in the control group larger positive amplitudes were observed in illusory compared with control trials.

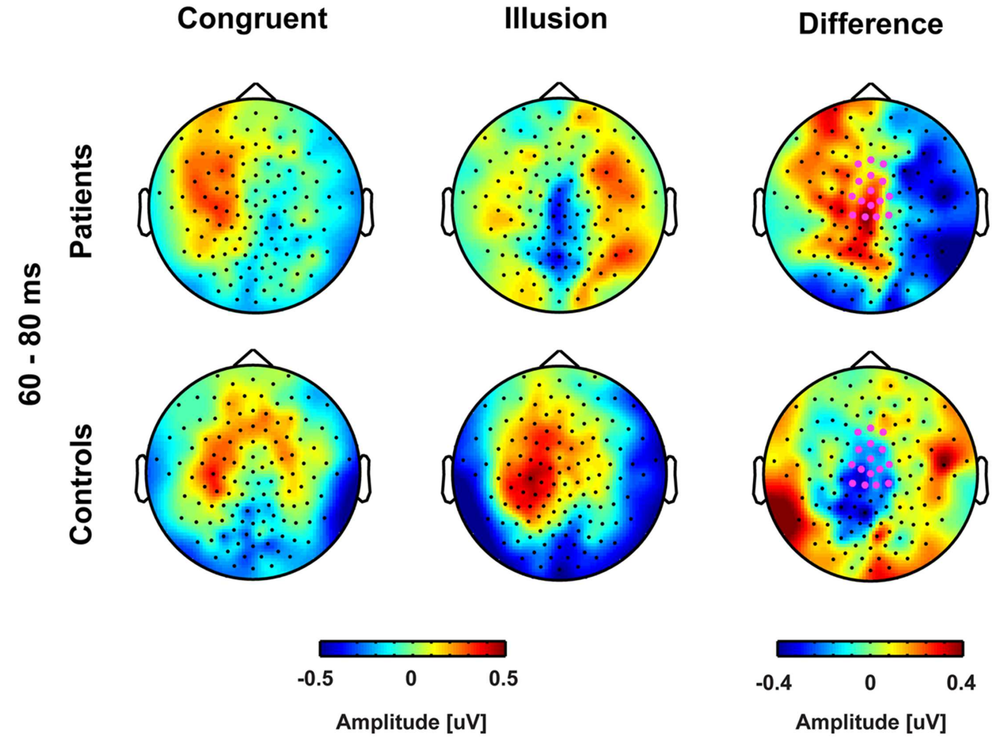

FIGURE 3 | Topographies of early event-related potentials for congruent and McGurk illusion trials. The bold magenta dots in the right panel denote the electrode ROI for which significant group by condition effects in the time interval 60–80 ms were found.

there are topographic differences in evoked activity between conditions and groups. To this end, within each subject, we calculated the Global Dissimilarity Index (GDI, Murray et al., 2008) between the congruent and illusion trials as a measure of difference in the topographies between both conditions. Subsequently, we compared the individual GDI values between groups with an independent *t*-test. This analysis did not reveal any significant effects in GDI, indicating that the topographies did not substantially differ between groups and conditions.

Power of Oscillatory Activity

Aside from strictly time-locked event-related processes, the time-varying signatures of audiovisual processing of congruent and illusory percepts were of interest. The focus of this analysis was on oscillatory activity that differentiates between the varying percepts, although in both trial types (i.e., control and McGurk trials) identical auditory stimuli were presented. Therefore, oscillatory activity in response to control trials and McGurk illusion trials was compared. The running 2 × 2-factorial ANOVA, which was computed for each sample point in the 0–850 ms interval for the frequency range of 4–40 Hz, did not reveal significant main effects. However, a significant Group by Condition interaction was found in the alpha-band (i.e., 8–10 Hz) between 550 and 700 ms [$F(2,26) = 5.47$, $p = 0.027$; **Figure 4**]. In ScZ patients, alpha-band power was stronger suppressed in congruent ($-0.24\ \mu V^2$) compared with illusion trials [$-0.17\ \mu V^2$; $t(13) = -1.80$, $p = 0.095$; **Figure 5**]. The reversed pattern was observed in matched controls [illusion trials $= -0.28\ \mu V^2$, congruent trials $= -0.22\ \mu V^2$; $t(13) = 1.64$, $p = 0.124$]. Visual inspection of alpha-band power time course indicated that ScZ patients and matched controls primarily differed in the illusion condition. Following visual inspection of the alpha-band topography, we additionally explored possible alpha-band effects at right posterior electrodes ($n = 7$). For these electrodes we found a main effect of condition [$F(2,26) = 7.20$, $p = 0.01$] between 550 and 700 ms, which revealed less alpha-band power in congruent compared with illusion trials (**Figure 6**). Additionally, we investigated oscillatory power for McGurk trials, irrespective of subjective percept and congruent control trials. This analysis, in which 19 ScZ patients and 19 matched controls were entered, did not reveal significant interactions or main effects of Group. However, the ANOVA revealed a main effect of Condition in the theta-band (4 Hz) between 50 and 350 ms [$F(2,36) = 8.04$, $p = 0.001$]. In both groups, theta-band power was larger in congruent compared with McGurk trials, suggesting that incongruent visual information modulates early audiovisual processing (Supplementary Material and Supplementary Figures S2 and S3).

Relationships Between EEG Data, Illusion Rates, and Clinical Symptoms

The correlations between ERP amplitudes and McGurk illusion rates for ScZ patients and matched controls were not significant. Moreover, the Bonferroni-corrected correlations

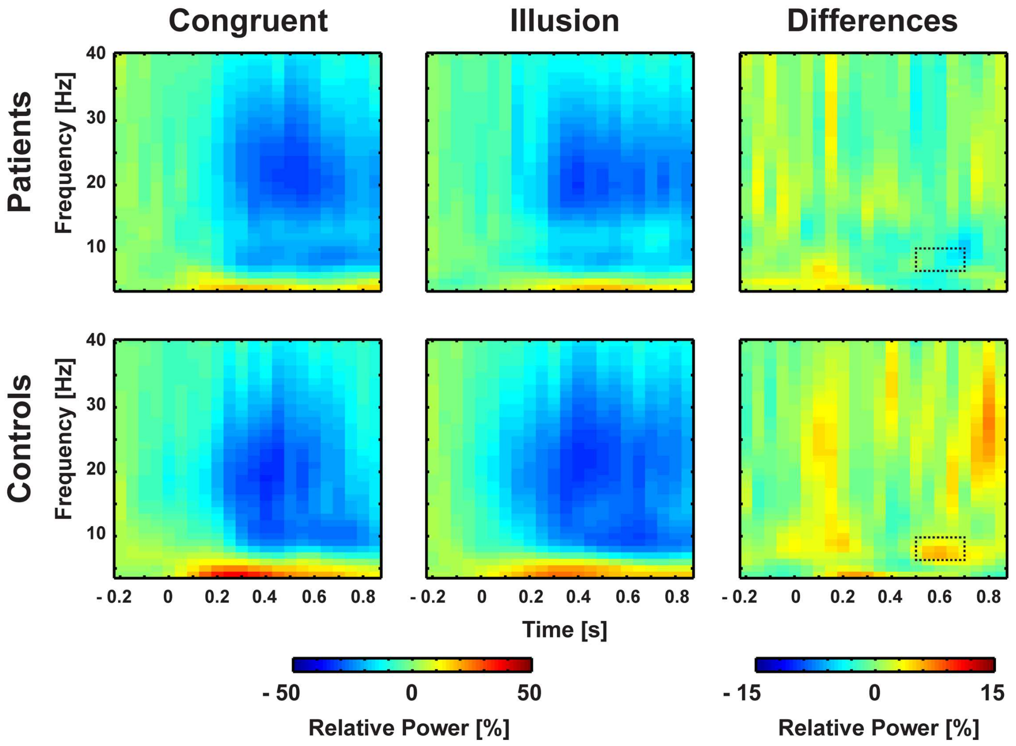

FIGURE 4 | Time-frequency responses of oscillatory responses at medio-central electrodes. Time zero denotes the onset of the auditory syllable. Significant group by condition interactions were found in the alpha-band (8–10 Hz) after 550–700 ms. Significant cluster are marked by dashed line.

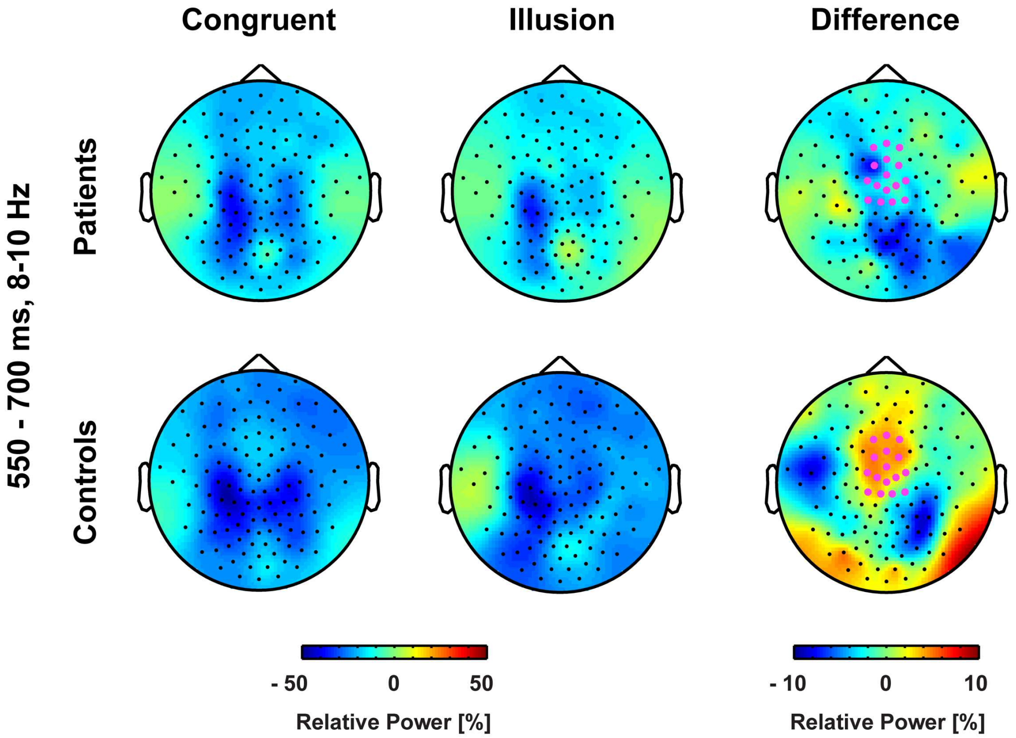

FIGURE 5 | Topographies of medio-central alpha-band power. The bold magenta dots in the right panel denote the electrode ROI for which significant group by condition effects in the time interval 550–700 ms were found.

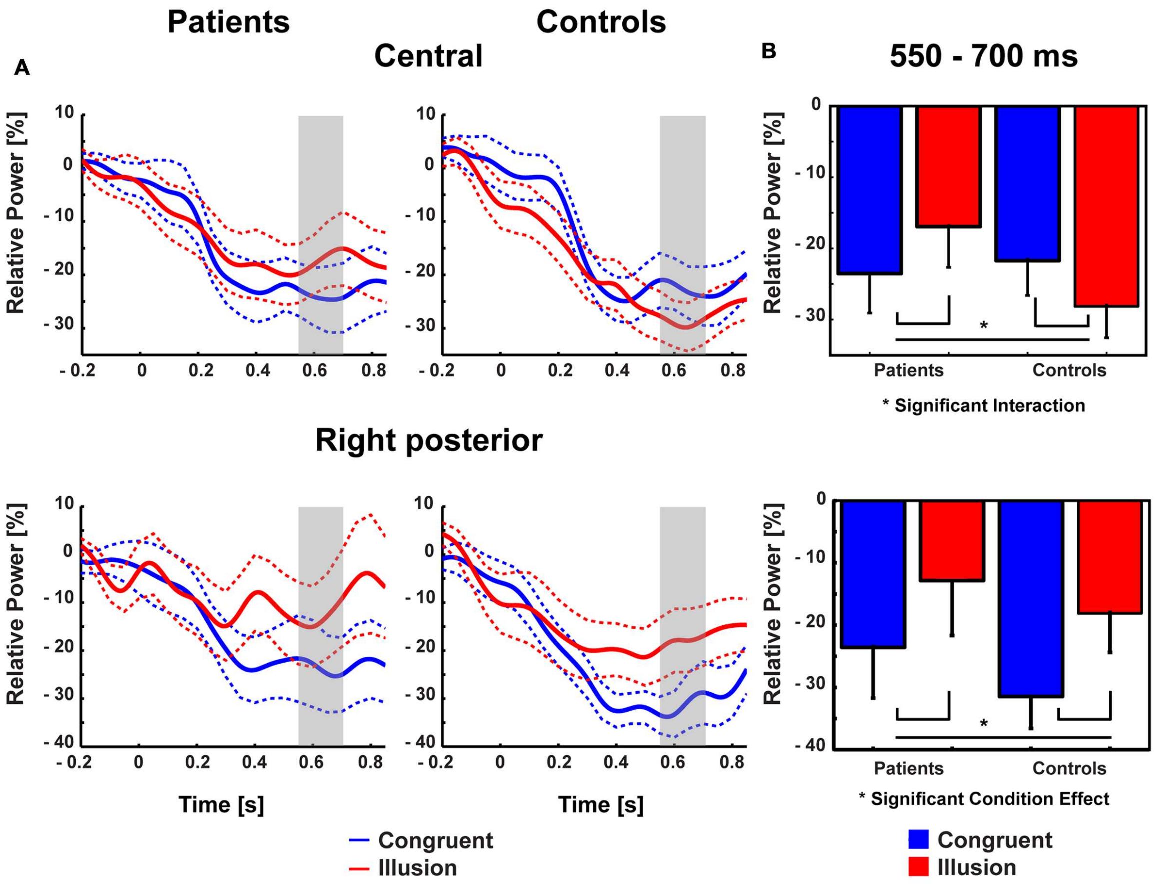

FIGURE 6 | Traces and amplitudes of medio-central alpha-band power. (A) Traces of alpha-band power in patients (left) and controls (right) for illusion (red line) and congruent (blue line) trials at central (upper row) and right posterior (lower row) electrodes. Dashed lines represent standard error of mean. Interactions between group and condition were found after 550–700 ms indicated by gray trace. At right posterior electrodes a similar time course of alpha-band power was found. Time zero denotes the onset of the auditory syllable. Note that negative amplitude values represent stronger suppression. **(B)** Mean alpha-band power at medio-central (upper row) and right posterior (lower row) scalp regions in the 550–700 ms time interval with error bars (standard error of mean). In patients stronger alpha-band suppression was found in congruent compared with illusory trials. In controls stronger alpha-band suppression was observed in illusory compared with congruent trials. For right posterior alpha-band power both groups showed stronger suppression during congruent compared with illusion trials.

between ERP amplitudes and alpha-band power were not significant (ScZ patients: $r = -0.131$, $p = 0.671$; matched controls: $r = 0.441$, $p = 0.202$). Additionally, none of the correlations between ERP amplitudes, alpha power and PANSS subscale scores were significant (all p-values >0.05). However, in ScZ patients a significant negative correlation between alpha-band power and illusion rate was found (partial correlation $r = -0.756$, $p = 0.004$). Interestingly, in matched controls there was a positive, yet not significant relationship between alpha-band power and illusion rate ($r = 0.30$, $p = 0.29$; **Figure 7**). The Pearson correlation coefficients differed significantly between groups ($Z = 3.04$, $Zkrit = 1.65$). Exploratory analysis of right posterior alpha-band power and illusion rate revealed no significant correlations for ScZ patients ($r = -0.06$, $p = 0.85$) and matched controls ($r = -0.1$, $p = 0.74$). Furthermore, the correlations between PANSS subscale scores and McGurk illusion rates were not significant.

DISCUSSION

In this electroencephalography study, we examined the McGurk illusion in ScZ. We observed altered ERPs and alpha-band suppression effects in ScZ patients compared with matched controls in McGurk illusion compared with congruent audiovisual syllable trials. Our behavioral analysis did not reveal group differences in McGurk illusion rates. Some studies have reported reduced illusion rates in ScZ (Gelder et al., 2002; White et al., 2014). However, in line with the present observation, other studies did not find group differences (Myslobodsky et al., 1992; Martin et al., 2013). The inconsistencies in findings could be due to differences in criteria for the definition of illusion rates, group ages, and heterogeneity of the clinical samples.

Our analysis of ERPs analysis revealed an early (60–80 ms) interaction at central electrodes. In matched controls, a larger positive deflection was found in illusion compared with congruent trials. By contrast, in ScZ patients, a larger negative

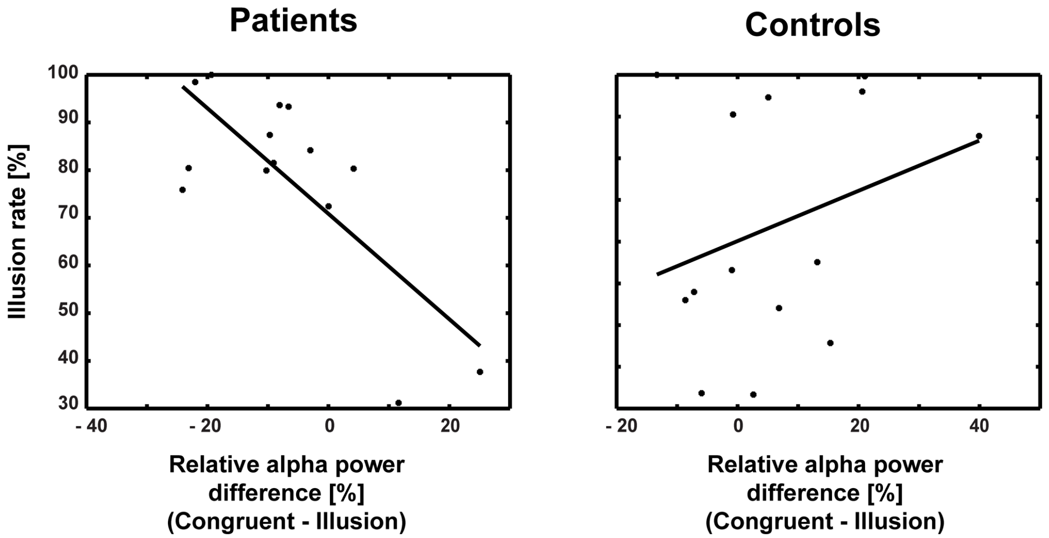

FIGURE 7 | Correlations between alpha-band power and illusion rate for patients and controls. The correlation coefficients differed significantly between patients ($r = -0.756$) and controls ($r = 0.30$).

deflection was found in illusion compared with congruent trials. The group differences in ERP amplitudes were primarily found in McGurk illusion trials, indicating altered early audiovisual processing of these trials in ScZ (**Figure 2**). A previous study in healthy participants has shown early processing differences between congruent and incongruent audiovisual syllables (Lebib et al., 2003). The authors found larger positive deflections for incongruent compared with congruent syllables. They suggested that the amplitude enhancement in incongruent syllables reflects an early detection of non-matching audiovisual information. Hence, the absence of early amplitude enhancement in McGurk illusion trials in ScZ patients might be due to a deficit in the early detection of non-matching audiovisual syllables. A study by Magnée et al. (2009), using an audiovisual P50 repetition-suppression paradigm, also revealed altered early ERPs in ScZ. Thus, our finding indicates altered early processing of sensory information across modalities in ScZ. Notably, when we compared congruent trials with all McGurk trials, independent of the subjective percept, there were no differences in early ERPs, neither in ScZ patients, nor in matched controls. Hence, the observation of early interaction effects in McGurk illusion trials indicates that there is specific processing deficit of these trials in ScZ patients. This assumption requires further empirical testing.

Another interesting finding in both groups were larger negative deflections (175–195 ms) in McGurk illusion compared with congruent trials. Moreover, the GFP also revealed differences between conditions at this latency. Previous studies in healthy participants found more negative auditory evoked P2 amplitudes for incongruent compared with congruent audiovisual syllables (Stekelenburg and Vroomen, 2007; Knowland et al., 2014). Similarly, another study using a McGurk oddball paradigm found a McGurk stimulus induced mismatch negativity at a similar latency (Saint-Amour et al., 2007). Hence, auditory evoked components could be a marker for congruency-detection and competition between sensory inputs during the processing of incongruent stimuli. A further finding was a larger positive deflection (190–250 ms) in ScZ patients compared with matched controls. Similar effects have been reported by Stekelenburg et al. (2013). The authors suggested that these larger ERPs could reflect multisensory processing deficits in the patient group. In summary, early ERPs in McGurk illusion trials in ScZ patients might be caused by a deficit in early incongruence detection in audiovisual syllables. In contrast, the results from late ERPs suggest no deficits in incongruence-detection but impaired mismatch-resolution during later processing stages of the McGurk illusion in ScZ.

Contrary to ERPs, the analysis of oscillatory responses did not reveal any early effects. The key finding in oscillatory responses was an interaction in long-latency (550–700 ms) alpha-band power: In ScZ patients medio-central alpha-band suppression was stronger in congruent compared with McGurk illusion trials. The pattern of suppression effects was reversed in the control group. Notably, the time course of medio-central alpha-band suppression in congruent trials was similar in both groups. By contrast, in illusion trials the suppression of later medio-central alpha-band power was more pronounced in matched controls compared with ScZ patients (**Figure 6**). This could reflect a lower signal-to-noise ratio in the auditory system during the processing of McGurk illusion in ScZ patients. Klimesch et al. (2007) and Klimesch (2012), hypothesized that suppression of alpha-band power is a neural signature of active processing in task relevant networks. Hence, the stronger alpha-power suppression presumably indicates a better signal-to-noise ratio, because irrelevant information (noise) is inhibited. In matched controls we found stronger medio-central alpha-band suppression in McGurk illusion compared with congruent trials. In contrast, ScZ patients showed less medio-central alpha-band suppression in illusion trials, which could indicate impaired integrative processing. Further, the reduced medio-central alpha-band power suppression in ScZ patients during illusion trials indicates a state, in which irrelevant information is not appropriately inhibited and the signal to noise ratio in the auditory system during the processing of McGurk illusion might be lower. Alpha-band suppression effects have been recently found in auditory illusion paradigms (Müller et al., 2013; Leske et al., 2014). Müller et al. (2013) observed a positive relationship between alpha-band suppression and illusory perception of music in an auditory continuity paradigm. Hence, the altered alpha-band suppression in ScZ patients might reflect abnormal processing of the auditory aspect of the McGurk illusion.

The medio-central topography of the alpha-band effect indicates an involvement of auditory and/or frontal structures. Auditory oddball tasks (Koh et al., 2011) and auditory gating paradigms (Popov et al., 2011) revealed altered alpha-band suppression in ScZ. In addition, alpha-band power modulations in the visual cortex have been found to contribute to multisensory illusions, such as the sound-induced flash illusion (Lange et al., 2014). Low alpha-band power indicates the increased excitability of visual areas and determines stimulus perception by regulating the incoming flow of information, within and between sensory areas, such as visual and auditory cortex.

Interestingly, in our study alpha-band power suppression over right posterior areas was stronger in congruent compared with McGurk illusion trials. This effect in right posterior alpha-band power was similar in both groups, indicating intact processing of McGurk illusion trials in visual areas in ScZ patients (**Figure 6**). Thus, the processing of McGurk illusion trials in ScZ seems to be specifically altered in auditory and/or frontal areas. The less pronounced medio-central alpha-band suppression in illusion trials could mirror reduced auditory processing, possibly due to an increased ambiguity in the encoding of auditory information. In contrast, in matched controls there might be a stronger processing of both auditory and visual stimuli in the illusion trials. As shown in this study, the modulations in alpha-band power presumably mirror a process that differentiates between patients and matched controls during the formation of the McGurk illusion.

In a previous study, we investigated the McGurk illusion in healthy subjects and revealed modulations in late beta-band activity over left temporal and frontal areas (Roa Romero et al., 2015). We suggested that the suppression of late beta-band power fosters the formation of a coherent, subjectively congruent percept, namely the McGurk illusion. The absence of differences in the late beta-band power in the present study could indicate that the process of perception formation itself is not altered in ScZ patients. This could lead to identical behavior reflected by similar illusion rates between both groups.

Another finding in our study was that the medio-central alpha-band suppression in ScZ patients was negatively correlated with the McGurk illusion rate, while it tended to be positively associated with the illusion rate in matched controls. This supports the notion that less pronounced alpha-band suppression in McGurk illusion in ScZ patients reflects altered multisensory integration. Moreover, we found no correlation between the effects of early ERP and late alpha-band power. This indicates that the effects might reflect distinct aspects of audiovisual processing in the McGurk illusion.

In this study, we obtained oscillatory responses in the EEG, but did not find clear modulations in gamma-band oscillations. MEG compared to the EEG has a higher sensitivity in the measurement of high frequency oscillations (Muthukumaraswamy and Singh, 2013). Additionally, numerous studies have revealed reduced gamma-band power in ScZ patients (Gallinat et al., 2004; Leicht et al., 2010). These factors might have contributed to the absence of gamma-band modulations in the present study. It would be interesting to use MEG to examine gamma-band oscillations during the processing of the McGurk illusion in ScZ patients, to uncover the possible role of gamma-band power for illusory perception.

CONCLUSION

Taken together, our study revealed altered early and late processing of McGurk illusion trials in ScZ. The early ERP effect might reflect audiovisual processing deficits in ScZ patients. The altered late alpha-band suppression effects could reflect abnormal multisensory integration in auditory and/or frontal areas. Our study provides new insight into the processing of the McGurk illusion in ScZ and fosters the notion that alpha-band oscillations reflect altered multisensory integration in ScZ patients.

AUTHOR CONTRIBUTIONS

YRR, JK, and DS designed the experiment. YRR and JB recruited the patients and healthy controls and collected the data. YRR performed data analysis and prepared MS. JK assisted data analysis and manuscript preparation. DS, JG, and MN reviewed and edited the MS.

FUNDING

This work was supported by the European Union (ERC-2010-StG-20091209 to DS) and the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG-Grant: KE1828/2-1 to JK and SE1859/3-1 to DS).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2016.00041

REFERENCES

Butler, P. D., Martinez, A., Foxe, J. J., Kim, D., Zemon, V., Silipo, G., et al. (2007). Subcortical visual dysfunction in schizophrenia drives secondary cortical impairments. *Brain* 130, 417–430. doi: 10.1093/brain/awl233

Esser, S. K., Huber, R., Massimini, M., Peterson, M. J., Ferrarelli, F., and Tononi, G. (2006). A direct demonstration of cortical LTP in humans: a combined TMS / EEG study. *Brain Res. Bull.* 69, 86–94. doi: 10.1016/j.brainresbull.2005.11.003

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011

Gallinat, J., Winterer, G., Herrmann, C. S., and Senkowski, D. (2004). Reduced oscillatory gamma-band responses in unmedicated schizophrenic patients indicate impaired frontal network processing. *Clin. Neurophysiol.* 115, 1863–1874. doi: 10.1016/j.clinph.2004.03.013

Gardner, D. M., Pharm, D., Murphy, A. L., Donnell, H. O., Pharm, B. S., Centorrino, F., et al. (2010). International consensus study of antipsychotic dosing. *Am. J. Psychiatry* 167, 686–693. doi: 10.1176/appi.ajp.2009.09060802

Gelder, B., Vroomen, J., Annen, L., Masthof, E., and Hodiamont, P. (2002). Audio-visual integration in schizophrenia. *Schizophr. Res.* 59, 211–218. doi: 10.1016/S0920-9964(01)00344-9

Grützner, C., Wibral, M., Limin, S., Rivolta, D., Singer, W., Maurer, K., et al. (2013). Deficits in high-(>60 Hz) gamma-band oscillations during visual processing in schizophrenia. *Front. Hum. Neurosci.* 7:88. doi: 10.3389/fnhum.2013.00088

Guthrie, D., and Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology* 28, 240–244. doi: 10.1111/j.1469-8986.1991.tb00417.x

Kay, S. R., Fiszbein, A., and Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13, 261–276. doi: 10.1093/schbul/13.2.261

Keefe, R. S. E., Goldberg, T. E., Harvey, P. D., Gold, J. M., Poe, M. P., and Coughenour, L. (2004). The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophr. Res.* 68, 283–297. doi: 10.1016/j.schres.2003.09.011

Keil, J., Müller, N., Ihssen, N., and Weisz, N. (2012). On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states. *Cereb. Cortex* 22, 221–231. doi: 10.1093/cercor/bhr125

Kissler, J., and Herbert, C. (2013). Emotion, Etmnooi, or Emitoon? – Faster lexical access to emotional than to neutral words during reading. *Biol. Psychol.* 92, 464–479. doi: 10.1016/j.biopsycho.2012.09.004

Kissler, J., and Koessler, S. (2011). Emotionally positive stimuli facilitate lexical decisions — An ERP study. *Biol. Psychol.* 86, 254–264. doi: 10.1016/j.biopsycho.2010.12.006

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007

Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition – timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainresrev.2006.06.003

Knowland, V. C. P., Mercure, E., Karmiloff-smith, A., Dick, F., and Thomas, M. S. C. (2014). Audio-visual speech perception: a developmental ERP investigation. *Dev. Sci.* 1, 110–124. doi: 10.1111/desc.12098

Koh, Y., Soon, K., Sic, J., Choi, J., Kang, D., Hwan, J., et al. (2011). An MEG study of alpha modulation in patients with schizophrenia and in subjects at high risk of developing psychosis. *Schizophr. Res.* 126, 36–42. doi: 10.1016/j.schres.2010.10.001

Lange, J., Keil, J., Schnitzler, A., Dijk, H., Van Dijk, H., and Weisz, N. (2014). The role of alpha oscillations for illusory perception. *Behav. Brain Res.* 271, 294–301. doi: 10.1016/j.bbr.2014.06.015

Leavitt, V. M., Molholm, S., Ritter, W., Shpaner, M., and Foxe, J. J. (2007). Auditory processing in schizophrenia during the middle latency period (10 – 50 ms: high-density electrical mapping and source analysis reveal subcortical antecedents to early cortical deficits. *J. Psychiatry Neurosci.* 32, 339–353.

Lebib, R., Papo, D., Bode, S., De Bode, S., and Baudonnie, P. (2003). Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neurosci. Lett.* 341, 185–188. doi: 10.1016/S0304-3940(03)00131-9

Lee, T. W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441. doi: 10.1162/089976699300016719

Leicht, G., Kirsch, V., Giegling, I., Karch, S., Hantschk, I., Möller, H., et al. (2010). Reduced early auditory evoked gamma-band response in patients with schizophrenia. *Biol. Psychiatry* 67, 224–231. doi: 10.1016/j.biopsych.2009.07.033

Leske, S., Tse, A., Oosterhof, N. N., Hartmann, T., Müller, N., Keil, J., et al. (2014). The strength of alpha and beta oscillations parametrically scale with the strength of an illusory auditory percept. *Neuroimage* 88, 69–78. doi: 10.1016/j.neuroimage.2013.11.014

Magnée, M. J. C. M., Oranje, B., Engeland, H., Kahn, R. S., and Kemner, C. (2009). Neuropsychologia Cross-sensory gating in schizophrenia and autism spectrum disorder: EEG evidence for impaired brain connectivity? *Neuropsychologia* 47, 1728–1732. doi: 10.1016/j.neuropsychologia.2009.02.012

Martin, B., Giersch, A., Huron, C., and Van Wassenhove, V. (2013). Temporal event structure and timing in schizophrenia: preserved binding in a longer "'now"'. *Neuropsychologia* 51, 358–371. doi: 10.1016/j.neuropsychologia.2012.07.002

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

Müller, N., Keil, J., Obleser, J., Schulz, H., Grunwald, T., Bernays, R.-L., et al. (2013). You can't stop the music: reduced auditory alpha power and coupling between auditory and memory regions facilitate the illusory perception of music during noise. *Neuroimage* 79, 383–393. doi: 10.1016/j.neuroimage.2013.05.001

Murray, M. M., Brunet, D., and Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr.* 20, 249–264. doi: 10.1007/s10548-008-0054-5

Muthukumaraswamy, S. D., and Singh, K. D. (2013). Visual gamma oscillations: the effects of stimulus type, visual field coverage and stimulus motion on MEG and EEG recordings. *Neuroimage* 69, 223–230. doi: 10.1016/j.neuroimage.2012.12.038

Myslobodsky, M., Golderberg, T., Johnson, F., Hicks, L., and Weinberger, D. (1992). Lipreading in patients with schizophrenia. *J. Nerv. Ment. Dis.* 180, 168–171. doi: 10.1097/00005053-199203000-00004

Pearl, D., Yodashkin-Porat, D., Katz, N., Valevski, A., Aizenberg, D., Sigler, M., et al. (2009). Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups. *Compr. Psychiatry* 50, 186–192. doi: 10.1016/j.comppsych.2008.06.004

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37, 127–152. doi: 10.1111/1469-8986.3720127

Popov, T., Jordanov, T., Weisz, N., Elbert, T., Rockstroh, B., and Miller, G. A. (2011). Evoked and induced oscillatory activity contributes to abnormal auditory sensory gating in schizophrenia. *Neuroimage* 56, 307–314. doi: 10.1016/j.neuroimage.2011.02.016

Popov, T., Rockstroh, B., Popova, P., Carolus, A., and Miller, G. (2014). Dynamics of alpha oscillations elucidate facial affect recognition in schizophrenia. *Cogn. Affect. Behav. Neurosci.* 14, 364–377. doi: 10.3758/s13415-013-0194-2

Roa Romero, Y., Senkowski, D., and Keil, J. (2015). Early and late beta band power reflect audiovisual perception in the McGurk Illusion. *J. Neurophysiol.* 13, 2342–2350. doi: 10.1152/jn.00783.2014

Rosburg, T., Boutros, N. N., and Ford, J. M. (2008). Reduced auditory evoked potential component N100 in schizophrenia — A critical review. *Psychiatry Res.* 161, 259–274. doi: 10.1016/j.psychres.2008.03.017

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Molholm, S., Javitt, D. C., and Foxe, J. J. (2007). Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr. Res.* 97, 173–183. doi: 10.1016/j.schres.2007.08.008

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., and Foxe, J. (2007). Seeing voices: high-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia* 45, 587–597. doi: 10.1016/j.neuropsychologia.2006.03.036

Schurger, A., Cowey, A., Cohen, J. D., Treisman, A., and Tallon-Baudry, C. (2008). Distinct and independent correlates of attention and awareness in a hemianopic patient. *Neuropsychologia* 46, 2189–2197. doi: 10.1016/j.neuropsychologia.2008.02.020

Senkowski, D., Schneider, T. R., Foxe, J. J., and Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci.* 31, 401–409. doi: 10.1016/j.tins.2008.05.002

Spencer, K. M., Niznikiewicz, M. A., Shenton, M. E., and Mccarley, R. W. (2008). Sensory-evoked gamma oscillations in chronic schizophrenia. *Biol. Psychiatry* 63, 744–747. doi: 10.1016/j.biopsych.2007.10.017

Stekelenburg, J. J., Maes, J. P., Van Gool, A. R., Sitskoorn, M., and Vroomen, J. (2013). Deficient multisensory integration in schizophrenia: an event-related potential study. *Schizophr. Res.* 147, 253–261. doi: 10.1016/j.schres.2013.04.038

Stekelenburg, J. J., and Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* 19, 1964–1973. doi: 10.1162/jocn.2007.19.12.1964

Stephan, K. E., Baldeweg, T., and Friston, K. J. (2006). Synaptic plasticity and dysconnection in schizophrenia. *Biol. Psychiatry* 59, 929–939. doi: 10.1016/j.biopsych.2005.10.005

Stone, D. B., Coffman, B. A., Bustillo, J. R., Aine, C. J., Stephen, J. M., and Moran, R. J. (2014). Multisensory stimuli elicit altered oscillatory brain responses at gamma frequencies in patients with schizophrenia. *Front. Hum. Neurosci.* 8:788. doi: 10.3389/fnhum.2014.00788

Stone, D. B., Urrea, L. J., Aine, C. J., Bustillo, J. R., Clark, V. P., and Stephen, J. M. (2011). Unisensory processing and multisensory integration in schizophrenia: a high-density electrical mapping study. *Neuropsychologia* 49, 3178–3187. doi: 10.1016/j.neuropsychologia.2011.07.017

Surguladze, S. A., Calvert, G. A., Brammer, M. J., Bullmore, E. T., Giampietro, V., and David, A. S. (2001). Audio-visual speech perception in schizophrenia: an fMRI study. *Psychiatry Res.* 106, 1–14. doi: 10.1016/S0925-4927(00)00081-0

Szycik, G. R., Münte, T. F., Dillo, W., Mohammadi, B., and Samii, A. (2009). Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study. *Schizophr. Res.* 110, 111–118. doi: 10.1016/j.schres.2009.03.003

Tan, H. M., Lana, L., Uhlhaas, P. J., Spencer, K. M., and Boston, V. A. (2013). High-frequency neural oscillations and visual processing deficits in schizophrenia. *Front. Psychol.* 4:621. doi: 10.3389/fpsyg.2013.00621

Uhlhaas, P. J., and Singer, W. (2006). Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron* 52, 155–168. doi: 10.1016/j.neuron.2006.09.020

Uhlhaas, P. J., and Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nat. Rev. Neurosci.* 11, 100–113. doi: 10.1038/nrn2774

White, T. P., Wigton, R. L., Joyce, D. W., Bobin, T., Ferragamo, C., Wasim, N., et al. (2014). Eluding the illusion? Schizophrenia, dopamine and the McGurk effect. *Front. Hum. Neurosci.* 8:565. doi: 10.3389/fnhum.2014.00565

Williams, L. E., Light, G. A., Braff, D. L., and Ramachandran, V. S. (2010). Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia* 48, 3128–3136. doi: 10.1016/j.neuropsychologia.2010.06.028

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4 Third study: Reduced frontal theta oscillations underlie impaired crossmodal prediction error processing in schizophrenia

Roa Romero, Y.; Keil, J; Balz, J.; Gallinat, J. and Senkowski, D. (February 2016). Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia. *Journal of Neurophysiology* (*submitted* 2^{nd} *February*).

Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia

Yadira Roa Romero[1*], Julian Keil[1], Johanna Balz[1], Jürgen Gallinat[2], Daniel Senkowski[1]

[1]Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin Berlin Hospital, St. Hedwig Hospital, Große Hamburger Straße 5-11, 10115 Berlin, Germany

[2]Department for Psychiatry & Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20246 Hamburg, Germany

Running title: Crossmodal prediction error processing in schizophrenia

Keywords: Predictive Coding; Audiovisual Speech; Neural Synchrony; Multisensory; Oscillatory Activity

Number of words in the abstract = 250
Number of words in the main text = 3824
Number of tables = 3
Number of figures =8

*** Corresponding author:**

Yadira Roa Romero
Dept. of Psychiatry and Psychotherapy
Charité-Universitätsmedizin Berlin
St. Hedwig Hospital
Große Hamburger Str. 5-11
10115 Berlin, Germany
Phone: +49-30-2311-2739
Fax: +49-30-2311-2209
Email: yadira.roa-romero@charite.de

Introduction

Learning stimulus regularities allow us to correctly predict future events. In case of a mismatch between predicted and actual events neural activity that relates to the processing of the prediction error (PE) is observed (Rao & Ballard, 1999). Failures in correctly predicting future events likely contribute to the schizophrenia (SZ) psychopathology, including delusions and hallucinations (Fletcher & Frith, 2009). Previous studies using unisensory stimuli, such as auditory speech, have shown altered PE processing in SZ (Ford & Mathalon, 2012; G. K. Murray et al., 2008; Ross et al., 2007; Stekelenburg, Pieter, et al., 2013). In our environment, however, stimuli occur in different sensory modalities and predictions need to be generated across the senses. Whether SZP show alterations in crossmodal PE processing is unknown. Thus far, relatively few studies have investigated multisensory processing in SZ (H. Tseng et al., 2015). One behavioral study in SZP revealed a diminished benefit of viewing lip movements for the recognition of auditory speech that is presented at an intermediate noise level (Ross et al., 2007). Moreover, an electroencephalography (EEG) study suggested altered audiovisual processing, as reflected in event-related potentials (ERPs), using non-speech stimuli (Stekelenburg, Pieter, et al., 2013). Together, this suggests a multisensory integration deficit in this patient group ((Ross et al., 2007; Stekelenburg, Pieter, et al., 2013) but see (Martin et al., 2013)). A neural mechanism that has been linked to multisensory processing in healthy participants is synchronized oscillatory activity (Senkowski et al., 2008). Neural oscillations are involved in perceptual processing (Engel & Fries, 2010) and contribute to the computation of crossmodal PE in audiovisual speech (Arnal & Giraud, 2012; Arnal et al., 2011). Using magnetoencephalography, Arnal et al. (2011) observed an enhancement of theta-band (i.e. 4-7 Hz) phase synchrony in higher speech processing areas during the presentation of congruent audiovisual speech in healthy participants. The same study revealed elevated beta-band (i.e. 14-15 Hz) phase synchrony and gamma power (i.e. 60-80 Hz) in higher multisensory areas when crossmodal predictions were violated. Recently, Lange et al. (2013) found increased theta-band power in the auditory cortex during the processing of incongruent compared to congruent audiovisual speech. Interestingly, altered oscillatory responses in unisensory paradigms have been previously reported in SZP (Grützner, Wibral, Sun, et al., 2013; Popov et al., 2014; Popov, Wienbruch, et al., 2015; Uhlhaas & Singer, 2010). For example, SZP show altered frontal theta-band power during the processing of color-word incongruence (Popov, Wienbruch, et al., 2015). This finding is in line with the proposed crucial role of dysfunctional oscillations in the frontal cortex for the SZ psychopathology (Senkowski & Gallinat, 2015b). Notably, a recent study indicated that altered neural oscillations also contribute to multisensory processing deficits in SZ (Stone et al., 2014). Hence, dysfunctional neural oscillations might underlie altered crossmodal PE processing in SZ. In this high-density EEG study we investigated crossmodal PE processing in SZP and HC. We adapted an audiovisual speech paradigm for which robust crossmodal PE has been previously reported in healthy participants (Arnal et al., 2011). We investigated crossmodal PE processing in ERPs and neural oscillations. The analysis of ERPs revealed similar detection of audiovisual incongruence in the auditory cortex of SZP and HC. In addition, we found group differences in theta-band oscillations that were source localized in the frontal cortex and linked to the SZ psychopathology.

Methods

Participants

Twenty-two patients with the DSM-4 diagnosis schizophrenia were recruited from outpatient units of the Charité–Universitätsmedizin Berlin. Diagnosis was assessed by the treating psychiatrist or through chart review. Severity of symptoms was obtained by the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1987). In accordance with the 5-factor model, symptoms were grouped into factors “positive”, “negative”, “depression”, ”excitement” and “disorganization” (Wallwork, Fortgang, Hashimoto, Weinberger, & Dickinson, 2012). Due to an insufficient number of trials in EEG data (i.e. at least 30 trials per condition), five patients were excluded from the analysis. Data from seventeen patients (5 female, 35.24 ± 7.73 years) and seventeen education, handedness, gender, and age matched HC (4 female, 36.00 ± 8.29 years), who were screened for psychopathology with the German version of the Structured Clinical Interview for DSM-4-R Non-Patient Edition (SCID), were subjected to the analysis (Table 1). In all participants the Brief Assessment of Cognition in Schizophrenia (BACS) was assessed (Keefe et al., 2004). All participants gave written informed consent, had normal hearing as well as normal or corrected to normal vision, and no record of neurological disorders. None of them met criteria for alcohol or substance abuse. A random sample of 45 % of participants underwent a multi drug-screening test. The study was performed in accordance with the Declaration of Helsinki and the local ethics commission approved the study.

Table 1

	Patients		**Controls**		**Statistics**	
	Mean	**SD**	**Mean**	**SD**	**t-values**	**p-values**
Age (years)	35.24	7.73	36.00	8.29	-.278	.783
Education (years)	11.00	1.66	11.18	1.47	.329	.745
Illness duration (years)	8.24	4.47	-	-	-	-
Chlorpromazin Eq. (daily dosage/ mg)	381.12	191.65	-	-	-	-
	N		**N**			
Gender (m/f)	12/5		13/4		-	-
Handedness (r/l)	14/3		16/1		-	-
Antipsychotic Med.	17		-		-	-
Co-medication*	3		-		-	-
BACS						
Verbal Memory	41.24	13.80	48.06	10.63	-1.615	.116
Digit	19.76	4.47	20.29	4.55	-.342	.734
Motor	65.53	12.33	75.71	9.41	-2.705	.011
Fluency	47.18	13.87	53.24	14.06	-1.265	.215
Symbol coding	54.82	15.00	58.35	14.05	-.708	.484
ToL	17.12	3.04	17.82	2.60	-.727	.472
Total score	245.65	43.31	273.47	37.37	-2.005	.053
PANSS						
Negative	18.65	3.12	-	-	-	-
Positive	16.59	1.97	-	-	-	-
General	37.18	2.83	-	-	-	-
Total score	72.41	5.80	-	-	-	-

Table 1.

Overview of demographic data. *Co-medication of antipsychotics and mood stabilizers.

Stimuli and experimental design

Video clips of a female actress uttering the syllables /Pa/, /La/, /Ta/, /Ga/, and /Fa/ were recorded using a digital camera (Canon 60D, 50 frames per second, 1280x720px, 44.1 Khz stereo audio) and exported with 30 frames per second (Apple Quicktime Player, Version 7; Supplementary Material 1). The experiment consisted of 640 trials that were presented in 16 blocks. Each block had a duration of about 3 minutes. Stimuli were controlled via PsychToolbox (Brainard, 1997). Different types of congruent and incongruent audiovisual syllable combinations were presented (Table 2). Congruent syllable trials contained matching audiovisual syllables (e.g., visual /Pa/ and auditory /Pa/), whereas incongruent syllable trials contained non-matching audio-visual syllables (e.g., visual /Pa/ and auditory /Ta/). Pilot data verified that syllables differed with respect to the predictive power of visual on auditory stimuli (Supplementary Material 2). Predictability of visemes relates to the place of articulation of syllables in the vocal tract. In accordance with a previous study (Arnal et al., 2011) and in line with pilot data, we selected syllables in which the visual input was high (e.g., /Pa/ and /La/) or low predictive (e.g., /Ga/ and /Ta/) for the auditory input (Figure 1).

Table 2

Condition	**Visual**	**Audio**	**N**
Congruent HP	Pa	Pa	80
Congruent HP	La	La	80
Congruent LP	Ga	Ga	80
Congruent LP	Ta	Ta	80
Incongruent HP	Pa	La	80
Incongruent HP	La	Pa	80
Incongruent LP	Ga	Ta	80
Incongruent LP	Ta	Ga	80
Congruent /Fa/	Fa	Fa	48
Incongruent /Fa/	Pa	Fa	12
Incongruent /Fa/	La	Fa	12
Incongruent /Fa/	Ga	Fa	12
Incongruent /Fa/	Ta	Fa	12

Table 2.
Overview of presented congruent and incongruent, high and low predictive trials.

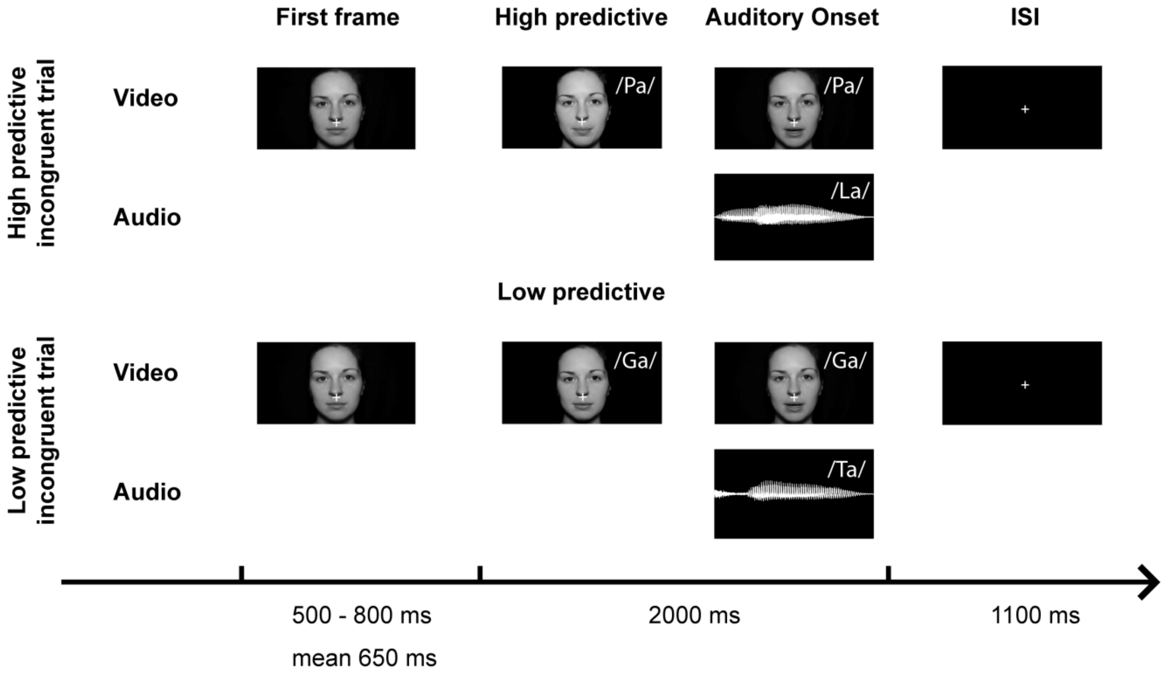

Figure 1

Trial sequence of high and low predictive incongruent audiovisual syllables. Each trial started with the presentation of the first static frame of a video clip that was presented for a random interval ranging from 500 to 800 ms. Following this frame the video clip, in which a female speaker uttered a syllable, was presented for 2000 ms. The participants' tasks were to respond to the occasional auditory syllable /fa/ and to an occasional change of the fixation cross, which turned into a white circle. The tasks ensured that participants were attending to the sensory inputs.

Four types of audiovisual stimuli were presented: high predictive congruent; high predictive incongruent; low predictive congruent; low predictive incongruent. We included audiovisual and visual only catch trials to warrant sustained auditory and visual attention. Audiovisual catch trials comprised 96 trials in which the visual syllable /Fa/ (N = 48), /Pa/ (N = 12), /La/ (N = 12), /Ga/ (N = 12), or /Ta/ (N = 12) was combined with the auditory syllable /Fa/. Visual only catch trials (N = 64) consisted of a white circle that appeared randomly for 270 ms during audiovisual syllable presentation. Participants had to press a button with their right index finger when they heard the syllable /Fa/ or when they detected the white circle.

Acquisition and processing of EEG data

Data were recorded using a 128 channel active EEG system (EasyCap, Herrsching, Germany), including one horizontal and one vertical EOG electrode (online recording: 1000 Hz sampling rate with a 0.016 – 250 Hz bandpass filter; offline filtering: downsampling to 500 Hz, 1 – 125 Hz FIR bandpass filtering and 49.1 – 50.2 Hz, 4th order Butterworth notch filtering). To correct for EOG and ECG artifacts, independent component analyses were conducted (Lee et al., 1999). After visual inspection, 14 ± 0.6 (standard error of mean) independent components for SZP and 15 ± 0.9 components for HC were rejected. Remaining noisy channels (SZP = 16 ± 1.2; HC = 15 ± 1.4) were interpolated using spherical interpolation. Extracted epochs of -1000 ms to 1000 ms around auditory onset were re-referenced to common average. For ERP analysis epochs were filtered (2 Hz high pass, 2nd order , 35 Hz low pass, 12th order two-pass Butterworth filter) and baseline corrected using an interval from -500 ms to -100 ms prior to sound onset. In the analysis of oscillations we explored a range from 4 to 100 Hz. Since we did not find robust modulations of high frequency (i.e. > 30 Hz) responses, we focused the analysis on 4 to 30 Hz. For the analysis of oscillations multitaper convolution transformation with frequency dependent Hanning window was computed in 1 Hz steps (time window: $\Delta t = 3/f$, spectral smoothing: $f = 1/\Delta t$). Averaged oscillatory responses were baseline corrected (relative change) from -500 to -100 ms prior to sound onset.

Analysis of behavioral data

For the analysis of reaction times (RTs) to audiovisual catch trials a repeated measures ANOVA with factors Group (SZP vs. HC) and Condition (congruent vs. incongruent) was conducted. The factor predictability was not investigated, due to a lack of systematic variation in audiovisual catch trials. For the analysis of RTs to visual catch trials a t-test was conducted between groups. Since hit rates (HRs) for audiovisual and visual catch trials were at ceiling level, non-parametric Mann-Whitney U tests were calculated to compare between groups. Wilcoxon-signed rank tests were computed to compare congruent and incongruent

HRs within groups. The significance level in these tests was set to a Bonferroni corrected p-value of 0.0125.

Analysis of EEG data

The analysis focused on global field power (GFP; (Esser et al., 2006)) and oscillatory responses to high and low predictive congruent and incongruent trials. For each condition 84 ± 10 trials for SZP and 88 ± 12 trials for HC were entered into the analysis. For the analysis GFPs and oscillatory responses to incongruent and congruent conditions for SZP and HC were subtracted (e.g., responses to high predictive congruent syllables were subtracted from responses to high predictive incongruent syllables). This ensured that physical differences between low and high predictive syllables cannot account for the results. The differences between congruent and incongruent syllables were entered into running repeated measures ANOVAs with the factors Group (SZP vs. HC) and Predictability (low vs. high). GFP was calculated across all scalp electrodes as a measure of electric field strength separately for all conditions. Running ANOVAs were computed for each sample point in a 0 to 1000 ms interval following auditory syllable onset. To account for multiple testing, a time stability criterion of at least 20 ms was applied (Picton et al., 2000). Main effects or interactions in the analysis were followed-up by a local autoregressive average (LAURA) analysis using Cartool (Brunet, Murray, & Michel, 2011). Grand average EEG data were fitted into a realistic head model comprising 4024 nodes. Source space was restricted to gray matter of the Montreal Neurological Institute's (MNI) average brain, divided into a regular source grid with 6 mm spacing. For each group and predictability level the source solutions of the congruent condition were subtracted from the incongruent condition.

To examine effects in neural oscillations we followed a protocol of our recent audiovisual speech processing study (Roa Romero et al., 2015). Oscillatory power (4-30 Hz) was analyzed for a region of interest, encompassing 10 fronto-central electrodes. Running ANOVAs with the same factors as in the GFP analysis were computed for each sample point and frequency in a 0 to 1000 ms interval. Due to low temporal resolution of the time-frequency transformation, a time stability criterion of at least 100 ms was applied. Source estimations of oscillatory responses were computed for significant time-frequency windows obtained in the ANOVA (i.e. 7.5 ± 2 Hz). For each participant oscillatory power was projected into source space using dynamic imaging of coherent sources (DICS) algorithm (Gross et al., 2001). All conditions were combined and sensor level cross-spectral density was calculated for an interval between 100 and 500 ms and a baseline interval between -500 to -100 ms prior to auditory syllable onset. A common spatial filter was computed on the combined data and boundary element models based on individual magnetic resonance images (MRI) were calculated. Each condition was projected into source space through the common filter. Finally, source activity was interpolated onto individual anatomical MRIs and normalized to MNI brain and averaged over subjects. Pearson correlations were computed between PANSS scores, anti-psychotic medication, and EEG data in SZP. To examine the influence of anti-psychotic medication, medication dosage was converted to chlorpromazine equivalent level (Gardner et al., 2010) and entered as covariate into partial correlation analyses**.** Bonferroni-correction was applied to control for multiple comparisons.

Results

Behavioral results

The ANOVA for RTs to audiovisual catch trials revealed a significant main effect of Group (F(1,32) = 11.24 , p = 0.002), due to faster responses in HC compared to SZP. Furthermore, there was a significant main effect of Condition (F(1,32) = 22.40 , p < 0.001), indicating faster RTs for congruent compared to incongruent syllables (t(33)= -4.80, p < 0.001). The analysis of RTs to visual only catch trials revealed no significant differences between SZP and HC (t(32) = -1.387 , p = 0.175). The HRs to audiovisual catch trials (average across conditions: SZP = 93.48 %, HC= 96.90 %) and visual only catch trials (average across conditions: SZP= 96.54 %; HC= 98.82 %) were at ceiling level. The analysis of HRs to audiovisual trials revealed significantly lower HRs in SZP compared to HC for congruent stimuli (U= 53.5, z = -3.29, p = 0.001). No significant group difference was found for incongruent stimuli (U= 134.5, z = - 0.35, p = 0.726). Within the HC group HRs were larger in congruent compared to incongruent audiovisual trials (Z= - 2.878, p = 0.004). No differences in HRs between congruent and incongruent trials were found in the SZP (Z= - 0.535, p = 0.593). Finally, HRs for visual catch trials did not significantly differ between groups (U= 94.5, z = -1.81, p = 0.07), indicating that both groups similarly attended to visual inputs.

Global field power

The running ANOVA with the factors Group (SZP vs. HC) and Condition (low vs. high predictability) revealed a significant main effect of Condition between 206 and 250 ms after auditory onset (F(1,32) =17.25, p < 0.001; Figure 2). This effect, which was observed in both groups, was due to larger GFP for high (average across groups = 0.70 uV^2) compared to low (0.65 uV^2) predictive syllables. ERP topographies for the time window of the significant GFP main effect revealed a stronger fronto-central activation during high compared to low predictive trials (Figure 3). LAURA analysis identified the sources of this effect in a region encompassing the auditory cortex (Figure 4). No other main effects or interactions were found.

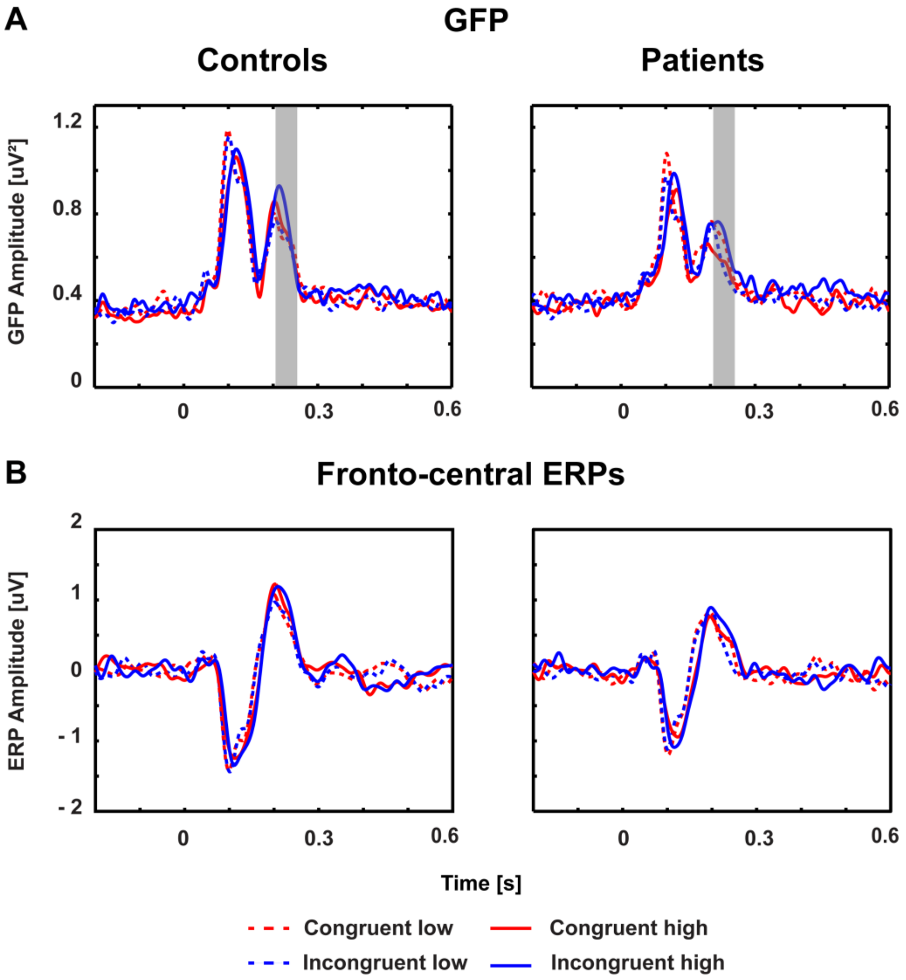

Figure 2

Global field power (GFP) and event-related potentials (ERPs) at fronto-central scalp electrodes. A. Traces of GFP in HC (left panel) and SZP (right panel) for congruent (red lines) and incongruent (blue lines) trials. Dashed lines represent low and solid lines represent high predictive trials. Time zero denotes the onset of the auditory syllable. The analysis of GFP revealed a significant condition effect in the 206 to 250 ms interval (highlighted in gray). B. Traces of ERPs in HC (left panel) and SZP (right panel) for congruent (red lines) and incongruent (blue lines) trials. Dashed lines represent low and solid lines high predictive trials. ERPs revealed a typical N1-P2 amplitude pattern to the auditory syllable onset, with smaller amplitudes in SZP compared with HC.

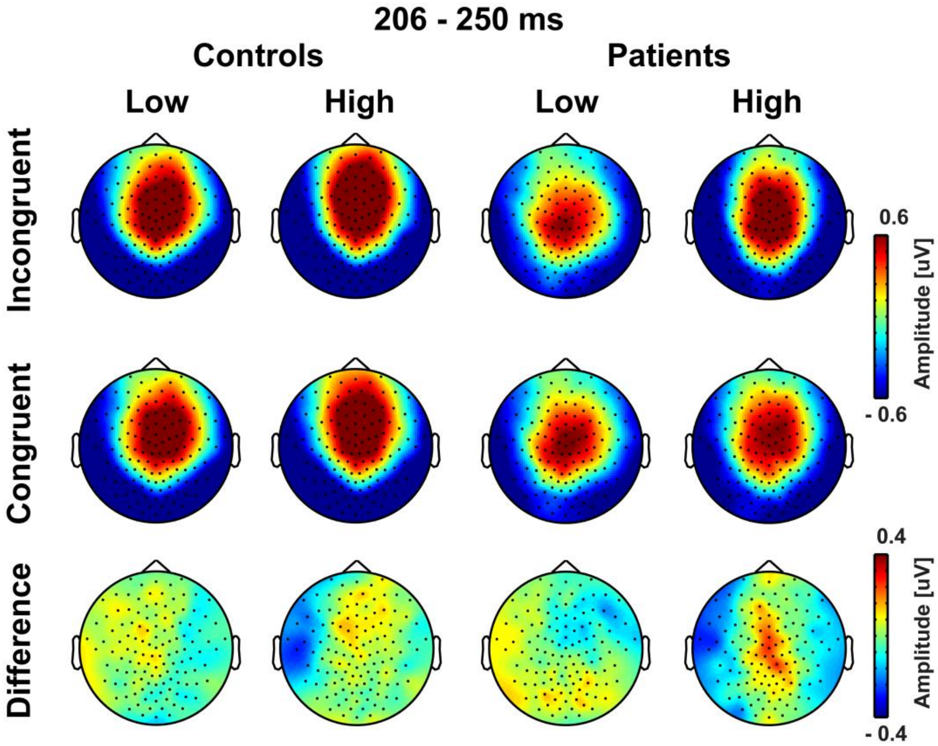

Figure 3

Topographies of event-related potentials to the four experimental stimulus types, as well as for the difference between incongruent and congruent trials during the 206 to 250 ms interval.

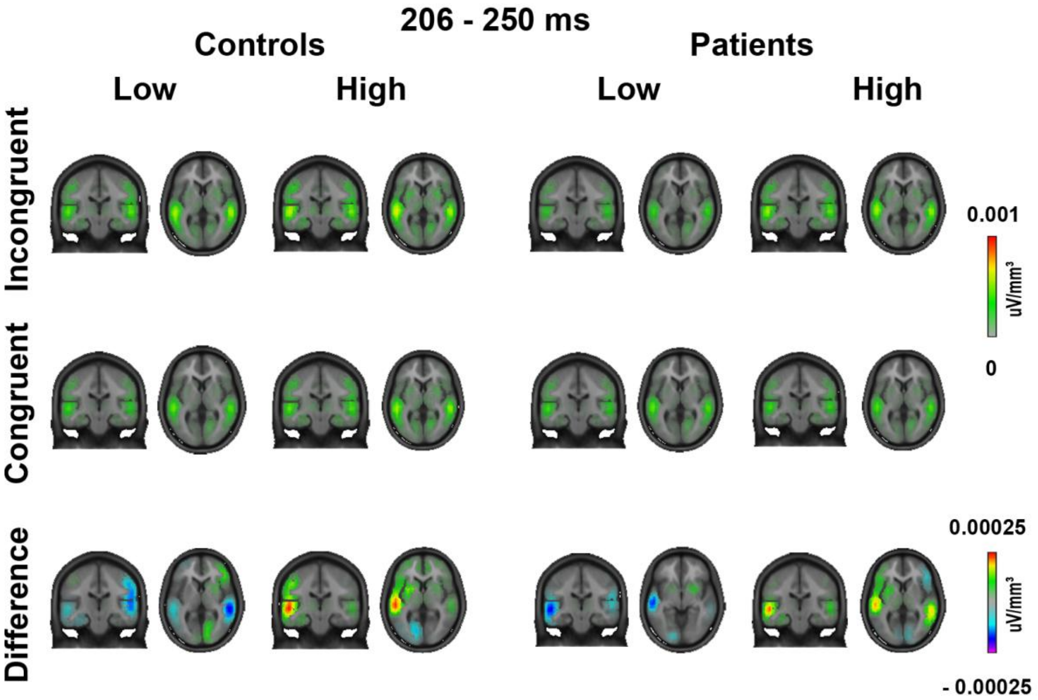

Figure 4

Local autoregressive average source estimation of ERPs for the four experimental stimulus types and the difference between incongruent and congruent trials during the 206 to 250 ms interval. The figure shows similar incongruence detection in SZP and HC that were localized in the auditory cortex.

Oscillatory activity

The running ANOVA with the factors Group (SZP vs. HC) and Condition (low vs. high predictability) revealed a significant main effect of Group between 230 and 370 ms in the theta-band, centered at 7.5 Hz (F(1,32) =7.66, p = 0.015). Across conditions, theta-band power was stronger in HC compared to SZP. The ANOVA revealed an additional main effect of Condition between 210 and 380 ms, centered at 7.5 Hz (F(1,32) = 13.73, p = 0.008). This effect was due to a stronger theta-band power in high compared to low predictive trials. Most relevant, the ANOVA revealed a Group x Condition interaction between 230 and 370 ms, centered at 7.5 Hz (F(1,32) = 7.42 ,p = 0.011; Figures 5 and 6). Follow-up analyses, which were conducted separately for each group revealed stronger theta-band power in high compared to low predictive trials, particularly in HC (t(16) = - 4.06, p < 0.001). No such effect was found in SZP (t(16) = - 0.67, p = 0.507). Follow-up analyses revealed that frontal theta-band power was larger in HC compared to SZP in high predictive (t(32) = 3.21, p = 0.003) but not in low predictive trials (t(32) = 0.26, p = 0.796). Finally, DICS source analysis revealed that the theta-band power effects are localized in frontal cortex, encompassing anterior cingulate cortex (ACC) and medial frontal cortex (MFC; Figure 7). Frontal theta-band power between high and low predictive trials differed in HC but not in SZP. Figure 8 provides an overview on the effects obtained in GFP and oscillatory responses.

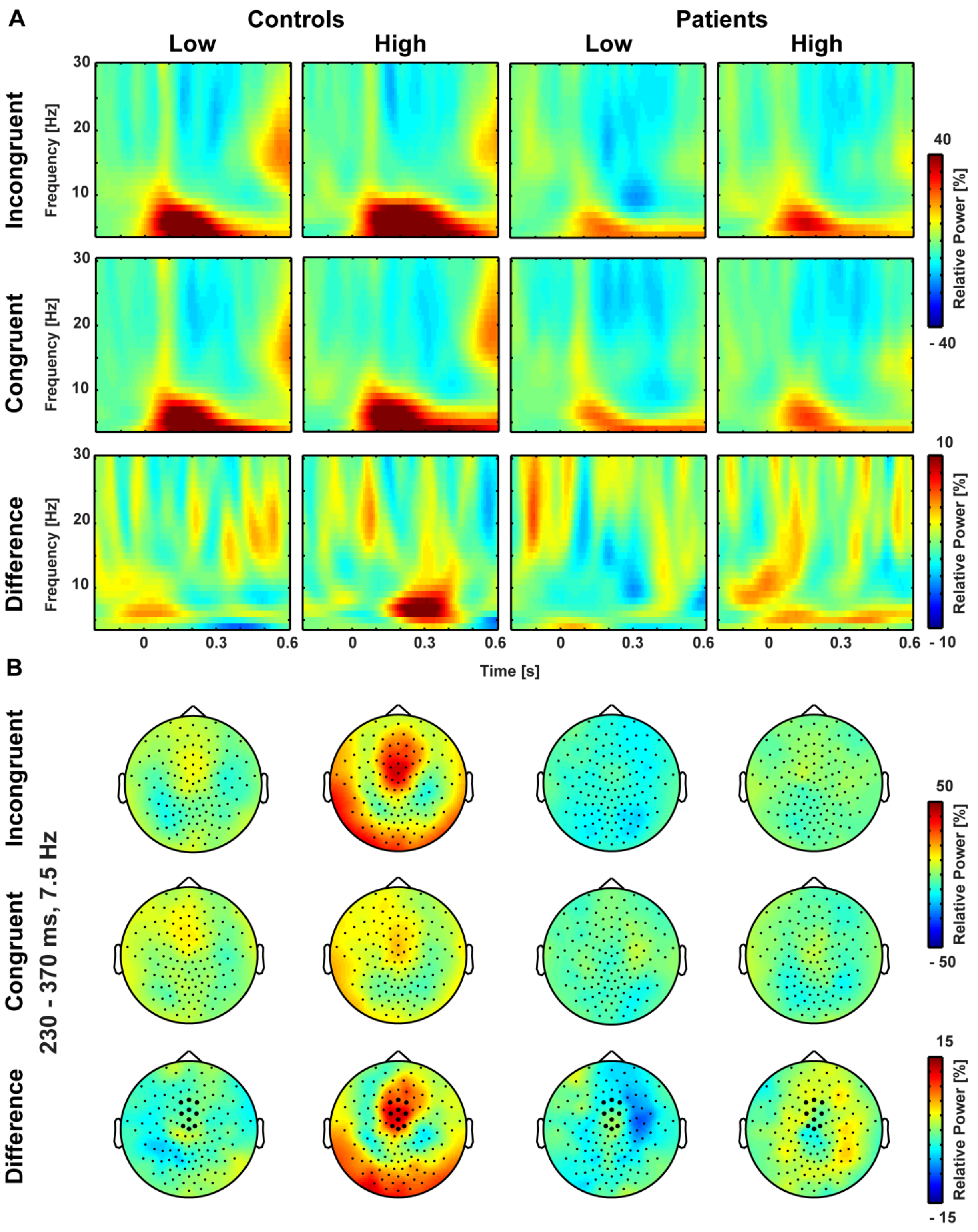

Figure 5

Time-frequency responses and topographies of oscillatory power. A. Time-frequency responses of 4 to 30 Hz power to the four experimental stimulus types and the difference between incongruent and congruent trials at fronto-central scalp electrodes. Time zero denotes the onset of the auditory syllable. The differences between incongruent and congruent trials (lower panel) revealed enhanced theta-band power (i.e. 7 Hz) around 300 ms, particularly in the control group. B. Topographies of fronto-central theta-band power. The bold dots in the lower panel (differences) denote the electrode group in which a significant group by condition interaction was found in the 230 to 370 ms interval. Theta-band power was enhanced in the control group, especially in the incongruent high condition.

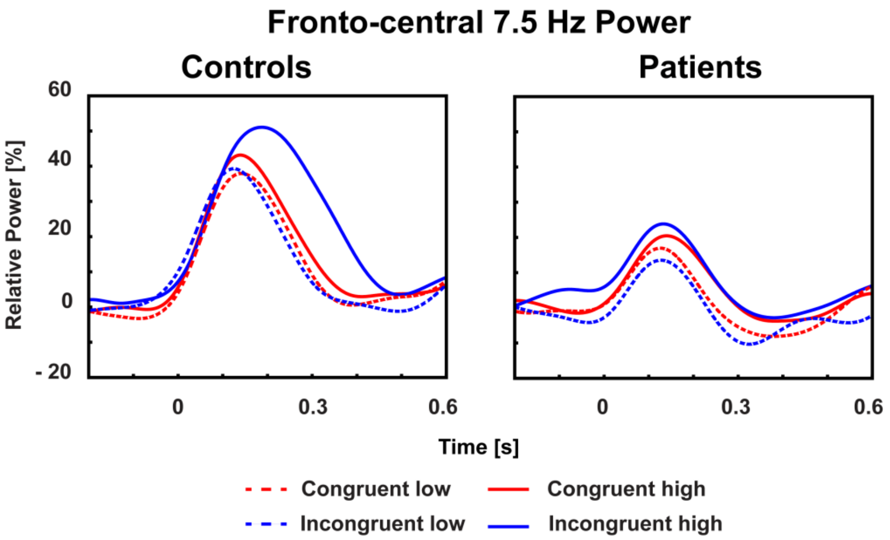

Figure 6

Traces of fronto-central theta-band power. Traces of theta-band power in HC (left panel) and SZP (right panel) for congruent (red lines) and incongruent (blue lines) trials at fronto-central electrodes. Dashed lines represent low and solid lines represent high predictive trials. The time course of theta-band power was similar in HC and SZP, while the amplitude was substantially reduced in SZP, most strongly in the incongruent high condition.

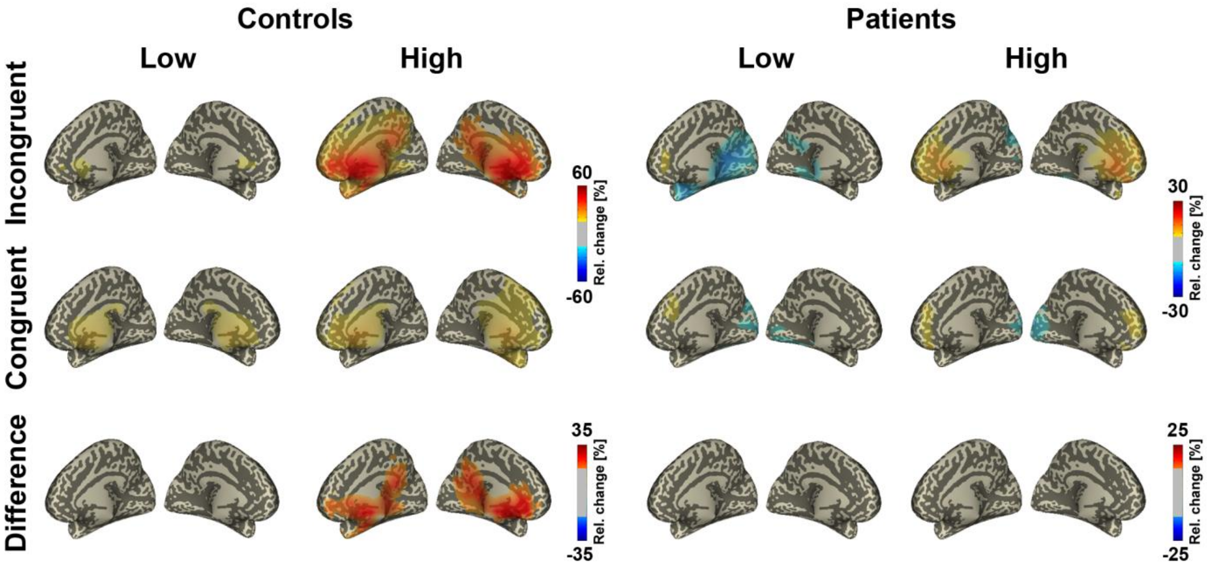

Figure 7

DICS source analysis of oscillatory power in the theta-band. The source of theta-band power was localized in the medial frontal gyrus (MFG) and anterior cingulate cortex (ACC). In HC crossmodal PE processing was found in frontal areas, particularly in the high predictive condition. No such effect was observed in SZP.

Relationships between EEG data, medication and clinical symptoms

The significance of the frontal theta-band power deficit for the SZ psychopathology was investigated by calculating partial correlations between the difference (i.e. incongruent minus congruent) of frontal theta-band power in the high predictive condition, medication dosage, and PANSS scores. Medication dosage did not correlate with frontal theta-band power ($p > 0.33$). Importantly, the partial correlation analyses between the theta-band power difference and PANSS scores revealed a trend towards a negative relationship with the factor Excitement ($r = -.593$, $p = 0.015$, Bonferroni corrected $\alpha = 0.01$). No other significant relationships were found (all p-values > 0.21).

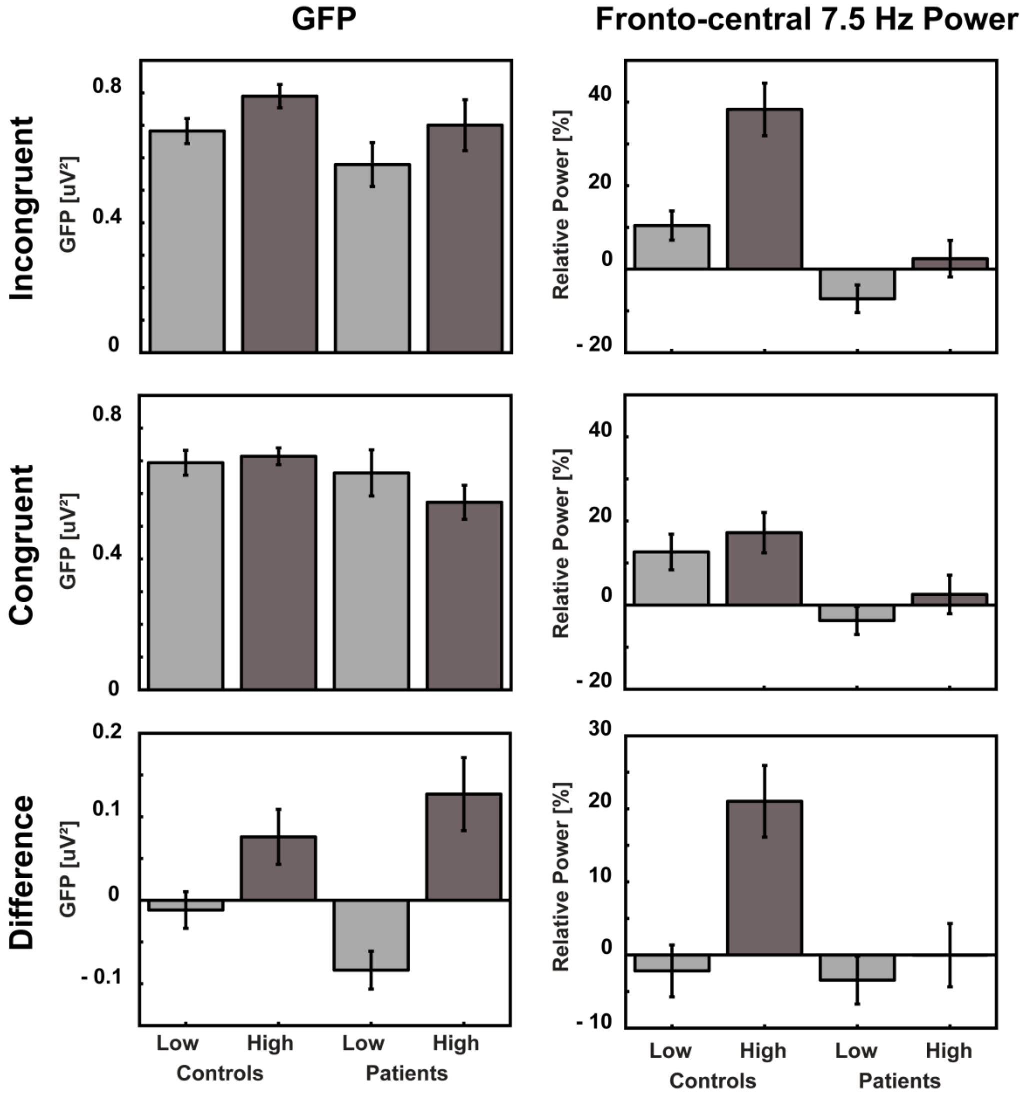

Figure 8

Summary of main results. Amplitudes of GFP (left panel) and 7.5 Hz power (right panel) at fronto-central electrodes. Amplitudes of GFP are depicted for the interval of 206 to 250ms. GFP differences (incongruent minus congruent trials) were positive in high and negative in low predictive trials. The pattern of effects did not differ between groups (lower left panel). Amplitudes of 7.5 Hz power are depicted for the 230 to 370 ms interval. Power differences between incongruent and congruent trials were stronger in high compared with low predictive trials, particularly in the HC group (lower right panel).

Discussion

In this study we examined the neural signatures underlying crossmodal PE processing in SZP and HC. In both groups we observed effects of incongruence detection in GFP, which were localized in the auditory cortex. Moreover, we obtained group differences in frontal theta-band power. Theta-band power was larger for PE processing of high compared with low predictive audiovisual syllables in HC, whereas no such effect was found in SZP.

Audiovisual stimulus congruence modulates behavioral performance in patients and controls

RTs to audiovisual target syllables were longer in SZP compared to HC. This is in line with previous audiovisual processing studies using simple (e.g., beeps and flashes) and complex (e.g., natural speech) audiovisual stimuli, showing prolonged RTs (Williams et al., 2010) and decreased HRs (Gelder et al., 2005; Jong, Hodiamont, Stock, & Gelder, 2009) in SZP. Additionally, we observed longer RTs to incongruent compared to congruent audiovisual targets in both groups. Previous studies in healthy participants have also reported facilitation effects for congruent compared to incongruent audiovisual speech stimuli (Grant & Seitz, 2000). Notably, only HC showed a higher HR to congruent compared to incongruent audiovisual targets. The HRs in SCZ did not differ between conditions. This indicates that stimulus incongruence had a stronger impact on behavioral performance in HC than in SZP, which is in line with the previous finding that SZP do not benefit from congruent audiovisual information in audiovisual speech perception (10).

Intact audiovisual incongruence detection in the auditory cortex of schizophrenia patients

The analysis of GFP revealed differences between high and low predictive audiovisual syllables at 206 ms to 250 ms after auditory onset. In both groups GFP amplitude differences between incongruent and congruent trials was larger in high compared with low predictive trials. Source estimation revealed that the effect of visual predictability on audiovisual speech processing is localized in the auditory cortex. Interestingly, the pattern of effects was similar in SZP and HC. This suggests that the strength of visually induced prediction influences the processing of audiovisual speech in auditory cortex similarly in both groups. Hence, our finding indicates an intact detection of audiovisual stimulus incongruence in the auditory cortex of SZP. Previously, Klucharev, Möttönen, & Sams (2003) examined late phonetic audiovisual interactions in healthy participants and found larger ERPs for incongruent compared with congruent audiovisual speech stimuli at 235 ms to 325 ms. Further studies revealed that sources of this late phonetic audiovisual interaction are localized in the auditory cortex (Möttönen, Krause, Tiippana, & Sams, 2002; Möttönen, Schürmann, & Sams, 2004). Our findings also correspond with evidence from an ERP study on audiovisual speech processing in SZP and HC (Stekelenburg, Pieter, et al., 2013). This study found no group differences in P2 amplitudes in response to congruent and incongruent audiovisual speech stimuli. Taken together, our data suggest that high and low predictive visual speech information influences stimulus processing in the auditory cortex similarly in SZP and HC. However, the degree of crossmodal prediction induced by visual speech information, i.e. high vs. low predictive, differentially modulates audiovisual processing in the auditory cortex. High predictive visual information leads to stronger auditory evoked responses to incongruent stimuli than low predictive visual information, indicating that predictive visual information facilitates audiovisual mismatch detection. Our study suggests that this process is widely intact in SZP.

Theta oscillations reflect impaired crossmodal prediction error processing in schizophrenia

Frontal theta-band power between 230 ms and 370 ms differed significantly between SZP and HC. We observed stronger theta-band power in the high compared to low predictive trials in HC, whereas no such effect was found in SZP. Specifically, SZP lacked on an enhanced frontal PE processing in the high predictive condition. The effect in HC was source localized in the frontal cortex, encompassing MFC and ACC. This observation is interesting considering the results of previous functional magnetic resonance imaging studies, which focused on the anatomical structures underlying crossmodal PE processing in healthy participants (Noppeney, Josephs, Hocking, Price, & Friston, 2007; Noppeney, Ostwald, &

Werner, 2010). Noppeney et al. (2007) proposed that crossmodal PE processing in audiovisual speech is expressed at multiple hierarchical levels, including the STG at a first phonological level, the angular gyrus at a semantic level, and the medial prefrontal cortex reflecting a higher conceptual level. Moreover, frontal areas including ACC and MFC play an important role in top-down control of attention, which is presumably reflected in modulations of theta-band oscillations (Cavanagh & Frank, 2014). Frontal theta-band oscillations have been related to cognitive control induced by conflicting and erroneous information that required behavioral adaptation (Cohen & Gaal, 2013; Hanslmayr et al., 2008). Hence, the absence of enhanced frontal theta-band effects during crossmodal PE processing in our study suggests altered top-down processing, i.e. insufficient update and resolution of violated predictions, in the frontal cortex of SZP. We also observed a trend level negative relationship between the PE induced frontal theta-band enhancements and the SZ psychopathology, a finding that requires validation in a larger sample. Nevertheless our data indicate a link between frontal theta-band oscillation deficits and SZ psychopathology.

Numerous fMRI studies have found relationships between altered frontal cortex activity and impairments in executive functions in SZP (Goghari, Sponheim, & MacDonald, 2011). With emphasis on audiovisual speech processing visual lip movements not only provide information about the content, i.e. “what”, but also about the onset, i.e. “when”, of the upcoming auditory input (Arnal & Giraud, 2012). A recent study compared the McGurk fusion rate in SZP and HC at varying onset asynchronies between visual and auditory inputs (Martin et al., 2013). The study showed that SZP compared with HC have a larger temporal window in which asynchronous audiovisual speech stimuli are perceived as synchronous. This is in agreement with the finding of enhanced simultaneity perception rates in SZP, obtained in simultaneity judgement tasks using basic, i.e. semantically meaningless, unisensory (Giersch et al., 2009; Giersch, Lalanne, Assche, & Elliott, 2013; Lalanne, Assche, & Giersch, 2012) and multisensory stimuli. Thus, the lack of frontal theta-band power enhancement in the present study might reflect, at least in part, temporal processing deficits in SZP. The mere detection of audiovisual incongruence, as reflected in auditory cortex activity, seems to be much less affected. Across conditions frontal theta-band power was reduced in SZP. Recently, Popov et al. (Popov, Wienbruch, et al., 2015) found reduced frontal theta-band oscillations in SZP during the processing of color-word incongruence. Another study revealed diminished frontal theta-band power in SZP during the processing of visual motion and top-down perceptual switching (Mathes et al., 2015). This suggests that altered theta-band oscillations play an important role in top-down processing deficits in SZP. Interestingly, an MEG study in healthy participants revealed violations of crossmodal predictions that were expressed in stronger coupling of beta-band (14-15 Hz) and gamma-band (60-80 Hz) oscillations (Arnal et al., 2011). In contrast this study we obtained oscillatory responses in the EEG and did not find clear modulations in gamma-band oscillations. MEG compared to the EEG has a higher sensitivity in the measurement of high frequency oscillations (Muthukumaraswamy & Singh, 2013). Additionally, numerous studies have revealed reduced gamma-band oscillations in SZP (Gallinat et al., 2004; Leicht et al., 2010). These factors might have contributed to the absence of gamma-band modulations in the present study. It would be interesting to use MEG to examine gamma-band oscillations during crossmodal PE processing in SZP.

Summary and conclusion

Our study demonstrates a crossmodal PE processing deficit in SZP. In auditory areas we observed similar processing of audiovisual stimulus incongruence in SZP and HC. The increase of GFP in high compared to low predictive syllables suggests that the crossmodal prediction strength modulates audiovisual incongruence detection in the auditory cortex. The key novel finding is that SZP lack on a crossmodal PE processing related enhancement of frontal theta-band oscillations. The reduced frontal theta-band power, which was associated with the SZ psychopathology, presumably reflects a top-down crossmodal processing deficit. Aberrant frontal theta-band oscillations underlie dysfunctional crossmodal PE processing in SZ and might contribute to the recently proposed multisensory processing deficits in this patient group.

5 General discussion:

The aim of the present thesis was to elucidate how changes in neural oscillatory activity contribute to multisensory speech processing in different paradigms in healthy participants and in the sample of patients with schizophrenia. Beside the existing evidence of multisensory speech processing reviewed above the novel aspects of the present thesis were: a) to evaluate how slow oscillatory activity changes during the perception of a multisensory speech illusion, such as the McGurk effect, in healthy participants, b) to investigate illusory multisensory processing, using the same design, in patients with schizophrenia and examine the differences in neural oscillations, c) to explore the oscillatory mechanisms of crossmodal prediction error in speech processing in schizophrenia. Taken together, the results from all three studies of the present thesis emphasize the role of slow oscillations in multisensory speech processing. Moreover, slow oscillatory activity seems to be a crucial marker for altered multisensory speech processing in patients with schizophrenia.

5.1 The contribution of slow oscillatory activity to illusory and non-illusory audiovisual perception in healthy participants and patients with schizophrenia

Investigating a classic multisensory integration phenomenon in speech (i.e McGurk effect), the first study revealed modulations in early and late beta-band activity over left temporal and frontal areas during the processing of the McGurk illusion. This paradigm incorporated the presentation of congruent and incongruent audiovisual syllables, which might induce the McGurk illusion. Hence, the task required the resolution of detected incongruence and the formation of a coherent illusory perception during the McGurk illusion. Precisely, in healthy participants this suppression of late beta-band power fosters the formation of a coherent, subjectively congruent percept, namely the McGurk illusion. However, when focusing on the differences in oscillatory activity during the perception of the McGurk illusion in patients with schizophrenia and matched controls there is no differences in beta-band power. The absence of differences in the late beta-band power, possibly indicates that the process of perception formation itself is not altered in patients with schizophrenia. Although speculative, the intact late beta-band suppression in patients might serve as a compensatory mechanism during the illusory perception formation process and also contributes to the absence of behavioral differences in the illusion rate between both groups. However, beside the absence of beta-band power, the role alpha-band power as a differentiating marker between patients and matched controls has been evidenced in the second study. The modulations in the alpha-band power presumably mirror another process, which differentiates between patients and matched controls during the formation of the McGurk illusion. Given that low alpha-band power reflects a state of increased excitability which favors perceptual processing (Klimesch et al., 2007), our findings of enhanced late alpha-band suppression during the McGurk illusion in matched controls point towards a stronger processing during the illusion in matched controls. Based on the medio-central topography of the alpha-band effect auditory and frontal areas might be involved in this processing. In contrast, patients show less alpha-band suppression during the McGurk illusion at medio-central topography, which could indicate an auditory processing deficit. Notably, alpha-band power over visual processing areas was similar for both groups during the McGurk illusion. This indicates similar intact processing of visual syllables in both groups and fosters the assumption that the processing of the McGurk illusion in schizophrenia is specifically altered in auditory or frontal areas. In combining aspects of multisensory integration in speech and crossmodal prediction error processing the third study revealed a modulation of theta-band power during the processing of a crossmodal prediction error in speech. This paradigm required attention to occasional visual and auditory congruent or incongruent targets. In addition, visually induced high or low prediction errors were examined. Hence, the paradigm requires the formation of visually induced predictions, generation of crossmodal prediction errors and a top-down mediated update of violated predictions. Interestingly, similar to the alpha power findings in second study, differences in theta-band power between patients and matched controls were observed at frontal areas. Again, this suggests that frontal processes, such as top down processing during the resolution of the crossmodal prediction error are altered in patients with schizophrenia. Moreover, this finding provides further evidence for the notion that frontal theta-band power plays a crucial role in cognitive control due to conflicting information, which requires behavioral adaptation (Cavanagh & Frank, 2014; Cohen & Cavanagh, 2011). In conclusion the presented data hint towards a specific role of slow oscillatory activity in a variety of multisensory speech processing tasks involving illusory and non-illusory perception. Furthermore, evidence is added that points towards a specific deficit in slow oscillations during multisensory processing in schizophrenia beside the well acknowledged reductions of fast oscillatory activity during unisensory processing (Uhlhaas & Singer, 2010).

5.2 A multistage integration model of multisensory integration in speech

The underlying precise mechanisms and the temporal sequence of processes involved in multisensory integration in the human brain are still not completely understood. Various models of multisensory integration that explain how audiovisual speech is integrated have been proposed over the last decades (Peelle & Sommers, 2015). The traditional view proposed that multisensory processing is an automatic process, organized in a hierarchical manner, which finally results in the stage of stimulus integration. Models corroborating such a view assume that the bottom-up sensory input converges first in the primary sensory cortices and later into association areas such as STG/STS or frontal areas. These convergence models assume that audio and visual information are processed largely independent and combined later in a distinct late integration process. The main advantage of late integration models is that the integration process is represented as a separate cognitive stage, which can account for individual differences in audiovisual speech perception that cannot be explained by unisensory processing differences. However, these models of late integration in audiovisual speech have been challenged by models of early integration (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008). The basic assumption of early integration models is that sensory input from one modality can influence the processing of another modality and operates not strictly in parallel. As such early integration models in audiovisual speech perception are supported by electrophysiological evidence which showed that visual speech information can already affect online auditory processing (Wassenhove et al., 2004). Beside models of early and late multisensory integration, a third type of model exists, which is more flexible and includes integration at multiple stages. These models combine the views from both early and late integration models. Thus, in multistage models, visual speech cues can influence auditory perception, but integration can also occur at later stages. Overall neurophysiological evidence, as reflected in ERPs and oscillatory activity are relevant to answer the question of whether multisensory integration in audiovisual speech occurs at early, late or multiple stages. With reference to the first study a reduction of early auditory ERPs, reflecting the auditory N1, was found during the McGurk illusion compared to congruent audiovisual condition. This finding is in line with previous evidence (Besle et al., 2004; Stekelenburg & Vroomen, 2007) and indicates the early influence of visual context on auditory speech processing. Likewise in the second study differences in early auditory ERPs at the latency around 100ms were found between patients with schizophrenia and matched controls in the same design. The positive amplitude deflections at this latency were enhanced during the McGurk illusion compared to congruent audiovisual speech in matched controls. In patients, this pattern was reversed, indicating an altered early audiovisual processing. Interestingly, the third study similar effects of visually induced predictions on auditory processing in global field power (GFP) were observed in both groups. However, visual prediction strength affected the processing of incongruent auditory speech information in a differential manner. Precisely, the GFP around 200ms was increased during high compared to low predictive visual syllables, suggesting intact audiovisual incongruence processing in both groups. In total, the evidence from early ERPs and GFP over all three studies indicates that visual speech impacts auditory processing and either contributes to the formation of an illusory perception (Study 1 and 2) or induces the generation of a crossmodal prediction error in speech (Study 3) at an early level. Beyond this, also oscillatory activity was found to reflect different processing stages during multisensory speech processing in the presented studies. In the first study late beta-band suppression was pronounced during the perception of the McGurk illusion compared to congruent audiovisual syllables. Notably, this pattern was also found when comparing the perception of the McGurk illusion and incongruent audiovisual syllables that did not result in an illusory percept. The early beta-band suppression following incongruent audiovisual stimulation matches with a recent framework (Engel & Fries, 2010) that regards beta-band activity as coding of enhanced processing demands, which is especially required during the processing of the McGurk illusion. Moreover, late beta-band suppression observed in the first study presumably mirrors enhanced audiovisual integrative processing, especially involving the formation of an illusory coherent percept, such as the McGurk illusion. Interestingly, early and late beta-band power effects in the first study were not correlated and therefore substantiate the notion that the two effects reflect different stages in the audiovisual integration process. In a similar vein, the second study used the same design and again found differences in late oscillatory activity between patients and matched controls. However, in contrast to the first study the diminished oscillatory activity at fronto-central electrodes was found in the alpha band and potentially displays another process than the late beta-band effect. In particular, the less pronounced alpha-band suppression during the McGurk illusion in the patient group, presumably points towards differences in late auditory processing. However, the behavioral outcome of this late integration process, namely the McGurk illusion rate, was not affected and no differences were found between patients and matched controls. The absence of this behavioral difference might indicate compensatory mechanisms during multisensory speech integration in schizophrenia. As mentioned above, the intact late beta-band suppression in patients might serve as a compensatory mechanism during the illusory perception formation process. Supporting evidence from behavioral studies, using non-speech audiovisual stimuli (Gelder et al., 2002a; Stone et al., 2011) also showed a lack of behavioral differences in multisensory integration performance. As discussed by a recent study (Martin et al., 2013) the absence of differences in the illusion rate might also be caused by the use of a strict criterion for fusion reports, which considered only fusion percepts as McGurk illusion rather than visually or auditory dominated perception. Finally, the third study found differences in late theta-band power, localized in the frontal cortex between patients and matched controls. In particular, these differences in frontal theta power were pronounced during high compared to low predictive visual syllables in patients. This presumably indicates deficits in crossmodal prediction error processing and altered top-down processing in patients with schizophrenia. Although speculative, it can be hypothesized that the reduction in oscillatory power as shown in the late beta and theta-band serves as a late incongruence resolution mechanism that either results in the formation of a coherent percept or helps to update violated crossmodal predictions. Hence, the findings in neural oscillation in the first and the third study imply that multisensory integration can also occur at late and thereby represents a separate cognitive stage.

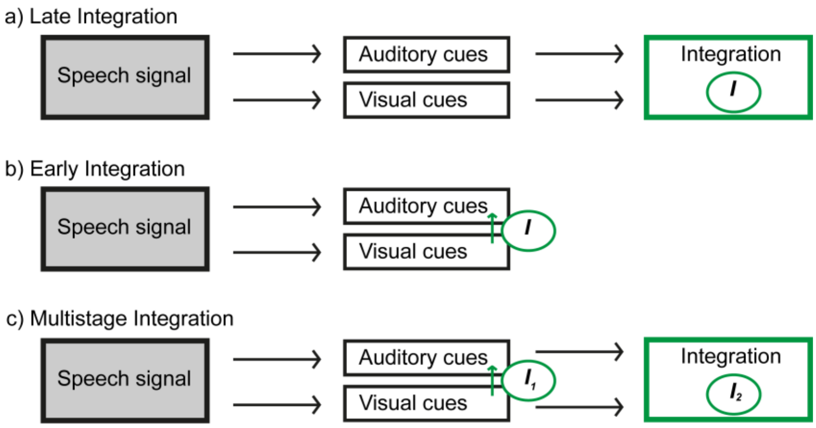

Figure 1 Overview of models from audiovisual speech integration.

a) The late integration view posits that audiovisual integration occurs at stage after modality-specific processing of visual and auditory. b) An early integration view holds that visual cues influence directly auditory processing at early processing stages and thereby integration happens concurrently with perception. c) Multistage models allow integration at multiple levels, as reflected by the green circles at early and late stages. Adapted figure from Peelle 2015.

5.3 Outlook and conclusion

The empirical evidence of slow oscillatory activity during multisensory speech processing presented in three studies may well be reconciled with the relevance of slow oscillations during long range connectivity (Palva & Palva, 2012). Certainly, future work is needed to elucidate the role of slow oscillatory activity for interareal communication, by means of connectivity measures (e.g. coherence, phase lag index) between distant cortical areas involved in multisensory speech processing. Related to this issue is the question of directed information flow during multisensory speech processing and how this information flow might be altered in patients with schizophrenia. Transcranial magnetic stimulation (TMS) is a non-invasive technique, which allows to investigate directed communication between brain areas by selectively inhibiting specific regions and hence disturbing information flow between brain areas (Bolognini & Maravita, 2011). TMS has already been successfully used to study multisensory phenomena, such as the McGurk illusion (Beauchamp et al., 2010) or the rubber hand illusion (Wold, Limanowski, Walter, & Blankenburg, 2014) in healthy participants. Previously neuro-anatomical evidence demonstrated that the STS is a key area for audio-visual speech integration (Beauchamp et al., 2004). In line with this finding, evidence from a fMRI guided TMS study found reduced susceptibility for McGurk illusion when left STS was stimulated compared to control region (Nath & Beauchamp, 2013). Consequently, TMS might be a suitable approach to modulate information flow, during multisensory speech processing in schizophrenia. Beyond temporally short lasting and artificially induced disruptions in information flow by TMS another valuable approach to modulate multisensory processing in schizophrenia are cognitive-perceptual trainings. The importance of cognitive and perceptual trainings in schizophrenia recently gained interest, but the effects of these trainings on functional measures as EEG are largely unexplored. First evidence from unisensory studies suggest that cognitive training affects auditory unisensory processing during a paired click-tone paradigm and results in oscillatory changes in alpha-band power (Popov, Rockstroh, Weisz, Elbert, & Miller, 2012; Popov et al., 2015). In the visual modality, patients with schizophrenia showed behavioral benefit in facial affect recognition associated with modulations in alpha-band power after cognitive and perceptual training (Popova et al., 2014). Based on these findings in unisensory processing further studies should explore to what extend training induced changes in oscillatory activity during multisensory speech processing could be achieved by specific multisensory speech trainings

However, these questions need to be substantiated by further data and might serve as inspiration for future projects.

References

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–8.

Arnal, L. H., Morillon, B., Kell, C. a, Giraud, A., (2009). Dual Neural Routing of Visual Facilitation in Speech Processing. *The Journal of Neuroscience*, *29*(43), 13445–53.

Arnal, L. H., Wyart, V., & Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, *14*(6), 797–801.

Bareither, I., Chaumon, M., Bernasconi, F., Villringer, A., & Busch, N. A. (2014). Invisible visual stimuli elicit increases in alpha-band power. *Journal of Neurophysiology*, *112*, 1082–1090.

Bastiaansen, M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research*, *159*(06), 179–96.

Bastiaansen, M., Magyari, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(7), 1333–47.

Baumgarten, T. J., Schnitzler, A., & Lange, J. (2014). Prestimulus Alpha Power Influences Tactile Temporal Perceptual Discrimination and Confidence in Decisions. *Cerebral Cortex*, 1–13.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809–23.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided Transcranial Magnetic Stimulation Reveals That the Superior Temporal Sulcus Is a Cortical Locus of the McGurk Effect. *The Journal of Neuroscience*, *30*(7), 2414–2417.

Besle, J., Fort, A., Delpuech, C., & Girad, A. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*(August), 2225–2234.

Boer-schellekens, L. De, Stekelenburg, J. J., Pieter, J., Gool, A. R. Van, & Vroomen, J. (2014). Acta Psychologica Sound improves diminished visual temporal sensitivity in schizophrenia. *ACTPSY*, *147*, 136–142.

Bolognini, N., & Maravita, A. (2011). Uncovering multisensory processing through non-invasive brain stimulation. *Frontiers in Psychology*, *2*(March), 1–10.

Boutros, N. N., Brockhaus-Dumke, A., Gjini, K., Vedeniapin, A., Elfakhani, M., Burroughs, S., & Keshavan, M. (2009). Sensory-gating deficit of the N100 mid-latency auditory evoked potential in medicated schizophrenia patients. *Schizophrenia Research*, *113*(2-3), 339–346.

Bowie, C. R., Leung, W. W., Reichenberg, A., Mcclure, M. M., Patterson, T. L., Heaton, R. K., & Harvey, P. D. (2009). Predicting Schizophrenia Patinets Real World Behavior with Specific Neurophysiological and Functional Capacity Measures. *Biological Psychiatry*, *63*(5), 505–511.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Brockhaus-Dumke, A., Mueller, R., Faigle, U., & Klosterkoetter, J. (2008). Sensory gating revisited: Relation between brain oscillations and auditory evoked potentials in schizophrenia. *Schizophrenia Research*, *99*, 238–249.

Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal Analysis of Multichannel EEG: CARTOOL. *Computational Intelligence and Neuroscience*, *2011*, 1–15.

Butler, P. D., & Javitt, D. C. (2007). Early-stage visual processing deficits in schizophrenia. *Current Opinion in Psychiatry*, *18*(2), 151–157.

Butler, P. D., Martinez, A., Foxe, J. J., Kim, D., Zemon, V., Silipo, G., … Javitt, D. C. (2007). Subcortical visual dysfunction in schizophrenia drives secondary cortical impairments. *Brain*, *130*, 417–430.

Buzsaki, G. (2006). *Rhythms of the Brain, Oxford University Press, New York.*

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

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421.

Cohen, M. X., & Cavanagh, J. F. (2011). Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict, *2*(February), 1–12.

Cohen, M. X., & Donner, T. H. (2013). Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. *Journal of Neurophysiology*, *110*, 2752–2763.

Cohen, M. X., & Gaal, S. Van. (2013). Dynamic Interactions between Large-Scale Brain Networks Predict Behavioral Adaptation after Perceptual Errors. *Cerebral Cortex*, *23*(May), 1061–1072.

Dang-vu, T. T., Schabus, M., Desseilles, M., Sterpenich, V., Vandewalle, G., Carrier, J., … Degueldre, C. (2008). Spontaneous neural activity during human. *PNAS*, *105*(39), 15160–15165.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolboxfor analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methodsethods*, *134*(1), 9–21.

Desantis, A., Mamassian, P., Lisi, M., & Waszak, F. (2014). The prediction of visual stimuli influences auditory loudness discrimination. *Experimental Brain Research*, *232*, 3317–3324.

Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B. A., & Javitt, D. C. (2002). Impaired Visual Object Recognition and Dorsal/Ventral Stream Interaction in Schizophrenia. *Arch Gen Psychiatry*, *59*(11), 1011–1020.

Driel, J. Van, Knapen, T., Es, D. M. Van, & Cohen, M. X. (2014). Interregional alpha-band synchrony supports temporal cross-modal integration. *NeuroImage*, *101*, 404–415.

Earls, H. A., Curran, T., & Mittal, V. (2015). Deficits in Early Stages of Face Processing in Schizophrenia: A Systematic Review of the P100 Component. *Schizophrenia Bulletin*, 1–9.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations--signalling the status quo? *Current Opinion in Neurobiology*, *20*(2), 156–65.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, *5*(1), 10749–10753.

Esser, S. K., Huber, R., Massimini, M., Peterson, M. J., Ferrarelli, F., & Tononi, G. (2006). A direct demonstration of cortical LTP in humans: A combined TMS / EEG study. *Brain Research Bullettin*, *69*, 86–94.

Fingelkurts, A. A., Fingelkurts, A. A., & Krause, C. M. (2003). Cortical operational synchrony during audio – visual speech integration. *Brain and Language*, *85*, 297–312.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews. Neuroscience*, *10*(January), 48–58.

Ford, J. M., & Mathalon, D. H. (2012). Anticipating the future: automatic prediction failures in schizophrenia. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *83*(2), 232–9.

Foucher, J. R., Lacambre, M., Pham, B., Giersch, A., & Elliott, M. A. (2007). Low time resolution in schizophrenia Lengthened windows of simultaneity for visual , auditory and bimodal stimuli. *Schizophrenia Research*, *97*, 118–127.

Freunberger, R., Klimesch, W., Griesmayr, B., Sauseng, P., & Gruber, W. (2008). Alpha phase coupling reflects object recognition. *NeuroImage*, *42*, 928–935.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*(10), 474–480.

Friston, K. (2002). Functional integration and inference in the brain. *Progress in Neurobiology*, *68*(October 2001), 113–143.

Gallinat, J., Winterer, G., Herrmann, C. S., & Senkowski, D. (2004). Reduced oscillatory gamma-band responses in unmedicated schizophrenic patients indicate impaired frontal network processing. *Clinical Neurophysiology*, *115*, 1863–1874.

Gardner, D. M., Pharm, D., Murphy, A. L., Donnell, H. O., Pharm, B. S., Centorrino, F., … Al, E. T. (2010). International Consensus Study of Antipsychotic Dosing. *American Journal of Psychiatry*, *167*(June), 686–693.

Gelder, B. De, Vroomen, J., Annen, L., Masthof, E., & Hodiamont, P. (2002a). Audio-visual integration in schizophrenia, *59*, 211–218.

Gelder, B. De, Vroomen, J., Annen, L., Masthof, E., & Hodiamont, P. (2002b). Audio-visual integration in schizophrenia. *Schizophrenia Research*, *59*, 211–218.

Gelder, B. De, Vroomen, J., Jong, S. J. De, Masthoff, E. D., Trompenaars, F. J., & Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics. *Schizophrenia Research*, *72*, 195–203.

Giersch, A., Lalanne, L., Assche, M. Van, & Elliott, M. A. (2013). On disturbed time continuity in schizophrenia: an elementary impairment in visual perception? *Frontiers in Psychology*, *4*(May), 1–10.

Giersch, A., Lalanne, L., Corves, C., Seubert, J., Shi, Z., Foucher, J., & Elliott, M. A. (2009). Extended Visual Simultaneity Thresholds in Patients With Schizophrenia. *Schizophrenia Bulletin*, *35*(4), 816–825.

Goghari, V., Sponheim, S., & MacDonald, W. (2011). The Functional Neuroanatomy of Symptom Dimensions in Schizophrenia: A Qualitative and Quantitative Review of a Persistent Question. *Neuroscience and Biobehavioral Reviews*, *34*(3).

Grant, K. W., & Seitz, P. F. (2000). The use of visual speech cues for imporving auditory detection of spoken sentences. *The Journal of the Acoustic Society of America*, *108*, 1197–208.

Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *PNAS*, *86*(5), 1689–1702.

Gross, J., Kujala, J., Hamaleinen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *PNAS*, *98*(2), 694–9.

Grützner, C., Wibral, M., Limin, S., Rivolta, D., Singer, W., Maurer, K., & Uhlhaas, P. (2013). Deficits in high-(>60 Hz) gamma-band oscillations during visual processing in schizophrenia. *Frontiers in Human Neuroscience*, *7*(88).

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

Hanslmayr, S., Pasto, B., Ba, K., Gruber, S., Wimber, M., & Klimesch, W. (2008). The Electrophysiological Dynamics of Interference during the Stroop Task. *Journal of Cognitive Neuroscience*, *20*(2), 215–225.

Hazlett, E. A., Goldstein, K. E., & Kolaitis, J. C. (2012). A Review of Structural MRI and Diffusion Tensor Imaging in Schizotypal Personality Disorder. *Current Psychiatry Reports*, *14*, 70–78.

Hipp, J. F., Engel, A. K., & Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron*, *69*(2), 387–96.

Horga, G., Schatz, K. C., Abi-dargham, A., & Peterson, B. S. (2014). Deficits in Predictive Coding Underlie Hallucinations in Schizophrenia. *The Journal of Neuroscience*, *34*(24), 8072–8082.

Howes, O. D., Kambeitz, J., Kim, E., Stahl, D., Slifstein, M., Abi-dargham, A., & Kapur, S. (2012). The Nature of Dopamine Dysfunction in Schizophrenia and What This Means for Treatment. *Arch Gen Psychiatry*, *69*(8), 776–786.

Javitt, D. C. (2009). Sensory Processing in Schizophrenia: Neither Simple nor Intact. *Schizophrenia Bulletin*, *35*(6), 1059–1064.

Jensen, O., Kaiser, J., & Lachaux, J. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neuroscience*, *30*(7).

Jong, J. J. De, Hodiamont, P. P. G., Stock, J. Van Den, & Gelder, B. De. (2009). Audiovisual emotion recognition in schizophrenia: Reduced integration of facial and vocal affect. *Schizophrenia Research*, *107*(2-3), 286–293.

Jung, J., Hudry, J., Ryvlin, P., Royet, J., Bertrand, O., & Lachaux, J. (2006). Functional Significance of Olfactory- induced Oscillations in the Human Amygdala. *Cerebral Cortex*, *16*(January), 1–8.

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin*, *13*(2), 261–276.

Keefe, R. S. E., Goldberg, T. E., Harvey, P. D., Gold, J. M., Poe, M. P., & Coughenour, L. (2004). The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophrenia Research*, *68*, 283–297.

Keil, J., Müller, N., Ihssen, N., & Weisz, N. (2012). On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states. *Cerebral Cortex*, *22*(January), 221–231.

Kissler, J., & Herbert, C. (2013). Emotion, Etmnooi, or Emitoon? – Faster lexical access to emotional than to neutral words during reading. *Biological Psychology*, *92*(3), 464–479.

Kissler, J., & Koessler, S. (2011). Emotionally positive stimuli facilitate lexical decisions — An ERP study. *Biological Psychology*, *86*(3), 254–264.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617.

Klimesch, W., Fellinger, R., & Freunberger, R. (2011). Alpha oscillations and early stages of visual encoding. *Frontiers in Psychology*, *2*(May), 1–11.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition – timing hypothesis. *Brain Research Reviews*, *53*, 63–88.

Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, *18*, 65–75.

Knebel, J. F., Javitt, D. C., & Murray, M. M. (2012). Impaired early visual response modulation to spatial information in chronic schizophrenia. *Psychiatry Research*, *193*(3), 168–176.

Knowland, V. C. P., Mercure, E., Karmiloff-smith, A., Dick, F., & Thomas, M. S. C. (2014). Audio-visual speech perception: a developmental ERP investigation. *Developmental Science*, *1*, 110–124.

Koh, Y., Soon, K., Sic, J., Choi, J., Kang, D., Hwan, J., … Soo, J. (2011). An MEG study of alpha modulation in patients with schizophrenia and in subjects at high risk of developing psychosis. *Schizophrenia Research*, *126*(1-3), 36–42.

Lalanne, L., Assche, M. Van, & Giersch, A. (2012). When Predictive Mechanisms Go Wrong: Disordered Visual Synchrony Thresholds in Schizophrenia.*Neuropsychologia*, *38*(3), 506–513.

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

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

Leavitt, V. M., Molholm, S., Ritter, W., Shpaner, M., & Foxe, J. J. (2007). Auditory processing in schizophrenia during the middle latency period (10 – 50 ms: high-density electrical mapping and source analysis reveal subcortical antecedents to early cortical deficits. *Journal of Psychiatry and Neuroscience*, *32*(5), 339–353.

Lebib, R., Papo, D., Bode, S. De, & Baudonnie, P. (2003). Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neuroscience Letters*, *341*, 185–188.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation*, *11*(2), 417–41.

Leicht, G., Kirsch, V., Giegling, I., Karch, S., Hantschk, I., Möller, H., … Mulert, C. (2010). Reduced Early Auditory Evoked Gamma-Band Response in Patients with Schizophrenia. *Biological Psychiatry*, *67*(3), 224–231.

Leske, S., Tse, A., Oosterhof, N. N., Hartmann, T., Müller, N., Keil, J., & Weisz, N. (2014). The strength of alpha and beta oscillations parametrically scale with the strength of an illusory auditory percept. *NeuroImage*, *88*, 69–78.

Lopes da Silva, F. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalography and Clinical Neurophysiology*, *79*, 81–93.

Luo, Y., Zhang, Y., Feng, X., & Zhou, X. (2010). Electroencephalogram oscillations differentiate semantic and prosodic processes during sentence reading. *Neuroscience*, *169*(2), 654–64.

Lynn, S. K., & Salisburg, D. F. (2009). Attenuated Modulation of the N170 ERP by Facial Expressions in Schizophrenia. *Clin EEG Neurosci*, *39*(2), 108–111.

Magnée, M. J. C. M., Oranje, B., Engeland, H. Van, Kahn, R. S., & Kemner, C. (2009). Neuropsychologia Cross-sensory gating in schizophrenia and autism spectrum disorder: EEG evidence for impaired brain connectivity? *Neuropsychologia*, *47*, 1728–1732.

Magnotti, J. F., & Beauchamp, M. S. (2014). The noisy encoding of disparity model of the McGurk effect. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-014-0722-2

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–90.

Martin, B., Giersch, A., Huron, C., & Wassenhove, V. Van. (2013). Temporal event structure and timing in schizophrenia: Preserved binding in a longer "'now'". *Neuropsychologia*, *51*(2), 358–371.

Mathes, B., Schmiedt-fehr, C., Kedilaya, S., Strüber, D., Brand, A., & Basar-Eroglu, C. (2015). Theta response in schizophrenia is indifferent to perceptual illusion. *Clinical Neurophysiology*. doi:10.1016/j.clinph.2015.02.061

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

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early Cross-Modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion. *The Journal of Neuroscience*, *27*(15), 4120–4131.

Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain ResearchBrain Res*, *13*, 417–425.

Möttönen, R., Schürmann, M., & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. *Neuroscience Letters*, *363*, 112–115.

Müller, N., Keil, J., Obleser, J., Schulz, H., Grunwald, T., Bernays, R.-L., … Weisz, N. (2013). You can’t stop the music: reduced auditory alpha power and coupling between auditory and memory regions facilitate the illusory perception of music during noise. *NeuroImage*, *79*, 383–393.

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, A. D., & Honey, G. (2008). Substantia nigra / ventral tegmental reward prediction error disruption in psychosis. *Molecular Psychiarty*, *13*(3), 1–18.

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP Analyses: A Step-by-Step Tutorial Review. *Brain Topographie*, *20*, 249–264.

Muthukumaraswamy, S. D., & Singh, K. D. (2013). Visual gamma oscillations: The effects of stimulus type, visual field coverage and stimulus motion on MEG and EEG recordings. *NeuroImage*, *69*, 223–230.

Myslobodsky, M., Golderberg, T., Johnson, F., Hicks, L., & Weinberger, D. (1992). Lipreading in Patients with Schizophrenia. *The Journal of Nervous and Mental Disease*, *180*(3), 168–171.

Nasrallah, H., Tandon, R., & Keshavan, M. (2015). Beyond the facts in schizophrenia: closing the gaps in diagnosis , pathophysiology , and treatment. *Epidemiology and Psychiatric Science*, *20*(4), 317–327.

Nath, A. R., & Beauchamp, M. S. (2013). A Neural Basis for Interindividual Differences in the McGurk Effect, a Multisensory Speech Illusion. *Neuroimage*, *59*(1), 781–787.

Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2007). The Effect of Prior Visual Information on Recognition of Speech and Sounds. *Cerebral Cortex*, *18*, 598–609.

Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual Decisions Formed by Accumulation of Audiovisual Evidence in Prefrontal Cortex. *The Journal of Neuroscience*, *30*(21), 7434–7446.

Obleser, J., Lahiri, A., & Eulitz, C. (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *NeuroImage*, *20*(3), 1839–1847.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869.

Palva, S., & Palva, J. M. (2012a). Discovering oscillatory interaction networks with M / EEG: challenges and breakthroughs.*Trends in Cognitive Sciences*, *16*(4), 219–230.

Palva, S., & Palva, J. M. (2012b). Discovering oscillatory interaction networks with M/EEG: challenges and breakthroughs. *Trends in Cognitive Sciences*, *16*(4), 219–230.

Pearl, D., Yodashkin-Porat, D., Katz, N., Valevski, A., Aizenberg, D., Sigler, M., … Kikinzon, L. (2009). Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups. *Comprehensive Psychiatry*, *50*(2), 186–92.

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *CORTEX*, 1–13. doi:10.1016/j.cortex.2015.03.006

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

Pfurtscheller, G., Zalaudek, K., & Neuper, C. (1998). Event-related beta synchronization after wrist , finger and thumb movement. *Electroencephalography and Clinical Neurophysiology*, *109*, 154–160.

Phillips, J. M., Vinck, M., Everling, S., & Womelsdorf, T. (2014). A Long-Range Fronto-Parietal 5- to 10-Hz Network Predicts “Top-Down” Controlled Guidance in a Task-Switch Paradigm. *Cerebral Cortex*, *24*(August), 1996–2008.

Piantoni, G., Werf, Y. D. Van Der, Jensen, O., & Someren, E. J. W. Van. (2015). Memory Traces of Long-Range Coordinated Oscillations in the Sleeping Human Brain. *Human Brain Mapping*, *84*(August 2014), 67–84.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., … Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*, 127–152.

Popov, T., Carolus, A., Schubring, D., Popova, P., Miller, G. A., & Rockstroh, B. S. (2015). Targeted training modifies oscillatory brain activity in schizophrenia patients. *NeuroImage:Clinical*, *7*, 807–814.

Popov, T., Jordanov, T., Weisz, N., Elbert, T., Rockstroh, B., & Miller, G. A. (2011). Evoked and induced oscillatory activity contributes to abnormal auditory sensory gating in schizophrenia. *NeuroImage*, *56*(1), 307–314.

Popov, T., Rockstroh, B., Popova, P., Carolus, A., & Miller, G. (2014). Dynamics of alpha oscillations elucidate facial affect recognition in schizophrenia. *Cognitive Affective Behavioral Neuroscience*, *14*, 364–377.

Popov, T., Rockstroh, B., Weisz, N., Elbert, T., & Miller, G. A. (2012). Adjusting Brain Dynamics in Schizophrenia by Means of Perceptual and Cognitive Training. *PloS One*, *7*(7), 1–9.

Popov, T., Wienbruch, C., Meissner, S., Miller, G. A., & Rockstroh, B. (2015). A mechanism of deficient interregional neural communication in schizophrenia. *Psychophysiology*, *52*(2), 648–656.

Popova, P., Popov, T. G., Wienbruch, C., Carolus, A. M., Miller, G. A., & Rockstroh, B. S. (2014). Changing facial affect recognition in schizophrenia: Effects of training on brain dynamics. *NeuroImage:Clinical*, *6*, 156–165.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Roa Romero, Y., Senkowski, D., & Keil, J. (2015). Early and Late Beta Band Power reflect Audiovisual Perception in the McGurk Illusion. *Journal of Neurophysiology*, *13*(7), 2342–2350.

Rosburg, T., Boutros, N. N., & Ford, J. M. (2008). Reduced auditory evoked potential component N100 in schizophrenia — A critical review. *Psychiatry Research*, *161*(3), 259–74.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Molholm, S., Javitt, D. C., & Foxe, J. J. (2007). Impaired multisensory processing in schizophrenia: Deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophrenia Research*, *97*, 173–183.

Saha, S., Chant, D., Welham, J., & Mcgrath, J. (2005). A Systematic Review of the Prevalence of Schizophrenia. *Plos Medicine*, *2*(5), 413–433.

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*(3), 587–597.

Sauseng, P., Griesmayr, B., Freunberger, R., & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1015–1022.

Schepers, I. M., Schneider, T. R., Hipp, J. F., Engel, A. K., & Senkowski, D. (2013). NeuroImage Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing. *NeuroImage*, *70*, 101–112.

Schepers, I. M., Yoshor, D., & Beauchamp, M. S. (2014). Electrocorticography Reveals Enhanced Visual Cortex Responses to Visual Speech. *Cerebral Cortex*.

Schneider, T. R., Lorenz, S., Senkowski, D., & Engel, A. K. (2011). Gamma-Band Activity as a Signature for Cross-Modal Priming of Auditory Object Recognition by Active Haptic Exploration, *31*(7), 2502–2510.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neural oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*(3), 106–113.

Schulz, E., May, E. S., Postorino, M., Tiemann, L., Nickel, M. M., Witkovsky, V., … Ploner, M. (2015). Prefrontal Gamma Oscillations Encode Tonic Pain in Humans. *Cerebral Cortex*, 1–8.

Schurger, A., Cowey, A., Cohen, J. D., Treisman, A., & Tallon-Baudry, C. (2008). Distinct and independent correlates of attention and awareness in a hemianopic patient. *Neuropsychologia*, *46*(8), 2189–2197.

Sehatpour, P., Molholm, S., Javitt, D. C., & Foxe, J. J. (2006). Spatiotemporal dynamics of human object recognition processing: An integrated high-density electrical mapping and functional imaging study of "'closure'" processes. *NeuroImage*, *29*, 605–618.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, *47*, 277–287.

Senkowski, D., & Gallinat, J. (2015). Dysfunctional Prefrontal Gamma-Band Oscillations Reflect Working Memory and Other Cognitive Deficits in Schizophrenia. *Biological Psychiatry*, *77*(12), 1010–1019.

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

Spencer, K. M., Niznikiewicz, M. A., Shenton, M. E., & Mccarley, R. W. (2008). Sensory-Evoked Gamma Oscillations in Chronic Schizophrenia. *Biological Psychiatry*, *63*(8), 744–747.

Stein, B., Burr, D., Constantinidis, C., Laurienti, P. J., Meredith, M. A., Jr, T. J. P., … Roeder, B. (2010). Semantic confusion regarding the development of multisensory integration: a practical solution. *European Journal of Neuroscience*, *31*, 1713–1720.

Stein, B. E., Meredith, M. A., & Wallace, M. T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Progress in Brain Research*, *95*, 79–90.

Stekelenburg, J. J., Maes, J. P., Van Gool, A. R., Sitskoorn, M., & Vroomen, J. (2013). Deficient multisensory integration in schizophrenia: An event-related potential study. *Schizophrenia Research*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23707640

Stekelenburg, J. J., Pieter, J., Gool, A. R. Van, Sitskoorn, M., Vroomen, J., Maes, J. P., … Vroomen, J. (2013). Deficient multisensory integration in schizophrenia: An event-related potential study. *Schizophrenia Research*, *147*(2-3), 253–261.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural Correlates of Multisensory Integration of Ecologically Valid Audiovisual Events. *Journal of Cognitive Neuroscience*, *19*, 1964–1973.

Stephan, K. E., Baldeweg, T., & Friston, K. J. (2006). Synaptic Plasticity and Dysconnection in Schizophrenia. *Biological Psychiatry*, *59*(10), 929–939.

Stephan, K. E., Friston, K. J., & Frith, C. D. (2009). Dysconnection in Schizophrenia□ : From Abnormal Synaptic Plasticity to Failures of Self-monitoring. *Schizophrenia Bulletin*, *35*(3), 509–527.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–1223.

Stone, D. B., Coffman, B. A., Bustillo, J. R., Aine, C. J., Stephen, J. M., & Moran, R. J. (2014). Multisensory stimuli elicit altered oscillatory brain responses at gamma frequencies in patients with schizophrenia. *Frontiers in Human Neuroscience*, *8*(November), 1–14.

Stone, D. B., Urrea, L. J., Aine, C. J., Bustillo, J. R., Clark, V. P., & Stephen, J. M. (2011). Unisensory processing and multisensory integration in schizophrenia: a high-density electrical mapping study. *Neuropsychologia*, *49*(12), 3178–3187.

Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, *8*(May), 1–7.

Surguladze, S. A., Calvert, G. A., Brammer, M. J., Bullmore, E. T., Giampietro, V., & David, A. S. (2001). Audio-visual speech perception in schizophrenia: an fMRI study. *Psychiatry Research*, *106*, 1–14.

Szycik, G. R., Münte, T. F., Dillo, W., Mohammadi, B., & Samii, A. (2009). Audiovisual integration of speech is disturbed in schizophrenia: An fMRI study. *Schizophrenia Research*, *110*(1-3), 111–118.

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

Tan, H. M., Lana, L., Uhlhaas, P. J., Spencer, K. M., & Boston, V. A. (2013). High-frequency neural oscillations and visual processing deficits in schizophrenia. *Frontiers in Psychology*, *4*(October), 1–19.

Taylor, S. F., & Tso, I. F. (2014). GABA abnormalities in schizophrenia: A methodological review of in vivo studies. *Schizophrenia Research*. doi:10.1016/j.schres.2014.10.011

Tseng, H.-H., Bossong, M. G., Modinos, G., Chen, K.-M., McGuire, P., & Allen, P. (2015, May). A systematic review of multisensory cognitive-affective integration in schizophrenia. *Neuroscience and Biobehavioral Reviews*. doi:10.1016/j.neubiorev.2015.04.019

Turetsky, B. I., Bilker, W. B., Siegel, S. J., Kohler, C. G., & Gur, R. E. (2009). Profile of auditory information-processing deficits in schizophrenia. *Psychiatry Research*, *165*(1-2), 27–37.

Uhlhaas, P. J., Phillips, W. A., Mitchell, G., & Silverstein, S. M. (2006). Perceptual grouping in disorganized schizophrenia B, *145*, 105–117.

Uhlhaas, P. J., & Singer, W. (2006). Neural Synchrony in Brain Disorders: Relevance for Cognitive Dysfunctions and Pathophysiology. *Neuron*, *52*, 155–168.

Uhlhaas, P. J., & Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nature Reviews. Neuroscience*, *11*(2), 100–113.

Uhlhaas, P. J., & Singer, W. (2015). Oscillations and Neuronal Dynamics in Schizophrenia: The Search for Basic Symptoms and Translational Opportunities. *Biological Psychiatry*, *77*(12), 1001–1009.

Van Atteveldt, N., Murray, M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron*, *81*(6), 1240–1253.

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

Wallwork, R. S., Fortgang, R., Hashimoto, R., Weinberger, D. R., & Dickinson, D. (2012). Searching for a consensus five-factor model of the Positive and Negative Syndrome Scale for schizophrenia. *Schizophrenia Research*, *137*(1-3), 246–250.

Wassenhove, V. Van, Grant, K. W., & Poeppel, D. (2004). Visual speech speeds up the neural processing of auditory speech Neurophysiological Basis of Multisensory Integration. *PNAS*, *102*, 1181–1186.

Weiss, S., & Mueller, H. M. (2012). “Too Many betas do not Spoil the Broth”: The Role of Beta Brain Oscillations in Language Processing. *Frontiers in Psychology*, *3*(June), 201. doi:10.3389/fpsyg.2012.00201

Weisz, N., Hartmann, T., Müller, N., Lorenz, I., & Obleser, J. (2011). Alpha rhythms in audition: cognitive and clinical perspectives. *Frontiers in Psychology*, *2*(April), 73.

White, T. P., Wigton, R. L., Joyce, D. W., Bobin, T., Ferragamo, C., Wasim, N., … Shergill, S. S. (2014). Eluding the illusion? Schizophrenia, dopamine and the McGurk effect. *Frontiers in Human Neuroscience*, *8*(August), 1–12.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–84.

Williams, L. E., Light, G. A., Braff, D. L., & Ramachandran, V. S. (2010). Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia*, *48*(10), 3128–3136.

Wold, A., Limanowski, J., Walter, H., & Blankenburg, F. (2014). Proprioceptive drift in the rubber hand illusion is intensified following 1 HzTMS of the left EBA. *Frontiers in Human Neuroscience*, *8*(June), 1–6.

Supplementary material

Study 2

Supplementary Methods

Participants included in the analysis of all McGurk trials

Since all McGurk trials, irrespective of perception, were used for the analyses only two participants per group were excluded due to an insufficient number of EEG trials after artifact rejection. We computed the same analyses of ERPs and oscillatory power as described in the main article. The analyses included 19 patients (7 female, mean age = 34.58 y) and 19 controls (5 female, mean age = 34.37 y). Supplementary Table 1 provides an overview on the demographic data, cognitive performance, and clinical scores.

Supplementary Results

Comparison of McGurk illusion rates between 19 patients and 19 controls that were included in the EEG data analysis of all McGurk trials

The illusion rates between ScZ patients (66.33 %) and matched controls (53.14 %;) did not significantly differ (t (36) = 1.191, p = 0.242).

Analysis of EEG data to all McGurk trials

The running 2 x 2 ANOVA for the ERPs in response to all McGurk trials (i.e. independent of the percept) revealed a significant main effect of Condition between 246 and 266 ms (F(2,36) = 8.12, p = 0.007) and between 338 and 368 ms (F(2,36) = 5.57, p = 0.023). The 246 to 266 ms effect of Condition indicates larger positive amplitudes in congruent trials compared with McGurk trials. The latter effect of Condition showed the reversed pattern: larger positive amplitudes in McGurk trials compared with congruent trials (Supplementary Figure 1). No significant interactions or main effects of Group were found. The running 2 x 2 ANOVA for oscillatory responses revealed a main effect of Condition in the theta-band (4 Hz) between 50 and 350 ms (F(2,36) = 8.04, p = 0.001). In both groups, theta-band power was larger in congruent compared with McGurk trials, suggesting that incongruent visual information modulates early audiovisual processing (Supplementary Material and Supplementary Figures S2 and S3). No significant interactions or main effects of Group were observed.

Supplementary Discussion

Outcome of the analysis of all McGurk trials

For congruent control and McGurk trials we found no significant interactions and no main effects of Group in ERPs, which survived the correction for multiple comparison (Figure S1). However, we found two significant main effects of Condition between 246 and 266 ms and 338 and 368 ms. The main effect in the 246 to 266 ms interval indicated larger positive amplitudes in congruent trials compared with McGurk trials. The latter effect in the 338 to 368 ms interval had the reversed pattern, indicated larger positive amplitudes in McGurk trials compared with congruent trials. Similarly to our late ERP effect, Arnal et al. (2011) found increased amplitudes of event-related fields during the presentation of incongruent compared to congruent audiovisual syllables. This suggests a longer latency processing of audiovisual stimulus incongruence.

Analogous to the analysis of oscillatory responses in McGurk illusion trials, we found no significant interactions and no main effects of Group. However, we found a significant main effect of Condition in the theta-band (4 Hz) power between 50 and 350 ms (see Figure S2 & S3). This effect was due to an enhanced theta-band power for congruent compared with McGurk trials, which was similarly observed in both groups. Interestingly, the theta-band effect occurred in a similar latency and in a similar frequency as a previously reported by Lange et al. (2013). Lange and colleagues investigated the processing of congruent and incongruent audiovisual speech vowels. In this study, theta-band power between 50 and 250 ms was increased during the presentation of incongruent compared to congruent audiovisual speech. Albeit the direction of the theta-band effects differed between the present and the Lange et al. (2013) study, the effects in early theta-band power suggests that incongruent visual information modulates early audiovisual processing.

Supplementary Figure S1

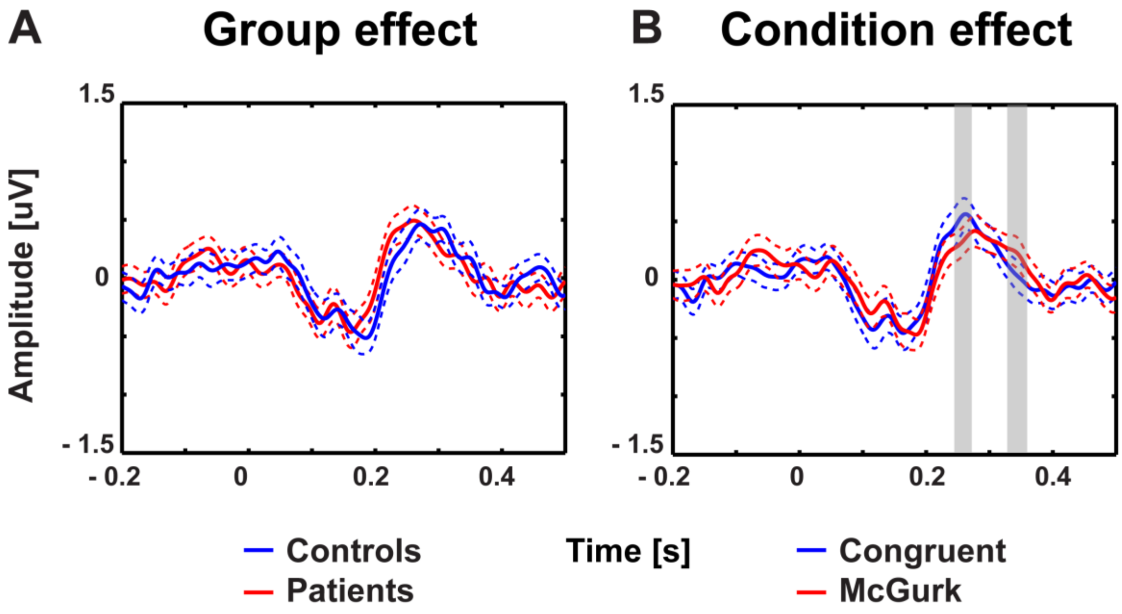

Figure S1. A) Traces of medio-central event-related potentials for all ScZ patient (red line), and matched controls (blue line). Dashed lines reflect traces with standard errors of mean. There were no differences between groups and no interaction. B) Traces of medio-central event-related potentials for all McGurk trials (red line), irrespective of perception, and congruent control trials (blue line). Dashed lines reflect traces with standard errors of mean. The analysis revealed main effects of condition at 246 to 266 ms and 338 to 368 ms marked with gray traces.

Supplementary Figure S2

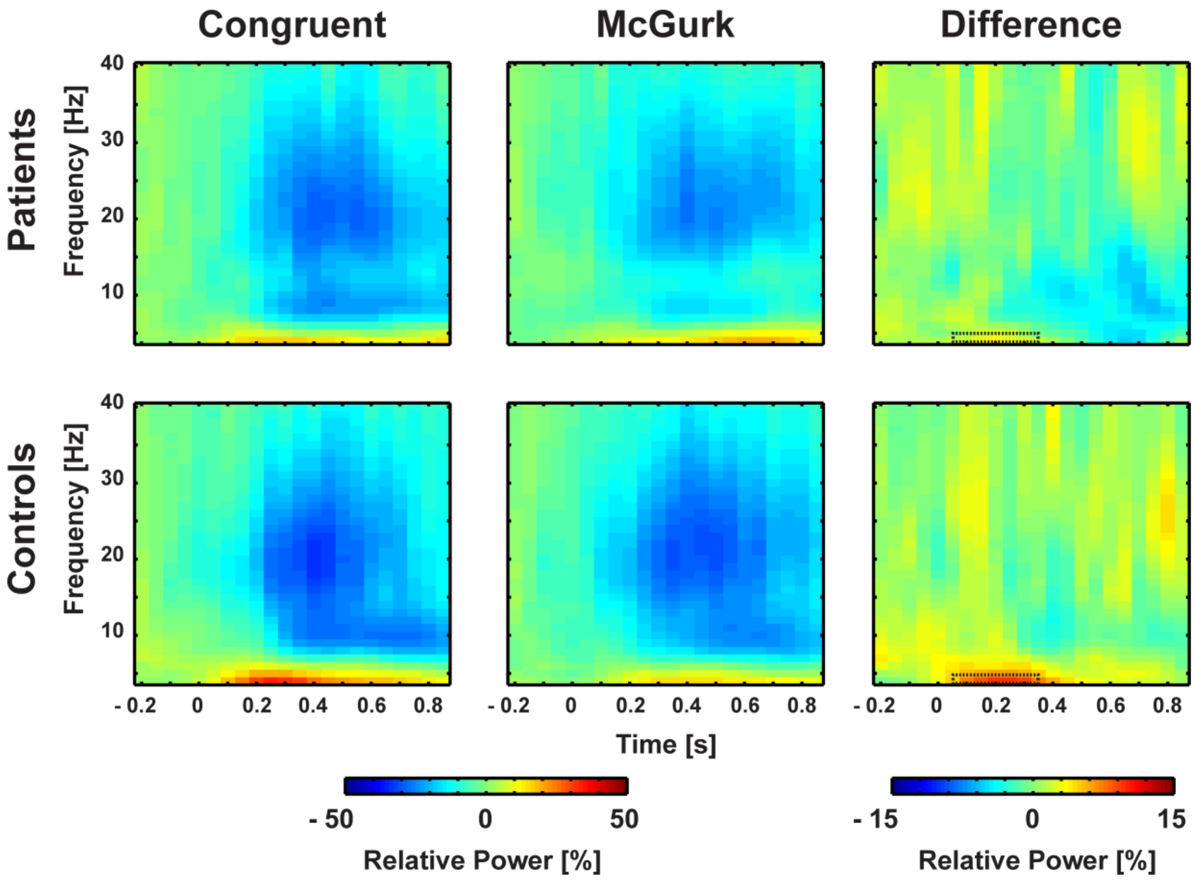

Figure S2. Time-frequency planes of oscillatory responses at medio-central electrodes. The dashed squares highlight the significant theta-band condition effect at 50 to 350 ms. Time zero denotes the onset of the auditory syllable.

Supplementary Figure S3

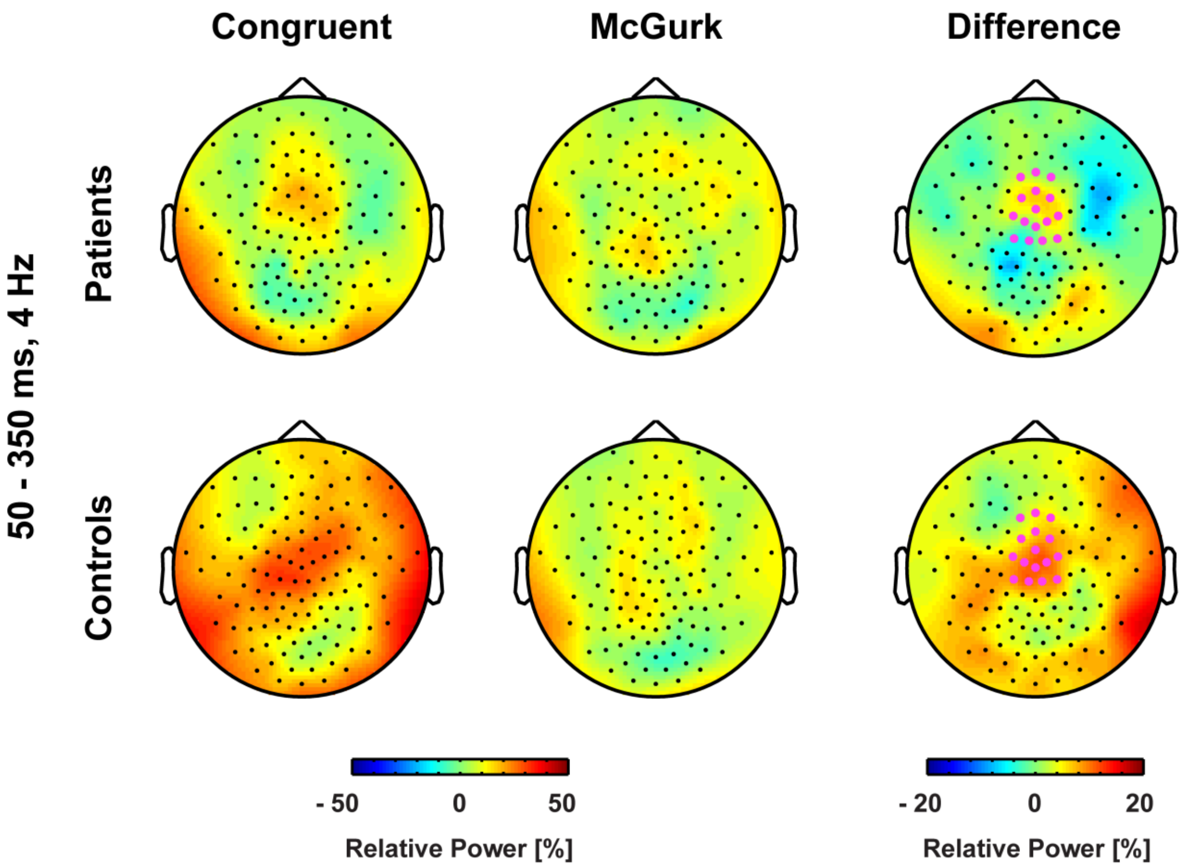

Figure S3. Topographies of medio-central theta-band power in the 550 to 700 ms interval. The bold magenta dots in the right panel denote the medio-central electrode group.

Supplementary Figure S4

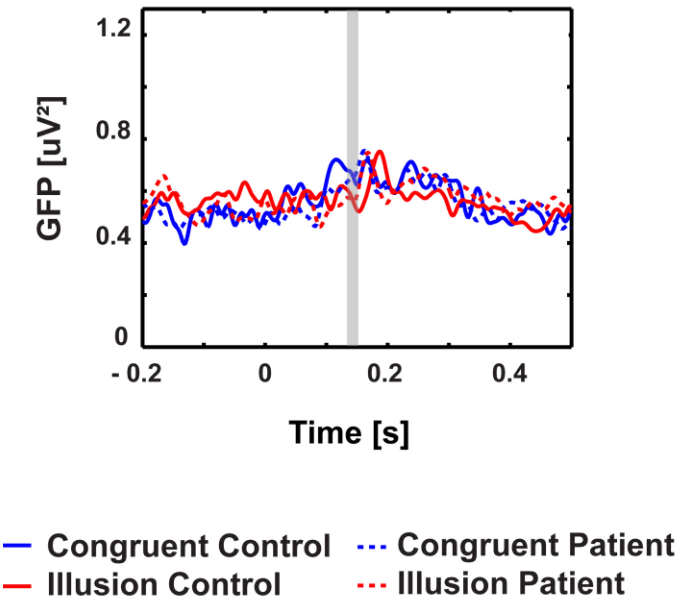

Figure S4. Global field power traces for patients (dashed lines) and matched controls (solid lines) for McGurk illusion trials (red lines), and congruent control trials (blue lines). Significant condition effect was found after 140-160 ms. Significant time interval are marked by gray shading. There were no further differences between groups and no interaction between these factors.

Supplementary Table 1

	Patients		Controls		Statistics	
	Mean	**SD**	**Mean**	**SD**	**t-values**	**p-values**
Age (years)	33.63	6.77	34.37	7.46	-.319	.752
Education (years)	10.95	1.58	10.95	1.72	0	1.00
Illness duration (years)	9.84	4.84	-	-	-	-
Chlorpromazin Eq. (daily dosage/ mg)	381.71	184.62	-	-	-	-
	N		**N**			
Gender (m/f)	12/7		14/5		-	-
Handedness (r/l)	15/4		16/3		-	-
Antipsychotic Med.	19		-		-	-
Co-medication*	5		-		-	-
BACS						
Verbal Memory	43.74	13.24	46.89	9.95	-.831	.411
Digit	19.74	4.28	20.11	4.41	-.261	.795
Motor	67.79	11.58	75.37	10.03	-2.52	.038
Fluency	48.26	14.50	52.90	14.66	-.979	.334
Symbol coding	56.68	13.45	57.84	14.42	-.256	.799
ToL	17.89	2.60	17.58	2.27	.399	.692
Total score	254.10	42.31	270.68	37.95	-1.27	.212
PANSS						
Negative	18.63	3.30	-	-	-	-
Positive	16.68	2.79	-	-	-	-
General	38.11	3.33	-	-	-	-
Total score	73.42	6.96	-	-	-	-

Table 1. Overview of demographic data. *Co-medication of antipsychotics and mood stabilizers

Supplementary Table 2

Syllable	**/Pa/**	**/Ka/**	**/Ga/**
Place of articulation	bilabilal	velar	velar
Visual duration	858 ms	1221 ms	891 ms
Auditory duration	329 ms	429 ms	348 ms
Latency difference between visual motion onset and auditory onset	337 ms	594 ms	429 ms

Table 2. Overview of the stimulus durations of the presented syllables and differences between visual motion onset and auditory syllable onset. Note that the latency difference between visual motion onset and auditory onset refers to congruent audiovisual syllable combinations (e.g., visual /Pa/ and auditory /Pa/).

Study 3

Stimuli

Video sequences of the syllables /Pa/, /La/, /Ta/, /Ga/, and /Fa/ were taken in frontal view displaying the face (visual angle = 5.95° x7.36°), in front of a black background. The clips of the syllables were converted to greyscale and equalized in their luminance using the SHINE toolbox (Willenbockel et al., 2010). To minimize eye movements, a small white fixation cross above the mouth at the philtrum was added to all clips (Figure 1). Visual stimuli were presented on a 21-inch CRT screen at a distance of 120cm and auditory stimuli were presented via a single, centrally positioned speaker (Bose Companion® 2). Syllables were presented with the real audiovisual onset asynchrony at 65dB (SPL) (Supplementary Table 1).

Selection of high and low predictive syllables

We conducted two pilot studies in healthy participants that did not participate in the main experiment to select high and low predictive visual syllables. In the first pilot study (n = 8, 4 female, mean age = 27.63 y) we presented 45 mute visual only trials in which a female speaker that was displayed on the screen voiced the syllable /Ba/, /Pa/, /Da/, /Ta/, /Ga/, /Ka/, /La/, /Ma/, or /Fa/. Importantly these syllables differed in terms of their place of articulation (bilabial, alveolar, velar) and manner of articulation (stop, fricative, nasal and liquid). Participants were instructed to indicate by button press on a labeled keyboard which syllable they perceived. A repeated measure ANOVA in which all nine syllables were entered revealed significant differences in recognition rates ($F(8,56) = 8.28$, $p < 0.001$). The two syllables with the highest recognition rates were /Pa/ and /La/. These syllables are generated at bilabial (/Pa/) and alveolar (/La/) position of the vocal tract, making them highly predictive. The two syllables with the lowest recognition rates were /Ta/ and /Ga/, which both are generated at alveolar (/Ta/) and velar (/Ga/) position of the vocal tract, making them low predictive. Follow up t-test revealed that recognition rates for the highly predictive /Pa/ and

/La/ syllables (average = 60 %) were much higher than the recognition rates for the less predictive /Ta and /Ga/ syllables (average = 31.2 %; (t(7) = 3.54, p = .01)). We selected the four syllables for further testing in the second pilot study (n = 6, 5 female, mean age = 27 y). In this second pilot study we examined how the visual predictability influenced subjective perception of congruent and incongruent audiovisual syllables. We presented 120 trials with the high (/La/ and /Pa/) and low (/Ta/ and /Ga/) predictive audiovisual syllables and instructed participants to rate their congruency perception on a 4 point scale by a button press on the keyboard (1 = congruent, 2 = little congruent, 3 = little incongruent, 4 = incongruent). Audiovisual syllables were presented in congruent and incongruent pairs. Importantly, we excluded all incongruent combinations that could evoke the McGurk illusion (Roa Romero et al. 2015). The repeated measure ANOVA for congruency ratings with the within-subject factors Predictability (high vs. low) and Congruency (congruent vs. incongruent) revealed a significant Predictability x Congruency interaction (F(1,5) = 220.6 , p < 0.001). Follow-up t-test, which were conducted separately for congruent and incongruent audiovisual syllables revealed that the incongruence perception in incongruent trials was much stronger in high compared with low predictive syllables (t(5) = 6.48, p = .001). No such effect was observed for the congruent audiovisual trials. Hence, this study shows that the high predictive syllables induce a stronger incongruence perception in incongruent audiovisual trials than the low predictive syllables.

Supplements

A) Curriculum vitae

B) Publications

C) Statement of authorship

D) Conducted studies and own contribution

A) Curriculum vitae

For reasons of data protection,
the curriculum vitae
is not included in the online version.

B) Publications

2016 Roa Romero, Y.; Keil, J; Balz, J.; Gallinat, J. and Senkowski, D. (February 2016). Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia. *Journal of Neurophysiology* (*submitted 2nd February*).

2016 Roa Romero, Y.; Keil, J; Balz, J.; Niedeggen, M.; Gallinat, J. and Senkowski, D. (2016). Alpha-band oscillations reflect altered multisensory processing of the McGurk Illusion in Schizophrenia. *Frontiers in Human Neuroscience,*10 (41).

2016 Balz, J.; Keil, J; Roa Romero, Y.;Mekele, R.; Schubert, F.; Aydin, S.; Ittermann, B.;Gallinat, J. and Senkowski, D. (2016). GABA concentration in superior temporal sulcus predicts gamma power and perception in the sound-induced flash illusion. *NeuoImage,*15 (125)*:7*20-734.

2015 Roa Romero, Y.; Senkowski, D. and Keil, J., (2015). Early and Late Beta Band Power reflects Audiovisual Perception in the McGurk Illusion. *Journal of Neurophysiology,*13 (7)*:*2342-2350.

2014 Roa Romero, Y.; Miltner, W.H.R. and Weiss, T., (2014). Pain-induced attention allocation effects versus distraction from pain: Competition over attention resources. *Der Schmerz;* 28(4):

2013 Romero, Y. R.; Straube, T.; Nitsch, A.; Miltner, W.H.R. and Weiss, T., (2013). Interaction between stimulus intensity and perceptual load in the attentional control of pain. *Pain,* 154(1): 135-140.

C) Statement of authorship

I herby certify that this dissertation has been composed by me and is based on my own work, unless otherwise stated. Ideas and thoughts cited directly or indirectly from other work have been cited accordingly.

Date signed	**Signature**

D) Conducted studies and own research contribution

The studies of this thesis were coauthored and supported by a number of colleagues.

They are listed below together with my own research contribution.

Study 1:

Roa Romero, Y.; Senkowski, D. and Keil, J., (2015). Early and Late Beta Band Power reflects Audiovisual Perception in the McGurk Illusion. *Journal of Neurophysiology,13* (7)*:*2342-2350.

I supported the planning, design and implementation of the experiment. Further I recruited participants and carried out the behavioral and EEG experiment, performed the analysis, interpreted the results and drafted the manuscript.

Study 2:

Roa Romero, Y.; Keil, J.; Balz, J.; Niedeggen, M.; Gallinat, J. and Senkowski, D. (2016). Alpha-band oscillations reflect altered multisensory processing of the McGurk illusion in schizophrenia. *Frontiers in Human Neuroscience,10* (41)*.* epub ahead of print. doi:10.3389/fnhum.20016.00041

Accepted in *Frontiers of Human Neurosciene* (25.01.2016 current status under production)

I supported the planning, design and implementation of the experiment. I recruited the schizophrenia patients and matched controls, carried out the EEG experiment and collected the behavioral and EEG data, performed the analysis, interpreted the results and drafted the manuscript.

Study 3:

Roa Romero, Y.; Keil, J.; Balz, J.; Gallinat, J. and Senkowski, D. (submitted). Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia. *Journal of Neurophysiology.*

Currently submitted in *Journal of Neurophysiology* (02.02.2016 current status under review)

I supported the planning, design and implementation of the experiment, recruited matched controls and patients, carried out the EEG experiment and collected the behavioral and EEG data, performed the analysis, interpreted the results and drafted the manuscript.