Sarah Jessen: Emotion Perception in the Multisensory Brain. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences, 2012 (MPI Series in Human Cognitive and Brain Sciences; 140)

Emotion Perception in the Multisensory Brain

Impressum

Max Planck Institute for Human Cognitive and Brain Sciences, 2012

Druck: Sächsisches Druck- und Verlagshaus Direct World, Dresden

Titelbild: (http://en.wikipedia.org/wiki/File:Penguins_walking_-Moltke_Harbour,_South_Georgia,_British_overseas_territory,_UK-8.jpg); © Liam Quinn

ISBN 978-3-941504-24-0

Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin

Emotion Perception in the Multisensory Brain

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

Vorgelegt von
Master of Science in Neural and Behavioral Sciences
Jessen, Sarah

Berlin, 2011
Tag der Verteidigung: 25. Mai 2012

Erstgutachter:

Prof. Dr. Arthur Jacobs

Zweitgutachterin:

Prof. Dr. Sonja Kotz

Acknowledgments

This thesis is the result of the time I spent at the Max Planck Institute for Human Cognitive and Brain Sciences as a PhD student. There a many people who helped and supported me during this time and to whom I am deeply grateful.

First of all, I would like to thank Sonja Kotz for being a great supervisor throughout my time as a PhD student. Thank you for your encouragement, your advice, and for all the things I was able to learn during the past years. Thank you for giving me so many great opportunities and for always having time for all the smaller and bigger problems during my PhD project.

Second, I would like to thank Arthur Jacobs for making it possible for me to work on my PhD thesis, for his support and advice.

Furthermore, I would like to thank the graduate school of the cluster of excellence “Languages of Emotion”. First, for the stipend, which allowed me to pursue my PhD project, but also for a great and inspiring environment which I was allowed to enjoy during my time there. Many thanks in particular to Gisela Klann-Delius und Markus Edler for their support, and to my fellow graduate students for many fruitful, interesting, and enjoyable discussions.

Thank you to all my colleagues at the Max Planck Institute who helped me during various stages of my project. In particular, many thanks to the people who helped me with all my questions about fMRI analysis, especially Jöran Lepsien, Kathrin Rothermich, and Anna Mestres-Misse.

Special thanks also to the “Neurocognition of Rhythm in Communication”-group; I had a great time with you!

Thank you to Cornelia Schmidt, Heike Boethel, Anke Kummer, and Simone Wipper for helping me with the data acquisition. Thanks to the graphics department, especially to Kerstin Flake and Stephan Liebig, for help with creating the figures and answering all my questions about recording, cutting, and preparing videos.

Great thanks to all my friends who were always there for me during all the ups and downs of my PhD; I couldn't have done without you. Thanks to the band, for many enjoyable evenings and a great distraction from work. And thank you especially to Iris Knierim, Lars Meyer, and Claudia Männel. For proof-reading my thesis, but much more for your friendship and the great times we spent together during the past years.

Last but not least, I am deeply grateful to my family for their never-ending love and support. Thank you for always believing in me, for being there for me, and for helping me whenever possible. Jonas, thank you for your love, your understanding, and for our shared hopes and plans. I cannot express how thankful I am to have found you.

Berlin, 2011

Contents

List of Figures

List of Tables

Abbreviations

A = auditory

ACC = anterior cingulate cortex

Ag-AgCl = silver–silver chloride

ANOVA = analysis of variance

ASD = autism spectrum disorder

AV = audiovisual

BA = Brodman area

BOLD = blood-oxygenation-level dependent

cm = centimeter

DC = direct current

dB = decibel

DICS = dynamic imaging of coherent sources

DSM-IV = Diagnostic and Statistic Manual of Mental Disorders IV

EEG = electroencephalography

EBA = extrastriate body area

EOG = electrooculography

ERP = event-related potential

FBA = fusiform body area

fMRI = functional magnetic resonance imaging

GLM = general linear model

(k)Hz = (kilo)Hertz

IE = inverse effectiveness

IFG = inferior frontal gyrus

kΩ = kiloΩ

LPC = late positive complex

MDD = major depressive disorder

MEG = magnetoencephalography

MNI = Montreal Neurological Institute

MPFC = medial prefrontal cortex

MR(I) = magnetic resonance (imaging)

ms = milliseconds

MTG = middle temporal gyrus

N = number (of participants)

n.s. = not significant

μV = micro Volt

OFC = orbitofrontal cortex

RMS = root mean square

s = seconds

SAM = self-assessment mannequin

SD = standard deviation

SE = standard error

SNR = signal-to-noise ratio

STS = superior temporal sulcus

STG = superior temporal gyrus

V = visual

vmPFC = ventro-medial prefrontal cortex

VPP = vertex positive potential

English Summary

Perceiving and correctly recognizing others' emotions is a crucial aspect of everyday social interaction. Humans can do so very successfully via auditory information, such as speech content and voice, as well as visual information, such as facial and bodily expressions. In most situations, emotional information can be obtained from a number of these channels simultaneously. While numerous previous studies have investigated the interaction between two channels (Ethofer, Anders, Erb, Droll, et al., 2006; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Robins, Hunyadi, & Schultz, 2009), the interplay between more than two channels has hardly been addressed. Furthermore, most studies on multisensory emotion perception focus on facial expressions and semantic information (Ethofer, Anders, Erb, Droll, et al., 2006; Dolan, Morris, & de Gelder, 2001), while vocal information and body expressions are often neglected. The aim of my thesis is therefore to investigate the neural basis underlying the perception of emotional information from face and especially voice and body expressions. Furthermore, I focus on the investigation of dynamic visual information, thereby allowing for an investigation of multisensory perception approximating ecologically valid settings as close as possible.

Emotional voice expressions (i.e. interjections such as "ah") are particularly important for several reasons. First, they occur very spontaneously, thereby providing a direct link to a person's emotional state. Second, they allow for a reliable recognition of someone's emotional state (Dietrich, Hertrich, Alter, Ischebeck, & Ackermann, 2008). Third, from an empirical point of view, they allow the investigation of voice perception with only a minimal semantic confound.

Likewise, body expressions also play an important role in the perception of someone's emotions. Not only are we able to tell someone's emotion (Atkinson, Dittrich, Gemmell, & Young, 2004), but also the action someone is intending to take in response to that emotion.

Body expressions, unlike facial expression, thereby provide a direct link between emotion and action (de Gelder, 2009).

Previous studies regarding the neural substrate underlying the perception of emotional information have revealed a broad network involved at various stages of emotion perception. While early sensory areas (e.g. the superior temporal gyrus and sulcus (STG/STS) and the fusiform gyrus) are involved in the perceptual aspects of emotion perception, hierarchically later areas (e.g. insula, anterior cingulate cortex) are rather involved in evaluating emotional information (Allison, Puce, & McCarthy, 2000; Ethofer, Anders, Erb, Droll, et al., 2006). Multisensory specific activations can be observed primarily in the STG and the thalamus (Kreifelts et al., 2007; Peelen, Atkinson, & Vuilleumier, 2010).

To investigate both the temporal as well as the spatial aspects of multisensory emotion perception, two electroencephalographic (EEG) studies and one functional magnetic resonance imaging (fMRI) study were conducted.

Electroencephalography measures electric brain responses on the scalp, thereby offering a high temporal but only a low spatial resolution. The EEG analysis was focused on the one hand on event-related potentials (primarily the auditory N100) and on the other hand on brain oscillations in the alpha- and beta-band range. Functional MRI in contrast to EEG relies on changes in the blood-oxygenation-level dependent (BOLD) signal, which allows for a low temporal but high spatial resolution. Besides the imaging studies, several behavioral pilot studies were conducted to evaluate the stimulus material used.

The experimental paradigm used is comparable for all imaging studies; participants were presented with emotional information (anger, fear, and a neutral control condition) in three different modality conditions. In an auditory condition, they heard emotional or neutral interjections; in a visual condition, they saw short video clips showing emotional or non-emotional body expressions; and in an audiovisual condition, they saw video clips and heard matching interjections.

In the first EEG study, the auditory N100 was strongly reduced in amplitude for audiovisual compared to auditory stimuli. Based on comparable results by Stekelenburg and Vroomen (2007), this reduction can be interpreted as a very early facilitation in processing due to an interaction between predictive visual and auditory information. Likewise, a reduction in N100 amplitude was found for emotional compared to neutral stimuli, again indicative of facilitated processing (Besle, Fort, Delpuech, & Giard, 2004; van Wassenhove, Grant, & Poeppel, 2005). An interaction between both factors revealed that in the auditory but not in the audiovisual condition, an additional differentiation between anger and fear stimuli occurred, providing first evidence that unimodal stimuli may necessitate a more detailed processing in comparison to audiovisual stimuli. In addition to these ERP results, we computed a time–frequency-analysis. This analysis revealed a highly similar processing pattern in the alpha- and beta-band for visual and audiovisual stimuli.

While a very early influences of modality as well as affective content was observed in the first EEG study, one can only draw limited conclusions about involved brain areas due to the limited spatial resolution of EEG. Hence, a very similar paradigm was implemented in an fMRI study in order to further investigate the neural networks involved. For both unimodal conditions, activations were observed in areas commonly reported in emotion processing from body and vocal expressions, respectively (Grèzes, Pichon, & de Gelder, 2007; Pichon, de Gelder, & Grèzes, 2008, 2009). An increased BOLD response for emotional compared to nonemotional body expressions was found in occipital areas, the fusiform gyrus, the extrastriate body area, and the hippocampus. In the auditory condition, emotional stimuli elicited an increased response in the amygdala, the inferior frontal gyrus, and the anterior cingulate cortex.

However, in the audiovisual condition, activation was strongly decreased for emotional stimuli, especially in the bilateral STG/STS. Concomitantly, activity in the fusiform gyrus increased for emotional stimuli compared to neutral stimuli. Hence, the preceding and ongoing visual information seems to result in a reduced auditory activation for emotional stimuli.

In sum, the results of the fMRI study extend previous studies on audiovisual emotion perception by showing that a combination of emotional and multimodal processing might result predominantly in a decrease, rather than an increase in BOLD response.

The first EEG study and the fMRI study both show a clear interaction between auditory and visual processing of body and voice information. However, it has been a longstanding debate in the field of multisensory perception research how such an interaction between two modalities can be qualified as an integration, that is, as a truly combined process rather than two parallel processing streams. It has been suggested that a true integration should be enhanced under suboptimal sensory conditions, a criterion that has been termed “inverse effectiveness” (B. Stein & Meredith, 1993). While this criterion has been employed in several fMRI studies (Holle, Obleser, Rueschemeyer, & Gunter, 2010; Werner & Noppeney, 2010b), only one EEG study has used it in investigating the multisensory integration of perceptually less complex material (Senkowski, Saint-Amour, Höfle, & Foxe, 2011).

In the second EEG study, the signal quality in the auditory stimuli was therefore manipulated to investigate audiovisual interactions at high and low levels of noise. Only in the high-noise condition, earlier N100 peak latencies were observed for audiovisual compared to auditory stimuli. This observation confirms the predictions of the inverse effectiveness principle, suggesting an actual integration of body and voice information. Further support for this assumption arises from the pattern observed in the beta-band suppression. Larger suppression differences between audiovisual and auditory stimuli were found in high-noise emotional conditions compared to low-noise emotional conditions. No such difference occurred for neutral (i.e. non-emotional) stimuli. Not only is this observation predicted by the inverse effectiveness principle, it additionally suggests a modulation of the integration process by emotional content.

Overall, data demonstrate an interaction between body and voice information, which occurs early in processing and which is characterized by an integration between both modalities. Furthermore, a clear effect of emotional content on the processing of body and voice information was found. On a neural level, this emotion processing invokes a network consisting of early modality-specific as well as later, modality-independent areas involved in the evaluation of emotional content. Furthermore, the data suggest different processes underlying the multisensory perception of emotional and non-emotional information.

Several interesting follow-up studies based on the presented data are possible. First, an application to clinical settings would be of high relevance. Various psychiatric but also developmental disorders are characterized by deficits in social skills in general and emotion perception in particular. Additionally, several patient groups, such as patients suffering from schizophrenia (see chapter 7) and patients with autism spectrum disorders, are assumed to have deficits in sensory perception. Especially in these patient groups, a closer investigation of the interaction between emotional and sensory deficits would therefore be of high importance for a better understanding of the respective symptoms.

Second, emotion perception presents an interesting testbed for the investigation of multisensory perception in general. Not only is emotion perception a highly familiar and behaviorally relevant process; emotional information is present naturally in several modalities at the same time so that multisensory interactions can be observed in an ecologically valid setting. For instance, how is the amplitude- and phase-coupling of neural oscillations (two mechanisms that have been implied in multisensory integration) affected by emotional content of the signal? How important is the temporal predictability as well as the predictability of the emotional content for successful multisensory integration? And how may multisensory integration differ from the presented findings for positive or social emotions? These are some of the questions that will have to be addressed in future studies to achieve a better understanding of the influence emotional information may have on multisensory integration.

In sum, the data presented here provide a valuable starting point in the investigation of the brain mechanisms underlying complex multisensory emotion perception.

Deutsche Zusammenfassung

Die Fähigkeit emotionale Zustände und Empfindung Anderer wahrzunehmen und korrekt zu interpretieren ist eine essentielle Grundlage sozialen Zusammenlebens. Dabei können Menschen die Emotionen Anderer anhand deren Stimme oder etwaigen sprachlichen Inhalts, aber auch anhand von Gesichtsausdrücken und Körperbewegungen zuverlässig beurteilen. Normalerweise ist emotionale Information nicht nur in einem, sondern in mehreren dieser Informationskanäle synchron enthalten. Während zahlreiche vorhergehende Studien die multisensorische Emotionswahrnehmung aus zwei dieser Informationskanäle untersucht haben (Ethofer, Anders, Erb, Droll, et al., 2006; Kreifelts et al., 2007; Robins et al., 2009), weiß man nur wenig über das Zusammenspiel von drei oder mehr Kanälen. Des Weiteren konzentrierten sich bisherige Studie vor allem auf die Emotionswahrnehmung anhand von Gesichtsausdrücken und semantischem Inhalt der Sprache (Ethofer, Anders, Erb, Droll, et al., 2006; Dolan et al., 2001), wohingegen Emotionswahrnehmung aus Stimme und Körperbewegungen weitaus weniger untersucht wurde. Das Ziel meiner Arbeit ist daher die Untersuchung der neuralen Vorgänge, die der simultanen Emotionswahrnehmung aus Gesichtsausdrücken und insbesondere Stimme und Körperbewegungen zugrunde liegen. Darüber hinaus konzentriere ich mich auf die Verarbeitung dynamischer visueller Informationen, da dies die visuelle Wahrnehmung in einer natürlichen Umgebung widerspiegelt und so die Untersuchung multisensorischer Emotionswahrnehmung unter ökologisch validen Bedingungen ermöglicht.

Emotionale Lautäußerungen (bspw. Interjektionen, wie “ah”) sind in diesem Zusammenhang von besonderem Interesse. Zum einen werden sie sehr spontan geäußert und ermöglichen dadurch einen direkten Ausdruck des aktuellen emotionalen Zustands einer Person. Zum Anderen zeigen Studien, dass Emotionen sehr zuverlässig anhand solcher kurzen Äußerungen erkannt werden könne (Dietrich et al., 2008). Darüber hinaus erlauben Interjektionen die

empirische Untersuchung emotionaler Stimmwahrnehmung ohne den möglicherweise konfundierenden Faktor Semantik.

Körperausdrücke spielen eine ähnlich wichtige und interessante Rolle in der Emotionswahrnehmung. Auch sie erlauben das korrekte Erkennen emotionaler Zustände (Atkinson et al., 2004), und lassen desweiteren Schlussfolgerungen über beabsichtigte Handlungen zu. Im Gegensatz zu Gesichtsausdrücken sind Körperbewegungen somit ein Bindeglied zwischen Emotionen und Handlungen (de Gelder, 2009).

Emotionswahrnehmung kann in verschiedene Schritte unterteilt werden, die mit verschiedenen Hirnarealen assoziiert sind, sodass ein breites neurales Netzwerk entsteht. Während frühe sensorische Areale (z.B. im Gyrus temporalis superior und im Gyrus fusiformis) für die perzeptuelle Verarbeitung emotionaler Informationen zuständig sind, findet eine evaluative Verarbeitung emotionalen Inhalts in hierarchisch späteren Arealen statt (z.B. der Cortex insularis und anteriore cinguläre Cortex) (Allison et al., 2000; Ethofer, Anders, Erb, Droll, et al., 2006). Spezifische Aktivierungen für multisensorische Verarbeitung lassen sich vor allem im Gyrus temporalis superior und im Thalamus beobachten (Kreifelts et al., 2007; Peelen et al., 2010).

Um sowohl zeitliche als auch räumliche Aspekte multisensorischer Emotionswahrnehmung zu untersuchen wurden zwei elektroenzephalograpische (EEG) Studien sowie eine funktionelle Magnetresonanztomographie (fMRT) Studie durchgeführt. In der elektroenzephalographischen Untersuchung wird die elektrische Hirnaktivität auf der Kopfoberfläche gemessen, ein Verfahren, das eine hohe zeitliche jedoch nur begrenzte räumliche Auflösung aufweist. Die EEG-Analyse konzentrierte sich in den vorliegenden Studien zum einen auf die Berechnung ereigniskorrelierter Potentiale (EKPs), zum anderen auf Hirnoszillationen im Alpha- und Beta-Band-Bereich. Im Gegensatz zum EEG beruht die funktionelle MRT auf der Messung von Änderungen im Blutsauerstoffgehalt (blood-oxygenation-level dependent signal; BOLD), sodass diese Methode eine niedrige zeitliche, dafür aber hohe räumliche Auflösung ermöglicht. Neben diesen bildgebenden Studien wurden mehrere Verhaltensstudien zur Evaluierung des benutzten Stimulusmaterials durchgeführt.

Für alle drei bildgebenden Studien wurden ein sehr ähnliches Paradigma verwendet: Den Probanden wurden emotionale Informationen (Angst, Wut, sowie eine neutrale Kontrollbedingung) in drei verschiedenen Modalitätsbedingungen dargeboten. In der auditiven Bedingung wurden ihnen emotionale oder neutrale Interjektionen vorgespielt; in der visuellen Bedingungen sahen sie kurze Videoausschnitte in denen emotionale oder neutrale Körperbewegungen gezeigt wurden; und in der audiovisuellen Bedingungen wurden Videoausschnitte mit den dazu passenden Interjektionen präsentiert.

In der ersten EEG-Studie ließ sich eine starke Reduktion der Amplitude der auditiven N100 für audiovisuelle im Vergleich zu auditiven Stimuli beobachten. Basierend auf vergleichbaren Ergebnissen von Stekelenburg and Vroomen (2007) lässt sich diese Reduktion als eine sehr frühe Fazilitierung der neuralen Verarbeitung durch eine Interaktion zwischen prädiktiver visueller und auditiver Information interpretieren. Desweiteren fiel die N100-Amplitude für emotionale Stimuli niedriger aus als für neutrale Stimuli, was wiederum auf eine fazilitierte Verarbeitung hindeutet (Besle et al., 2004; van Wassenhove et al., 2005). Eine Interaktion zwischen den Faktoren Emotion und Modalität zeigte darüber hinaus, dass eine Differenzierung zwischen Angst- und Wut-Stimuli in der auditiven, nicht jedoch in der audiovisuellen Bedingung auftritt. Dies liefert erste Hinweise darauf, dass unimodale Stimuli möglicherweise eine elaboriertere Verarbeitung als multimodale Stimuli erfahren. Neben diesen EKP-Ergebnissen wurde eine Zeit–Frequenz-Analyse durchgeführt. Hier zeigten sich starke Ähnlichkeiten zwischen der Verarbeitung visueller und audiovisueller Stimuli.

Die Ergebnisse der ersten EEG-Studie deuten somit auf sehr frühe Verarbeitungseinflüsse sowohl durch den affektiven Gehalt als auch durch die involvierten Modalitäten hin. Jedoch lassen sich aufgrund der niedrige räumlichen Auflösung des EEGs kaum Rückschlüsse auf beteiligte Hirnareale ziehen. Daher wurde ein vergleichbares Experiment in einer fMRT-Studie implementiert um nähere Erkenntnisse über die neuralen Grundlagen der beteiligten Prozesse zu ermöglichen. In beiden unimodalen Bedingungen fanden sich Aktivierungen in Arealen, die bereits in vorherigen Studien mit der Verarbeitung emotionaler Informationen aus Stimme bzw. Körperbewegungen in Zusammenhang gebracht worden sind (Grèzes et al., 2007; Pichon et al., 2008, 2009). So zeigte sich eine erhöhte BOLD-Antwort für emotionale im Vergleich zu neutralen Körperbewegungen im occipitalen Cortex, dem Gyrus fusiformis, dem Hippocampus sowie der sogenannten “extrastriate body area” im extrastriatalem Cortex. In der auditiven Bedingung führten emotionale Stimuli zu einem verstärkten BOLD-Signal in der Amygdala, im Gyrus frontalis inferior sowie dem anterioren cingulären Cortex.

In der audiovisuellen Bedingung jedoch fand sich das BOLD-Signal im bilateralen Gyrus und Sulcus temporalis superior stark reduziert für emotionale, verglichen mit neutralen Stimuli.

Umgekehrt ließ sich eine verstärkte Aktivierung im Gyrus fusiformis für emotionale Stimuli beobachten. Prädiktive und anhaltende visuelle Information scheint somit zu einer reduzierten auditiven Aktivierung für emotionale Stimuli zu führen.

Zusammenfassend erweitern die Ergebnisse der fMRT-Studie den Wissensstand aus bisherigen Forschungsergebnissen dahingehend, dass die Kombination affektiven Inhalts mit multimodaler Wahrnehmung vornehmlich zu einer relativen Deaktivierung statt einer Aktivierung in auditiven Arealen führen kann.

Sowohl die erste EEG- als auch die fMRT-Studie belegen eine Interaktionen zwischen auditiver Stimmverarbeitung und visueller Verarbeitung von Körperbewegungen. Wann eine solche Interaktion jedoch als tatsächliche Integration zweier Modalitäten, d.h. eine kombinierte Verarbeitung im Gegensatz zu zwei parallelen Prozessen, gilt, ist eine vieldiskutierte Frage im Gebiet der multisensorischen Wahrnehmungsforschung. Ein mögliches Kriterium für einen Integrationsprozess ist das "Inverse Effectiveness"-Prinzip (B. Stein & Meredith, 1993), das besagt, dass sich für eine tatsächliche Integration eine Verstärkung der Effekte unter suboptimalen sensorischen Bedingungen beobachten lassen sollte.

Dieses Kriterium wurde in einer Vielzahl an fMRT-Studien verwendet (Holle et al., 2010; Werner & Noppeney, 2010b). Jedoch hat bisher lediglich eine EEG-Studie mit Hilfe des "Inverse Effectiveness"-Kriteriums die multisensorische Integration von perzeptuell relativ simplem Material untersucht (Senkowski et al., 2011). In der zweiten EEG-Studie wurde daher zusätzlich die Signalqualität der auditiven Stimuli manipuliert, um audiovisuelle Interaktionen unter optimalen und suboptimalen sensorischen Bedingungen zu untersuchen.

Eine kürzere N100-Latenz ließ sich nur in der suboptimalen, nicht jedoch in der optimalen sensorischen Bedingung beobachten. Diese Beobachtung bestätigt die Annahme des "Inverse Effectiveness"-Kriteriums, und deutet somit auf eine tatsächliche Integration von Körperbewegungen und Stimme hin. Auch die beobachteten Veränderungen der Hirnoszillationen im Beta-Band stützen diese Hypothese: Suboptimale sensorische Bedingungen führten zu einem größeren Suppressionsunterschied der Oszillationen zwischen audiovisuellen und auditiven Stimuli als optimale sensorische Bedingungen. Dies war jedoch nur für emotionale und nicht für neutrale Stimuli der Fall. Somit sind auch diese Beobachtungen in Übereinstimmung mit dem "Inverse Effectiveness"-Kriterium und deuten darüber hinaus auf eine Modulation der Integration durch den emotionalen Gehalt hin.

Insgesamt demonstrieren die Daten eine Interaktion zwischen Stimminformation und Information aus Körperbewegungen, die früh in der Verarbeitung stattfindet und durch eine Integration beider Modalitäten charakterisiert ist. Darüber hinaus zeigt sich ein klarer Einfluss des emotionalen Gehalts auf die Verarbeitung von Stimme und Körperbewegungen. Dabei wird ein Netzwerk im Gehirn aktiviert, das sowohl frühe, modalitätsspezifische als auch spätere, modalitätsunabhängige Areale beinhaltet. Die Ergebnisse deuten darauf hin, dass klare Unterschiede in der multisensorischen Verarbeitung zwischen emotionalen und nichtemotionale Stimuli bestehen.

Aus den dargelegten Ergebnissen ergeben sich verschiedene vielversprechende Nachfolgeuntersuchungen. Zum einen scheint ein Anwendung auf klinischem Gebiet äußerst interessant; verschiedene psychiatrischen Erkrankungen, aber auch Entwicklunsgstörungen, sind durch Defizite in sozialen Fähigkeiten allgemein sowie speziell im Bereich der Emotionswahrnehmung charakterisiert, wie Kapitel 7 argumentiert. Darüber hinaus finden sich bei einigen Patientengruppen, wie beispielsweise Schizophreniepatienten und Patienten mit Erkrankungen aus dem Autismus-Spektrum, Auffälligkeiten in der sensorischen Verarbeitung. Insbesondere bei diesen Patientengruppen wird es daher besonders wichtig sein, das Zusammenspiel beider Faktoren und die beobachteten Defizite besser zu verstehen.

Zum anderen ist die Untersuchung multisensorischer Emotionswahrnehmung ein spannender Testfall für multisensorische Wahrnehmung im Allgemeinen. Emotionswahrnehmung ist nicht nur ein alltäglicher und behavioral hochrelevanter Prozess; darüber hinaus werden Emotionen auch sehr oft in mehreren Modalitäten gleichzeitig ausgedrückt, sodass multisensorische Interaktionen in einem natürlichen Kontext untersucht werden können. Wie wird beispielsweise die Amplituden- oder Phasenkopplung von Hirnoszillationen, zwei Mechanismen, die mit der Integration multisensorischer Emotionen in Verbindung stehen, durch den emotionalen Gehalt beeinflusst? Wie wichtig ist die zeitliche Prädizierbarkeit sowie die Prädizierbarkeit des emotionalen Inhalts für eine erfolgreiche multisensorische Integration? Und wie unterscheidet sich die multisensorische Integration für positive oder soziale Emotionen? Dies sind einige der Fragen, deren Beantwortung in zukünftigen Studien ein besseres Verständnis des Zusammenspiels zwischen Emotionen und unterschiedlichen Modalitäten ermöglichen kann.

Zusammenfassend stellen die hier beschriebenen Resultate einen wichtigen Ausgangspunkt für die weiterführende Untersuchung der neuralen Mechanismen dar, die komplexer multisensorischer Emotionswahrnehmung zugrunde liegen.

Part I

Introduction

Chapter 1

General Introduction

Communication is an essential aspect of human life. We constantly interact with other people; we hear what they say and how they say something, we see what they do and how they do something, and we interpret this information to understand someone's feelings, intentions, and motivations. These mechanisms are the building blocks of any social interaction.

While speech has a primary role in human communication, other channels such as facial expressions, bodily expressions, and non-linguistic vocalizations[1] are essential as well. Especially when it comes to the communication of basic needs and emotions, the importance of non-verbal channels increases. This can be illustrated by a number of everyday examples. A newborn baby cannot speak and also cannot communicate abstract messages. One thing it can express though with great emphasis is his discontent with a situation. Even though we – as adults – are capable of speech, non-verbal communication does not lose its importance throughout our lifespan. Being in a foreign country, unable to speak one word of the local language, we may not be able to understand the meaning of the words the man whose car we just drove into is uttering. However, we will probably understand that he is angry (and can also infer contextually what he is angry about). Likewise, when we come home after a long day of work, our partner will be able to tell quite reliably whether it was a good or bad day before we have said a single word. Hence, while speech is essential for communicating detailed,

---

[1] While I use the term "non-linguistic" to refer to sound utterances that do not have any (or at least no clearly defined) semantic content, I use the term "non-verbal" to refer to the set of all communication channels that are not speech. In the current thesis, these include non-linguistic utterances, facial expressions, and body expressions.

complex, and abstract information, non-verbal expressions are ideal for rudimentary, basic, and often emotional communication.

Non-verbal emotional communication can be investigated from various starting points. One approach is to ask how the various non-verbal channels – facial, bodily, and vocal expressions – interact to form a coherent percept. The content of a percept can be emotional, but the general principles are assumed to hold true also for non-emotional communication.

Another approach is to ask how emotional content influences the communication process. Such communication can be unimodal, but as outlined already by Charles Darwin (1872), emotions are usually communicated via numerous modalities.

In recent years, researchers have started to bring these two approaches together to achieve a more coherent understanding of non-verbal emotional communication. This endeavor has additionally been fueled by the advent of modern neuroimaging techniques allowing for a non-invasive investigation of neural mechanisms in healthy populations.

While previous studies have greatly advanced our understanding of multisensory emotion perception, many open questions remain. One aspect that has rarely been investigated is the perception of emotions with ecologically valid, dynamic stimuli. Typically, multisensory emotion perception has been investigated by using emotional words or sentences combined with photographs of emotional facial expressions. While this allows for a well-controlled experimental setting, it cannot capture the complexity of emotion perception in a natural environment, including facial and vocal, but also bodily expressions, dynamic auditory and visual information, context information and previous knowledge.

The aim of the current thesis is therefore to unite research approaches from multisensory research as well as emotion perception research in order to investigate the neural basis of multisensory emotion perception in an ecologically valid form.

1.1 Multisensory Perception

Humans can perceive their environment via multiple senses, such as vision, audition, olfaction, and proprioception. The various modalities usually provide congruent information and can be clearly attributed as originating from the same or different sources: If we see and hear a person speaking, we effortlessly attribute the voice we hear to the face we see. Furthermore, the words said, the tone of the voice, and the facial expression will (apart from notable exceptions such as irony or sarcasm) provide converging information. At first glance, perceiving the same information content from various sources therefore seems rather redundant.

However, there are a number of reasons why this is by no means the case and multisensory perception can prove essential for survival, but also convenient for everyday social interactions and behavior.

While the information provided by different modalities may be quite similar, each sense allows for unique access to our surrounding world (Calvert, Brammer, & Iversen, 1998). Not only does each modality elicit a different sensation; each modality is also specialized in sampling a different aspect of our environment. While audition enables us to distinguish fine-grained temporal differences, vision provides higher spatial acuity (e.g. Calvert et al., 1998).

Furthermore, perceiving information via multiple sources simultaneously may, especially in more complex situations such as social interaction, allow for the distinction of finer nuances of the perceived information. Besides the already mentioned examples of sarcasm and irony, changes in facial expression may allow to better judge the sincerity of a verbal utterance. For similar reasons, speaking to someone on the phone may easily result in misunderstandings or misinterpretations of another person's statement that could have been prevented by simultaneously seeing that person's face.

Besides refining perception, multimodal information can also ensure reliable and successful perception under suboptimal conditions (Sumby & Pollack, 1954; Bertelson & de Gelder, 2004). If we can hardly understand someone we are talking to at a noisy train station, our understanding can be greatly improved by seeing that person's face. Likewise, recognizing someone at night in the street is by far easier if we also hear his or her voice.

Thus, multisensory information facilitates processing, especially in difficult perceptual conditions. At the behavioral level, people are faster in detecting targets that are presented bimodally rather than unimodally (e.g. Giard & Peronnet, 1999; Sperdin, Cappe, Foxe, & Murray, 2009), and they do so with a higher accuracy (e.g. Giard & Peronnet, 1999; Kreifelts et al., 2007). Similar facilitation processes can also be observed in brain responses, for instance in shorter event-related potential (ERP) latencies and smaller ERP amplitudes (e.g. van Wassenhove et al., 2005; Stekelenburg & Vroomen, 2007).

Overall, multisensory perception therefore seems to be essential for the successful interaction with our environment.

1.1.1 Multisensory Perception – When and How?

The question of how and at which stage in processing information from different modalities is perceptually combined has been of long-standing scientific interest (see Section 1.1.2 for a discussion of the difference between multisensory *integration* and *interaction*). As early as 1838, Johannes Müller described the ventriloquist effect as an example of the integration of two modalities (see Bertelson & de Gelder, 2004).

Today, most of our knowledge regarding the neural basis and the time course of multisensory integration in humans originates from modern neuroimaging techniques such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), but also from clinical work on patients who suffer from lesions in specific brain regions. These findings are complemented by electrophysiological studies in non-human animals, allowing for direct recording of neuronal responses.

In the following, the current state of knowledge regarding the neural mechanisms underlying the integration of auditory and visual information will be outlined. With respect to integration processes of other modalities, such as visual and tactile, partly overlapping but also partly distinct brain areas have been suggested (Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005).

One of the classical subcortical loci for early multisensory integration is the superior colliculus; electrophysiological studies in non-human animals show superadditive response properties (see below) at the cellular level (e.g. B. Stein & Meredith, 1993). Other subcortical structures assumed to play a role in multisensory perception are the medial geniculate body (Komura, Tamura, Uwano, Nishijo, & Ono, 2005), as well as the amygdala, particularly its medial part (Nishijo, Ono, & Nishino, 1988).

At the cortical level, several brain regions have been suggested to partake in the integration of information from different modalities. Firstly, studies suggest multisensory properties already in early, classically unimodal areas, such as the primary visual and primary auditory cortex (Calvert et al., 1998, 1999; Ghazanfar & Schroeder, 2006; C. Kayser & Logothetis, 2007). Secondly, several areas have been identified to be specifically involved in the combination of multisensory information, most prominently the anterior as well as posterior superior temporal sulcus (STS, e.g. Calvert et al., 1997; Calvert, Campbell, & Brammer, 2000; Callan et al., 2003; Beauchamp, Lee, Argall, & Martin, 2004; Beauchamp, 2005a), superior temporal gyrus (STG, Calvert et al., 2000; Callan et al., 2003) and middle temporal gyrus (MTG, Beauchamp et al., 2004). In particular, the STS/STG has been suggested to play a key role in the integration of content (or "what") information from different modalities (Amedi et al., 2005). However, also other cortical areas such as the intraparietal sulcus (Werner & Noppeney, 2010a), the posterior insula (Bushara, Grafman, & Hallett, 2001), and a number of frontal areas, for instance, the prefrontal cortex (Laurienti et al., 2003) and the ventral anterior cingulate cortex (Laurienti et al., 2003), are thought to play a role in multisensory processing.

Considering the time course of multisensory integration, most evidence arising from human EEG or magnetoencephalographic (MEG) studies suggests an early perceptual interaction between auditory and visual information. Studies using very simple stimulus material, such as geometrical shapes and sinusoidal tones, report interaction effects between auditory and visual processing within 40 to 60 ms after the onset of a stimulus (Giard & Peronnet, 1999; Molholm et al., 2002; Cappe, Thut, Romei, & Murray, 2010). A source localization of this early effect suggests primary auditory and visual areas as well as the posterior STS as its likely origin (Cappe et al., 2010).

In the investigation of face–voice or multisensory object perception, where more complex stimuli are used, first reliable effects are typically found around 100 ms after stimulus onset, as indicated by a reduction in amplitude (Besle et al., 2004; van Wassenhove et al., 2005; Stekelenburg & Vroomen, 2007) as well as latency (van Wassenhove et al., 2005; Stekelenburg & Vroomen, 2007) of the auditory N100 (see below). One interpretation of this effect is facilitation in auditory processing due to the preceding, predictive visual information. Via back-projections from the STS, temporally predictive visual information is thought to influence auditory processing (Besle et al., 2004; Stekelenburg & Vroomen, 2007). If the preceding visual information fails to reliably predict the auditory information, the N100 reduction disappears (Vroomen & Stekelenburg, 2010).

Besides these very early interaction effects, some studies also report differences associated with more cognitive processing at later stages. A mismatch between auditory and visual information may result in an increased P400 compared to matching stimuli, suggesting a more elaborate, cognitive processing of incongruent information (Puce, Epling, Thompson, & Carrick, 2007; B. Liu, Wang, & Jin, 2009).

Overall, the interaction between between auditory and visual information seems to be a very rapid process involving numerous cortical and subcortical areas.

1.1.2 Multisensory Interaction versus Multisensory Integration

One essential aspect in the investigation of multisensory integration is the distinction between multisensory *interaction* and multisensory *integration*. While one can reliably observe differences when contrasting unimodal and multimodal processing (e.g. in the N100 amplitude), these differences can be explained by a number of factors, only one of them being multisensory integration. Alternatively, such differences may also arise from the parallel processing of two modalities.

This can be demonstrated by considering a reduction in the auditory N100 amplitude, an effect commonly observed in audiovisual perception. In terms of multisensory integration, such an effect would be explained by the facilitatory influence of concurrent visual information on auditory processing. However, the same effect could also be explained by the separate processing of auditory and visual information, that in sum would be reflected as one change in the EEG signal, namely a reduction in the N100 amplitude. This observed interaction would however not reflect an underlying multisensory integration.

In recent years, several paradigms have been developed that aim at differentiating multisensory integration from other forms of multisensory interaction. While these principles should be applicable to any multisensory interaction, the most commonly investigated interaction by far is the one between auditory and visual information. In the following, I will therefore use these modalities as an example to outline different approaches in investigating multisensory integration.

One straightforward way to measure multisensory integration is to compare the activation (in a small group of neurons, but also of an entire brain area) in a bimodal condition to the sum of the activations in response to the unimodal conditions (AV vs. $A+V$[2]). If no integration occurs and the sensory input from both modalities is processed in parallel, the activation in the bimodal condition should by fully explicable by the summed activations in the unimodal

---

[2] A = auditory, V = visual, AV = audiovisual

conditions, i.e. $AV = A + V$. If, however, additional activation is observed, this *superadditivity* is considered as evidence for integration ($AV > A + V$). Conversely, a bimodal activation failing to reach the level expected by the sum of the unimodal activations (*subadditivity*) also provides evidence against a linear summation of unimodal responses and therefore for integration ($AV < A + V$). Integration may thus be characterized by a profile of subadditive and superadditive responses in different brain areas (Stanford & Stein, 2007)

Paradigms investigating super- or subadditivity have been employed in numerous studies (e.g. Calvert, Hansen, Iversen, & Brammer, 2001; Besle et al., 2004; Hagan et al., 2009; Joassin et al., 2011; Love, Pollick, & Latinus, 2011). However, this criterion has been heavily criticized in recent years (Beauchamp, 2005b; Laurienti, Perrault, Stanford, Wallace, & Stein, 2005; Ethofer, Anders, Erb, Droll, et al., 2006). One argument against the reliability of such a criterion is the presence of activity unrelated to the processing of the stimulus itself (Ethofer, Anders, Erb, Droll, et al., 2006). Assuming an approximately equal level of such background activity b in all conditions, the sum of the activation in response to the two unimodal conditions would also include the sum of this background activity ($AV + b$ vs. $(A + b) + (V + b)$). Hence, a difference in activation would be measured simply due to the summed background activity (b vs. $2b$), even if the activation in response to AV and $A + V$ is identical.

A different line of argument specifically questions the applicability of such a paradigm to noninvasive methods as are typically used in humans. Many paradigms investigating multisensory integration, among them super- and subadditivity, were first investigated using single-cell recordings in nonhuman animals (B. Stein & Meredith, 1993). Imaging techniques typically used in healthy human subjects, however, can at best allow inferences about activations in certain brain regions, but do not allow an assessment of the activity of specific neurons or smaller groups of neurons. Hence, even if super- or subadditivity is observed in a given brain area, it remains unknown whether single neurons or neuronal populations respond in a super- or subadditive way (Laurienti et al., 2005). Rather, such an area may comprise neurons responding to auditory as well as neurons responding to visual information, but no neurons with convergent input.

More recently, the super-/sub-additivity criterion has been extended by a number of additional features that characterize multisensory integration. Based again on neurophysiological work, B. Stein and Meredith (1993) posit three criteria for qualifying an interaction between different modalities as integration.

1. *Spatial Congruity.* Signals from two (or more) different modalities show a stronger interaction if they originate from the same location.
2. *Temporal Congruity.* Signals from two (or more) different modalities show a stronger interaction if they occur in temporal synchrony.
3. *Inverse Effectiveness.* Signals from two (or more) different modalities show a stronger interaction if one of the signals is unimodally least effective.

Studies based on these criteria typically investigate super- or subadditivity depending on one (or several) of these features (e.g. Bolognini, Leo, Passamonti, Stein, & Làdavas, 2007; Werner & Noppeney, 2010b). Taking inverse effectiveness as an example, such a study may vary the intelligibility or informativeness of one of the modalities, rendering the signal in this modality less effective.

Werner and Noppeney (2010b), for instance, investigated audiovisual integration in an fMRI study with healthy participants using pictures of everyday object manipulations (e.g. using a hammer) along with matching sounds. Either the visual, the auditory or both channels were degraded. To analyze differences in interaction, contrasts were computed for responses to audiovisual stimuli and to the sum of the unimodal conditions for both intact and degraded stimuli ($((AV + c > A + V)_{intact}$ vs.$(AV + c > A + V)_{degraded}$, where c is a control condition to account for summed activations unrelated to the actual stimuli). While the authors observed superadditive responses in the STS for degraded stimuli, responses to intact stimuli were mainly subadditive. Crucially, superadditivity was reliably predicting the participants' behavioral success, suggesting a functionally relevant role of the STS in multisensory integration.

Similar to the super-/sub-additivity criterion described above, these characteristics were originally developed and investigated in neurophysiological recordings. Hence, again the question remains, to what degree criteria developed based on response properties of local neuronal populations can be transferred to neuroimaging studies.

A different approach in the investigation of multisensory integration, especially of more complex information such as emotion or speech perception, are mismatch paradigms. If auditory and visual information provide conflicting information, differences in the brain response can be observed compared to audiovisual stimuli conveying congruent information. Even if a participant's attention is directed specifically to one modality, the other, incongruent and unattended modality can influence the perception of the first one. Such effects have been observed in audiovisual speech perception (McGurk & MacDonald, 1976) as well as audiovisual emotion perception (de Gelder & Vroomen, 2000).

Unlike the previously outlined criteria, the mismatch criterium allows for an investigation of multisensory integration by comparing one type of audiovisual stimulus to another, hence avoiding possible confounds resulting from the summation of unimodal stimuli. However, multisensory integration cannot be manipulated directly, but only inferred from violation responses to mismatching stimuli. Furthermore, such a manipulation can only be applied to stimuli with a clear relation between auditory and visual signal, so that a clear distinction between matching and mismatching stimuli can be drawn. For a more detailed overview about different ways to quantify multisensory integration, see for instance Ethofer, Anders, Erb, Droll, et al. (2006) or Love et al. (2011).

1.2 Emotion Perception in Communication

As outlined in the previous section, multisensory interaction offers several processing benefits that can be crucial for survival, such as facilitated processing and fast and reliable perception under suboptimal conditions. Furthermore, multisensory perception is an omnipresent phenomenon; healthy human beings (and in fact also most other animals) constantly sample their environment via multiple senses at the same time. One prototypical everyday context in which multisensory perception plays an important role is face-to-face communication.

Face-to-face communication, defined as any direct interaction between two or more people, is the most natural and widespread form of social interaction. It ranges from mother-infant-interaction to an encounter between total strangers to a physical fight between two opponents. The main information sources in human social interaction include auditory and visual information, while the role of other modalities is discussed more controversially (Shepherd, 2004). Auditory information carries into lexical and syntactic as well as prosodic aspects, all of which contribute to the final semantic information after an adequate decoding has taken place. In addition, auditory information may encompass non-linguistic utterances. Visual information encompasses facial as well as bodily expressions.

One crucial aspect of any interaction is the emotional information conveyed by the communication partners. Communication can have an obvious emotional content, such as a conversation about a person's emotional state, in which semantic content, prosody, facial and bodily expression provide congruent information about a person's current emotional state. However, also in less straightforward situations, when the content of a conversation may superficially be quite neutral, humans are very successful at detecting small cues in the voice, the face, or body expressions, transporting information about someone's current emotional state. Such emotion recognition can have a great influence on our response behavior, and is of high importance for maintaining intact social relations. In sum, reliably determining emotional states from such non-verbal emotional cues is of great importance to everyday communication.

Several other aspects render non-verbal emotional communication a particularly interesting topic of investigation. First, while language without doubt is an essential aspect of human communication, non-verbal communication is phylogenetically as well as ontogenetically older (Hauser, 2004). Human beings were able to communicate via non-verbal expressions before speech developed, and non-human animals successfully use body and vocal expressions in social interactions. Any infant can quite clearly communicate his liking or disliking before speaking his first word. Non-verbal communication is therefore the more basic and, in parts, almost reflex–like way to communicate. This becomes especially interesting in the case of emotional communication. Emotions are very fundamental experiences that can guide our behavior and ensure fast responses in threatening situations. Hence, spontaneous emotional responses have to be quick and realized in a way that does not depend on higher cognitive capacities such as language. Thus, non-verbal responses may provide a more direct link to someone's emotional experience offering an interesting testbed in the investigation of emotional communication.

Second, negative emotional content tends to affect responses observed in non-emotional (or positive) information. This can be observed at various levels. Angry faces, for example, are detected faster than happy faces (Fox et al., 2000). Brain responses such as the N100, indicating early auditory processing (Paulmann, Jessen, & Kotz, 2009), or the P200, associated with emotional salience detection (Paulmann & Kotz, 2008a), are reduced for emotional content. Areas typically involved in the processing of faces, such as the fusiform face area, or bodies, such as the extrastriate body area, show increased blood-oxygenation-level dependent (BOLD) responses for emotional compared to neutral stimuli (Vuilleumier, Armony, Driver, & Dolan, 2001; Peelen, Atkinson, Andersson, & Vuilleumier, 2007)[3]. Decreased ERP amplitudes and increased BOLD responses in this case reflect different aspects of sensory perception that can be modulated by emotional content.

Such a modulation in processing has been interpreted as facilitation through emotional content, a prerequisite for fast and reliable responses to emotional information (Kreifelts et al., 2007). In the context of multisensory interaction this feature of emotion perception becomes particularly interesting, as multisensory perception is also characterized by facilitated processing. Multisensory perception may therefore be one factor that ensures fast responses to emotional information. Furthermore, as emotional facilitation can be observed unimodally as well as multimodally, the question arises to what degree emotional and multisensory facilitation interact.

Overall, just as multisensory perception, emotion perception is an essential component of social interaction. While both, multisensory perception and emotion perception, can occur separately, they often also occur in combination, leading to multisensory emotion perception. Investigating the interplay between these two factors can therefore provide valuable information about the mechanisms underlying human communication.

Numerous previous studies have investigated subparts of multisensory emotion perception. A large body of literature exists on the interaction between facial and vocal expressions (e.g. de Gelder, Böcker, Tuomainen, Hensen, & Vroomen, 1999; de Gelder & Vroomen, 2000; Ethofer, Anders, Erb, Droll, et al., 2006; Brosch, Grandjean, Sander, & Scherer, 2009; Hagan et al., 2009; Paulmann et al., 2009), and some studies have also investigated the interplay between facial and body expressions (Meeren, van Heijnsbergen, & de Gelder, 2005; van den Stock, Righart, & de Gelder, 2007) or body and vocal expression (van den Stock et al., 2007; van den Stock, Grèzes, & de Gelder, 2008). However, rarely have these three sources of emotional information been combined in order to investigate multisensory emotion perception approximating natural social conditions (de Gelder, 2009).

---

[3] Note that directionality of ERP components, that is positive vs. negative deflection, as well as amplitude increase vs. reduction, and the relation between changes in ERP amplitude and BOLD amplitude are hard to interpret on an absolute level. The interpretation of the N100 amplitude reduction as well as changes in BOLD signal is addressed in more detail in the general discussion (p. 138ff).

1.2.1 Empirical Foundations of Emotion Perception

The perception of other people's emotions has classically been investigated by using photographs of facial expressions (de Gelder, 2009). Going back to Ekman's work (Ekman, Sorenson, & Friesen, 1969), six basic emotions have been suggested to be universally recognizable: fear, anger, happiness, sadness, disgust, and surprise, plus a non-emotional or neutral state (but see Russell (1994) for a discussion of the universality of basic emotions). In the following years, these emotional states have been amply investigated, both at the behavioral level and at the neural level.

At a formal level, emotional states can be classified by their valence; while anger, fear, sadness, and disgust possess a clear negative connotation, happiness is considered positive. A possible exception from this classification may be surprise (but see Oatley and Johnson-Laird (1987), who suggest that "surprise" can be a feature of several emotions rather than a proper emotion itself). Besides valence, arousal is another parameter commonly used to classify emotions (high arousal: anger, fear, happiness; low arousal: sadness). Furthermore, anger and fear can be considered opposing poles on an approach–avoidance continuum; anger involves the tendency to approach an opponent, fear, in contrast, rather elicits the tendency to retreat from a threat (de Gelder, Snyder, Greve, Gerard, & Hadjikhani, 2004; de Gelder, 2006b). While they therefore differ with respect to the elicited action tendency, both can be considered as signaling threat (Pichon et al., 2009). Due to these factors – high arousal, negative valence, strong behavioral tendencies, threat signaling – anger and fear are known to often elicit strong responses (de Gelder, 2006a; Pichon et al., 2008; Grèzes et al., 2007), a fact which makes them a preferred candidate for the investigation of emotion processing. Due to the strong approach–avoidance contrast, anger and fear furthermore provide a particularly interesting example for the investigation of action tendencies in response to an emotional state. Such action tendencies become particularly interesting in the investigation of emotional body expressions, as outlined below. In the following discussion, and also in the studies reported later on, I will therefore focus on these two emotional states, either in comparison to a non-emotional (neutral) state or in contrast to each other.

Typically, emotional stimuli are recognized faster and with a higher accuracy in comparison to non-emotional control stimuli (e.g. Hansen & Hansen, 1988; Burton et al., 2005). Very fast and efficient processing is also reflected in the brain response. As I will outline in more detail for body and voice expressions below, first processing differences can be observed in the EEG signal around 100 ms after stimulus onset (e.g. Pizzagalli, Regard, & Lehmann, 1999; Batty & Taylor, 2003; Eimer & Holmes, 2007). Later ERP components, such as the late positive complex (LPC, see below), indicate higher level processing, such as evaluation of emotional information (e.g. Marinkovic & Halgren, 1998; Schupp et al., 2000).

According to a model by Adolphs (2002b), the perception of emotions can be broken down into three basic steps; an early processing of salient stimuli (as indicated by the early EEG differences introduced above), a more detailed analysis of the emotional information, and an integration of the perceived information with previous conceptual knowledge. The first, sensory steps are assumed to encompass occipital areas as well as posterior temporal visual cortices involved in the processing of sensory information (e.g. Allison et al., 2000). Furthermore, a direct connection between the superior colliculus and the amygdala – thus "bypassing" visual cortex – can mediate fast emotional reactions (Morris, Ohman, & Dolan, 1999). In the second step, activations of the fusiform gyrus, the STG as well as the orbitofrontal cortex indicate processing of more detailed aspects of the stimuli as well as task relevant processing (Vuilleumier et al., 2001). Finally, the additional activation of somatosensory cortices (Adolphs, Damasio, Tranel, Cooper, & Damasio, 2000) as well as the insula (T. Singer, Critchley, & Preuschoff, 2009) suggest a matching or comparison with the perceived emotion and one's own emotional experience. It should be noted, however, that this model primarily captures the perception of emotions from facial expressions. For more details on emotion perception from other modalities, see below.

One important but often neglected feature of emotion perception from facial (but also bodily) expressions is the fact that they are dynamic, that is, constantly changing over time. Investigating emotion perception by means of static photographs hence introduces an artificial level of abstraction that has to be taken into account when interpreting the activation patterns described above. In fact, several studies have shown that the perception of emotions from dynamic compared to static information can lead to a clear processing advantage. Participants recognize emotions more accurately from dynamic compared to static stimuli (Atkinson et al., 2004) and dynamic displays of emotions are rated more arousing than static ones (Sato & Yoshikawa, 2007).

Similar differences are also reflected in brain activations for dynamic compared to static information. Stronger activations can for instance be observed in the fusiform gyrus and the amygdala for dynamic compared to static emotional facial expressions (LaBar, Crupain, Voyvodic, & McCarthy, 2003; Sato, Kochiyama, Yoshikawa, Naito, & Matsumura, 2004; Yoshikawa & Sato, 2006; Trautmann, Fehr, & Herrmann, 2009). Furthermore, while dynamic emotional stimuli elicit activations in various frontal brain regions, the perception of emotions from static stimuli causes an additional activation in premotor cortex and parietal regions associated with motor imagery (Kilts, Egan, Gideon, Ely, & Hoffman, 2003). Hence, successful processing of static stimuli may require additional motor simulation, increasing the processing load for emotion perception from static stimuli. Such additional processing steps are unnecessary in natural dynamic environments.

1.2.2 Emotion Perception from Voices

Auditory emotion perception can be separated into linguistic and non-linguistic aspects. While semantic information clearly falls in the realm of linguistic processing and vocalizations such as laughter or crying can be considered non-linguistic, prosody can cover linguistic as well as non-linguistic aspects. Sentence prosody conveys linguistic information (e.g. the distinction between a question and a declarative sentence), but also emotional information. Furthermore, interjections such as "ah" and "oh" contain close to no semantic information, but can clearly express emotions by a variation in the tone of voice. An "ah" can indicate both anger as well as fear, depending on its intonation. Hence, non-linguistic interjections offer an ideal possibility to investigate the processing of emotional voice information; the same interjection can be used to convey several emotions, allowing for a maximal control of physical stimulus parameters. At the same time, interjections occur naturally and reflect high ecological validity in contrast to pseudo-language or distorted speech, which is often used to investigate the perception of voice information. In addition, interjections occur very spontaneously (Scherer, 1995), which matches the properties described for non-verbal emotional signals.

Just as facial expressions, emotions can readily be recognized from voice information (Scherer, 1995). It has been argued that this ability is widespread, if not universal (Sauter, Eisner, Ekman, & Scott, 2010). Furthermore, emotion recognition from prosody seems to operate on a similar time scale as suggested for visual information.

As suggested by EEG studies, the processing of voice information is modulated by emotional content early in processing. First differences between emotional and non-emotional vocalizations have been observed in the P200, for vocalizations (Sauter & Eimer, 2010) as well as for sentence prosody (Alter et al., 2003; Paulmann & Kotz, 2008a). While the earlier N100 is associated to the extraction of acoustic cues, the P200 indicates the integration of such cues, for instance with regard to emotional content (Kotz & Paulmann, 2011).

These aspects are also reflected in a recent model proposed by Schirmer and Kotz (Schirmer & Kotz, 2006), which posits three basic steps involved in the perception and processing of prosodic emotional information: sensory processing, the integration of emotionally significant acoustic cues, and evaluation.

Regarding sensory processing, Schirmer and Kotz (Schirmer & Kotz, 2006) propose the involvement of bilateral primary and secondary auditory areas. Already at these early stages of processing, an enhanced activation in response to emotional vocalizations can be observed (Fecteau, Belin, Joanette, & Armony, 2007). Clearer effects, however, become evident in the next processing step, the integration of emotional cues. This integration seems to mainly take place in the right anterior STS and STG, where stronger activations for emotional compared to neutral vocalizations (Morris, Scott, & Dolan, 1999; Fecteau et al., 2007; S. Scott, Sauter, & McGettigan, 2009) and interjections (Dietrich et al., 2008) are observed. Following this integration, higher, rather cognitive processes take place, such as the evaluation of emotional content and the processing of accompanying semantic information. For emotional non-linguistic vocalizations, the superior frontal gyrus (Fecteau et al., 2007) and the ventral prefrontal cortex and insula (Morris, Scott, & Dolan, 1999) have been reported, for interjections the anterior cingulate gyrus (Dietrich et al., 2008). Furthermore, just as in emotional body language, several studies report an enhanced amygdala activation related to the processing of emotions in general, or also in response to threat in particular (S. Scott et al., 1997; Fecteau et al., 2007).

1.2.3 Emotion Perception from Bodies

While plenty of studies have investigated emotion perception from facial expressions, only few studies have addressed the perception of emotional information from body language. Yet, body language appears to be an equally reliable source of emotion information (Atkinson et al., 2004). Furthermore, it can provide information that cannot be obtained from facial expressions alone.

Firstly, we can recognize someone's emotional state in their body expression, even if that person is turning his/her back to us or is not clearly visible; therefore, body expressions can ensure emotion recognition at large distances or under suboptimal visual conditions where only rough shapes are visible (de Gelder, 2009).

Secondly, more than facial expressions, body expressions provide direct information about the action someone is intending to take in response to the experienced emotion (de Gelder, 2009). By someone's body movements, we can directly see whether he/she is preparing to attack or rather to flee. Furthermore, we can obtain more detailed information about the origin of the experienced emotion. Is that person afraid of me? In this case, I'm in a safe position and no action on my part is required. Or is the person rather afraid of something next to me? In this case that something may also be a threat to me, and it may be important for my survival to respond immediately. Studies manipulating the gaze direction of facial expressions have shown that observers are very sensitive to subtle direction cues (Adams, Gordon, Baird, Ambady, & Kleck, 2003; Adams & Kleck, 2005; Hadjikhani, Hoge, Snyder, & de Gelder, 2008), suggesting that humans indeed try to obtain direction information from someone else's emotional state. This aspect of emotional body expression may play an especially important role in the case of emotions eliciting strong action tendencies, such as anger or fear.

Studies investigating the neural mechanisms underlying the perception of emotional body expressions are rare (de Gelder, 2009). Several EEG studies suggest strong similarities between early processing of body and facial expressions. Like facial expressions, body expressions can elicit an N170 (Stekelenburg & de Gelder, 2004; Gliga & Dehaene-Lambertz, 2005), an early component associated with the configural processing of facial (or bodily) expressions (Bentin, Allison, Puce, Perez, & McCarthy, 1996). A study by Meeren, Hadjikhani, Ahlfors, Hämäläinen, and de Gelder (2008), however, suggests that the source of the N170 in response to bodies differs from the source of the N170 in response to faces. While in both cases the fusiform gyrus appears to be involved, the processing of faces involved additional activations in ventral stream areas (inferior occipital gyrus, fusiform gyrus), while the processing of body expressions rather activated areas in the dorsal stream (precuneus, posterior cingulate cortex) (Mishkin, Ungerleider, & Macko, 1983). Processing differences based on emotional content of body expressions can first be observed around 100 ms after stimulus onset. Emotional stimuli cause a reduction in latency of the P100 (van Heijnsbergen, Meeren, Grèzes, & de Gelder, 2007), a component which also robustly indicates emotion perception from facial expressions (Batty & Taylor, 2003).

Emotional body expressions have been found to elicit increased activation in the amygdala, as has been shown by studies comparing fearful (Hadjikhani & de Gelder, 2003; Grèzes et al., 2007; Pichon et al., 2009; van de Riet, Grèzes, & de Gelder, 2009) or angry body expression (Pichon et al., 2008, 2009) to neutral ones. Furthermore, activations in areas associated specifically with the processing of human bodies, such as the fusiform body area (FBA) or the extrastriate body area (EBA) (Peelen & Downing, 2005; Downing, Peelen, Wiggett, & Tew, 2006; Peelen & Downing, 2007), are modulated by the emotion expressed by the perceived body (Peelen et al., 2007). Interestingly, these differences correlate with amygdala activations, suggesting functional connections between the amygdala and EBA/FBA (Peelen et al., 2007). Furthermore, almost all studies contrasting emotional with neutral body language report an enhanced activation for emotional body language in the posterior STS (Grèzes et al., 2007; Pichon et al., 2008, 2009) as well as the dorsal premotor cortex (Grèzes et al., 2007; Pichon et al., 2008). This activation is related to the perception of body motion in general, as both areas are commonly involved in the processing of biological motion and action observation (Grèzes et al., 2007), as well as the preparation of action (Hoshi & Tanji, 2004).

Furthermore, several frontal cortical areas show enhanced activation to emotional stimuli, among them the inferior frontal gyrus, the lateral orbito-frontal cortex (Grèzes et al., 2007; Pichon et al., 2008, 2009), the ventro-medial prefrontal cortex (vmPFC) (Pichon et al., 2008), and the dorso-medial prefrontal cortex (Pichon et al., 2009). In general, these regions have been attributed various functions in the higher-level processing of emotions such as the integrations of emotions into a social context and regulation of behavioral responses.

Considering a contrast between angry and fearful body language, differential activations have been reported in the temporal pole and the anterior STS, which have been explained by the increased involvement of the observer of angry body language compared to fearful body language (Pichon et al., 2009). Furthermore, differences in activation were observed in the vmPFC and the posterior orbitofrontal cortex, suggesting a higher need for a behaviorally adequate response in the case of angry body language and thus a stronger activation of related areas (Pichon et al., 2009).

According to a model proposed by de Gelder (2006b) the areas outlined above are part of three complex interacting subsystems in the brain that subserve the perception and processing of emotional body expressions. A reflex-like system encompassing subcortical structures such as the amygdala and the superior colliculus is assumed to mediate automatic responses to threatening stimuli. A second system provides a cortical input system for information from emotional body expressions, involving a connection between the amygdala and fronto-parietal areas, as well as the STS. Its main role is to integrate the perceived emotion with previous experiences and memory. Both of these systems additionally feed information to a third system, which includes the somatosensory cortex, the insula, the ventromedial prefrontal cortex, and the anterior cingulate cortex, and is assumed to mediate emotional evaluation as well as bodily responses.

1.2.4 Multisensory Emotion Perception

In principle, multisensory integration of emotional information follows similar principles as multisensory integration of other types of complex information, such as audiovisual speech (e.g. van Wassenhove et al., 2005).

One important aspect for audiovisual speech perception and audiovisual emotion perception alike is the dynamic nature of the preceding visual information. As introduced above, facial and body movements are *movements*, and the use of static information creates a clear abstraction from our everyday experience with these movements. This problem becomes especially pressing in the investigation of multisensory perception; auditory information is dynamic and extending over time, also in an experimental context. Hence, static visual information combined with dynamic auditory information creates a mismatch, which may have a strong impact on the integration of the two modalities. To avoid such a mismatch, the studies reported here use only dynamic visual and auditory information.

One paradigm commonly employed in the investigation of multisensory emotion perception is the presentation of incongruent (or mismatching) auditory and visual information, and the analysis of the subsequent violation response (see above). Typically, emotional information in one modality is paired either with a different emotion or with non-emotional information in the other modality.

Using such mismatch paradigms, it has been shown that facial, bodily, and vocal expressions strongly influence each other (de Gelder & Vroomen, 2000; van den Stock et al., 2007, 2008). If, for instance, a happy face is accompanied by a sad voice, participants are more likely to judge a person as sad, as compared to a combination of happy face and happy voice (de Gelder & Vroomen, 2000). This effect is reduced but still visible if participants are instructed to actively ignore the voice, suggesting a partial though not complete voluntary control of the interaction between the modalities (de Gelder & Vroomen, 2000).

As has been shown in EEG studies, this interaction appears early in processing. A mismatch response can be observed around 100 ms after stimulus onset, as can be seen by an increased P100 for a mismatch between face and body information (Meeren et al., 2005), and a decreased N100 for a mismatch between face and voice information (Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000).

Regarding the underlying neural substrate, areas involved in multisensory integration in general, in particular posterior STS and STG, appear to be modulated by the affective content of the stimulus signal (Ethofer, Anders, Erb, Droll, et al., 2006; Kreifelts et al., 2007; Kreifelts, Ethofer, Shiozawa, Grodd, & Wildgruber, 2009; Robins et al., 2009; Peelen et al., 2010). Furthermore, areas activated in the unimodal processing of emotions are modulated by the concurrent perception of another modality, for instance the fusiform gyrus (Dolan et al., 2001; Ethofer, Anders, Erb, Droll, et al., 2006).

Two further areas that have been implicated in multimodal perception of emotions are the amygdala and the orbito-frontal cortex. As has been shown in previous studies both areas are activated in auditory (S. Scott et al., 1997; Belin, Fecteau, et al., 2008) as well as in visual emotion perception (Pichon et al., 2009). Therefore, it has been suggested that both, the amygdala and the orbito-frontal cortex, are involved in cross-modal processing, in which information from one modality is combined with information from a different modality (Dolan et al., 2001; O'Doherty, Rolls, & Kringelbach, 2004). In addition, various other frontal areas have been suggested for the higher-order multisensory representation of emotional information, such as the superior frontal cortex (Müller et al., 2011) or the medial prefrontal cortex (Peelen et al., 2010).

1.3 Present Studies

Starting from the current state of research as reviewed in the previous sections, I will present three studies investigating the complex interplay between modalities in multisensory emotion perception. In particular, these studies will focus on the following aspects:

Body, face, and vocal expressions. While most previous studies have focused on the interaction between two of these modalities, mainly faces and voices, I investigated the interaction between all three channels. This is a necessary step in the investigation of emotional communication in an ecologically valid setting, the reality of everyday social interaction.

Dynamic visual information. To overcome one of the shortcomings in many previous studies investigating multisensory emotion perception, I used dynamic visual information by means of a specifically created set of video clips. Using dynamic material not only increases the ecological validity of the results, but also avoids a possible conflict between static visual and dynamic auditory information.

Integration in congruent settings. A commonly used paradigm in the investigation of multisensory emotion integration are mismatch paradigms, in which conflicting auditory and visual information is presented (see p.9). This only allows for the investigation of responses to a mismatch in multisensory perception, thereby measuring integration only indirectly. In contrast, I investigated multisensory integration by using the inverse effectiveness principle (see p. 8), which has been commonly used in fMRI studies, but only rarely in the EEG. It offers a unique possibility to investigate multisensory integration in a congruent, naturalistic setting.

Before describing the studies and results of this series of experiments in detail, I will briefly describe the methods I employed in these studies. Furthermore, I will describe the stimulus set I developed as well as several pilot studies validating the stimulus-set. In the main part of the thesis, I will report three studies, two EEG and one fMRI study.

In the first study, I investigated the time course of the interaction of face, body, and vocal emotional information in a congruent setting. I measured the EEG signal while presenting video clips, either mute or accompanied by matching vocalizations, as well as while presenting the vocalizations without any dynamic visual information. By computing ERPs, I found early facilitation effects for multimodality as well as for emotionality. In addition, I performed a time–frequency analysis (see methods section) to monitor ongoing visual processing and to investigate oscillatory brain responses in multisensory processing. I observed strong similarities in processing between in the visual and audiovisual conditions.

To obtain closer insights into the neural substrates underlying the effects observed in the first study, I implemented a highly similar experimental design in an fMRI study. Emotional vocalizations elicited enhanced activations in the inferior frontal gyrus, the anterior cingulate cortex, and the amygdala, while emotional body expressions enhanced the BOLD response in the fusiform gyrus, the extrastriate body area, and the hippocampus. The combination of emotional information from both sources seems to be processed primarily visually, as shown by activity in the fusiform gyrus, while activations in auditory areas such as the STS/STG are reduced. These results hence confirm the results from the first EEG study, showing a strong visual influence on the processing of auditory emotional information.

As these first two studies provide clear evidence for an interaction between the involved modalities, I conducted a third study in order to differentiate between a mere *interaction* and an actual *integration* of multisensory information (as outlined in Section 1.1.2). To this end, I extended the design used in the previous two studies by adding an auditory as well as an audiovisual condition containing degraded auditory information. I was thereby able to investigate an audiovisual interaction under good as well as suboptimal sensory conditions as suggested by the inverse effectiveness criterion for multisensory integration. I recorded the EEG signal and observed stronger interaction effects under suboptimal sensory conditions, in the auditory N100 as well as the beta-band. This suggests that integration can be observed between body and voice information, and furthermore emphasizes the importance of using complementary analysis techniques, such as ERPs and time–frequency-analysis.

Following these studies, I will include a review article discussing the role of multisensory emotion perception and dynamic visual information (as well as task settings) in psychiatric patient populations.

In the following general discussion, I will discuss the results of the empirical studies in more general terms, as well as provide an outlook on open questions and possible future directions.

I will now briefly introduce the methods used in the following studies, namely electroencephalography, including the analysis of event-related potentials as well as time–frequency-analysis, and functional magnetic resonance imaging.

Chapter 2

Methods

2.1 Electroencephalography (EEG)

2.1.1 Biological Basis

Brain activity is electric activity. A signal is propagated from one neuron to the next by changes in the membrane potential, thereby creating an electric current. Potential changes at the post-synaptic neuron can last up to hundreds of milliseconds, which can result in a summation of the signals from large groups of neurons (S. Luck, 2005). In particular cortical pyramidal cells can thereby generate an electric field strong enough to be measured on the scalp (Westbrook, 2000). The recording of this signal via specialized electrodes is called *electroencephalography* (EEG). As the neuronal signal measured on the scalp is attenuated by interjacent tissue as well as the skull, the recorded EEG amplitudes are in the microvolt range.

EEG provides a direct measure of brain activity and thereby offers a temporal resolution in the range of milliseconds (S. Luck, 2005). The spatial resolution, however, is rather poor and limited by the number of electrodes used during the recording. Furthermore, the signal is measured on the surface of the head, while it is generated by one or, more likely, several sources in the brain. While numerous algorithms have been developed to localize the neural generators of a given EEG signal, the source localization remains challenging, as the same signal measured on the scalp can in principle be generated by any number of underlying sources. This problem is also known as the “inverse problem” (von Helmholtz, 1853).

2.1.2 Recording and Processing

The EEG signal is recorded via electrodes that are attached to the skull. Today, the electrodes are mostly implemented in a cap or a net to allow for easier and fast attachment as well as facilitate the arrangement according to a standardized electrode layout. The most commonly used layout is the 10-20 system and modifications thereof (Sharbrough et al., 1991). To decrease impedances (ideally below 5kΩ), an abrasive paste is applied to the head at the electrode positions, often in combination with a conductive gel.

Using this set-up, the EEG signal is recorded as changes in voltage arising from the difference between the recording electrodes and a references electrode. Typically, the recorded signal is amplified to allow for an accurate digitization, and an analog filter is applied to remove ultra-slow and very high frequencies, which are not of interest as they are not driven by sensory or cognitive processes. In the studies described later, a band-pass filter from DC to 140 Hz was applied.

In order to analyze the recorded signal, usually several preprocessing steps are taken. While the aim of recording EEG is to analyze brain activity, the recorded signal is often contaminated by unrelated artifacts unrelated to brain processes, such as eye blinks and muscle contractions. One way to counteract such artifacts is to instruct the participant during recording to blink only during certain periods and to sit in a comfortable and relaxed position while avoiding movements. Nevertheless, some artifacts often remain. Via detection algorithms and/or visual inspection, time periods containing artifacts can be marked and excluded from further analysis.

Furthermore, various filters can be applied in order to minimize drifts and other fluctuations that are not of interest.

The resulting signal can be analyzed in fundamentally different ways, two of which, event-related potentials and time–frequency-analysis, will be outlined in the following sections.

2.2 Event-Related Potentials (ERPs)

An event-related potential (ERP) is by definition any change in potential that is elicited by external or internal events (Otten & Rugg, 2005). Practically, these changes in potential are usually too small to be detected in a single trial against the ongoing background activity in any EEG signal. Hence, in a standard ERP experiment, several trials of the same type are recorded and averaged after the experiment, assuming that in an ideal case the background activity will cancel out or at least be reduced (S. Luck, 2005).

The averaged trials result in a waveform consisting of several negative and positive deflections, which are considered ERPs. Most ERPs are named according to their latency in relation to the eliciting event and by the direction of their deflection (positive or negative). An N100, for instance, is a negative deflection approximately 100 ms following an event. A different approach is to name ERP components according to their order of appearance, resulting in the N100 being called N1, as it is typically the first negative deflection (Otten & Rugg, 2005). By convention, ERPs are plotted “upside-down”, that is the negative scale up and the positive one down. Besides latency and deflection, an additional factor to distinguish ERP components is their topography, that is there distribution across the scalp.

Different ERP components are assumed to reflect different sensory and cognitive processes, and changes in deflection amplitude, latency, or topography can be considered when comparing experimental conditions. Early components, such as the N100, are interpreted as primarily driven by physical features of the stimulus, such as loudness. These components are also referred to as *exogenous components* (Otten & Rugg, 2005). In contrast, later components such as the LPC are associated with cognitive processing, and can be classified as *endogenous components*.

In the following, the three ERP components which will be relevant in the experiments described later on will be briefly discussed. Schematically, these ERP components are depicted in Figure 2.1.

2.2.1 Auditory N100

As introduced above, the auditory N100 is a negative deflection peaking around 100 ms after an event. It is primarily observed at fronto-central electrodes (Rosburg, Boutros, & Ford, 2008). It is elicited by the onset of a sound and assumed to result from the combined activation of various cortical areas, such as the primary auditory cortex and auditory association areas (Näätänen & Picton, 1987; Giard et al., 1994).

The auditory N100 indicates sensory processing of sound information, and is influenced by physical parameters of the sound, such as loudness (e.g. Harris, Mills, & Dubno, 2007) and

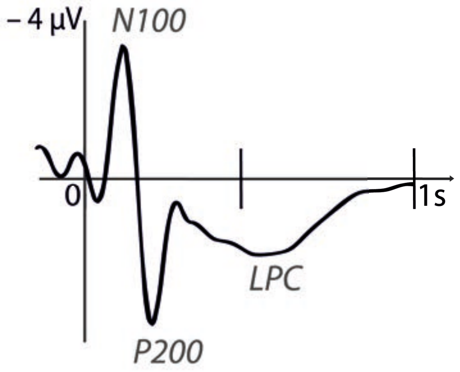

Figure 2.1: *Schematic Representation of ERP data.* The temporal unfolding of an ERP curve across 1 second after an event is depicted. First, a negative deflection, the N100 can be observed, followed by 2 positive peaks, the P200 and the LPC.

frequency (e.g. Dimitrijevic, Michalewski, Zeng, Pratt, & Starr, 2008), but also by attention (e.g. Hillyard, Hink, Schwent, & Picton, 1973).

Apart from the auditory stimulus features themselves, visual information can also have an influence on the N100 (e.g. Dekio-Hotta et al., 2009; Sugimoto, Nittono, & Hori, 2007). Emotional pictures, for instance, influence the N100 amplitude of concurrently presented neutral tones (Sugimoto et al., 2007). In particular, an N100 amplitude suppression can be observed if a sound is preceded by congruent, predictive visual information as it occurs in speech perception (van Wassenhove et al., 2005) but also in the perception of actions (Stekelenburg & Vroomen, 2007) and even in the observation of artificial stimuli (Vroomen & Stekelenburg, 2010). Furthermore, this amplitude reduction may also be accompanied by a reduction in peak latency (van Wassenhove et al., 2005). The predictive validity of the visual information appears to be of crucial importance for this mechanism, as no N100 changes are observed if the visual information does not reliably predict the sound onset (Stekelenburg & Vroomen, 2007; Vroomen & Stekelenburg, 2010).

Furthermore, the N100 seems to be modulated by the emotional quality of the stimulus, though the exact nature of such an influence is still under debate. G. G. Scott, O'Donnell, Leuthold, and Sereno (2009), for instance, report larger N100 amplitudes for negative compared to neutral low-frequency (i.e. rare in occurrence) words, but smaller N100 amplitudes for negative high-frequency words in contrast to neutral high-frequency words. Regarding multisensory emotion perception, Pourtois et al. (2000) show that mismatching information from emotional facial expressions and prosody results in a decreased N100 amplitude. Following a recent model by Schirmer and Kotz (2006), such early differences are associated with a first step in the processing of emotional prosodic information, namely the analysis of acoustic parameters (see also Kotz & Paulmann, 2011).

2.2.2 Auditory P200

Following the N100, a positive deflection called P200 can be observed. Though sometimes treated together as N100-P200-complex, the P200 can be differentiated from the N100 in several parameters, such as potential generators and response characteristics (Crowley & Colrain, 2004).

Similar to the N100, the P200 is affected by changes in physical stimulus parameters such as pitch (e.g. Wunderlich & Cone-Wesson, 2001) and loudness (e.g. Hegerl & Juckel, 1993). However, the P200 is associated with more elaborate, cognitive processing than the N100. Differences between neutral and emotional stimuli can reliably be observed, usually manifesting in larger P200 amplitudes for emotional compared to neutral stimuli (Sauter & Eimer, 2010; Paulmann et al., 2009). Furthermore, the auditory P200 is also affected by concurrent visual information, typically indicated by a reduction in amplitude (Stekelenburg & Vroomen, 2007).

In sum it is assumed that while the N100 is connected to the extraction of (emotionally) salient features, the P200 is assumed to indicate the integration of such cues (Paulmann & Kotz, 2008a).

2.2.3 Late Positive Complex (LPC)

As early components, both the N100 and the P200 are strongly related to sensory processing, combined with very early cognitive processes. Later components, such as the late positive complex (LPC), on the contrary, are related to later cognitive processing steps with little contribution of sensory aspects.

The LPC (sometimes also referred to as late positive potential; LPP) is a widely-distributed positive deflection usually starting after 500 ms. It is associated with the processing of (semantic) content of words (e.g. Herbert, Kissler, Junghöfer, Peyk, & Rockstroh, 2006; Kanske & Kotz, 2007) and pictures (e.g. Schupp et al., 2000; A. Keil et al., 2002). Perception of emotional stimuli, negative as well as positive ones, results in an increased LPC in comparison to neutral stimuli (e.g. Schupp et al., 2000; Herbert et al., 2006; A. Keil et al., 2002). Functionally, the LPC is associated with the activation of attentional and motivational resources, that are required to respond to emotional and highly arousing stimuli (Olofsson, Nordin, Sequeira, & Polich, 2008; Bradley & Lang, 2007; Hinojosa, Carretié, Valcárcel, Méndez-Bértolo, & Pozo, 2009). It will be therefore of interest whether the perception of non-linguistic dynamic, multisensory stimuli modulates the LPC.

2.3 Oscillatory Brain Activity

While ERPs are a common way to analyze EEG data, they capture only a certain aspect of the information contained in the signal. They can only reflect changes that are time-locked to a certain event (“evoked changes”), and thereby miss out on any change in the EEG that might be related to an event but varies in delay (“induced changes”). One way to asses such induced changes is the analysis of the oscillatory brain activity.

Like any continuous signal, the EEG can be decomposed into a number of sine waves that contribute to varying degrees to the observed signal. Depending on the sampling rate, the duration of the recording, and the analysis technique employed, such a decomposition can have different frequency resolutions. In principle, a frequency decomposition can be done by a classical Fourier transform. However, this would result in a loss of temporal information. A time–frequency-analysis in contrast retains, as the name indicates, time as well as frequency information. Below, the basic steps involved in a time–frequency-analysis will be roughly explained.

Traditionally, the oscillatory components of the EEG signal in humans are grouped into five different frequency bands: delta ($<$ 4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma ($>$ 30 Hz) (Buzsáki, 2006). However, these subdivisions are rather an orientation than fixed values, and different authors suggest slightly different frequency ranges. In addition, these bands can be further subdivided, for example into high and low beta (ranging from 13–25 and 25–35 Hz, respectively), as well as defined by topographical distribution and/or functional role, for instance mu rhythm and alpha rhythm (e.g. Perry, Troje, & Bentin, 2010), which both occur in the same frequency range, but are attributed different functional roles.

Each of these frequency ranges has been linked to various functions. In the studies conducted for my thesis, I focused on the alpha– and in particular the beta–range. Hence, I will, in the following, only discuss these two frequency bands in more detail.

One crucial feature of oscillatory brain activity is the possibility to assess changes in brain activity that are not time-locked. Another important aspect of brain oscillatory data is the fact that they can be analyzed along a number of dimensions. The aspect most akin to ERPs are changes in power (i.e. amplitude), that is, in the amount each frequency band contributes to the observed EEG signal. These power changes arise from the degree of synchronization or, conversely, desynchronization, of neural populations generating the respective frequency (Pfurtscheller & Silva, 1999). Therefore, increases in power are often termed synchronization, while a decrease in power is termed desynchronization.

Power changes, however, are only one way to analyze time-frequency data. Another approach is the analysis of changes in the phase in relation to an external stimulus. Note that phase-locking also provides a direct link to the observed ERP-components; a strong phase-locking around 10 Hz is necessary to observe an N100 response. Only trials where phase-locking occurs can contribute to the resulting ERP amplitude; by way of how ERPs are calculated, non-phase-locked aspects will be suppressed. However, analysis of phase-changes can go beyond ERPs, as it allows to detect more subtle shifts in phase that may occur at a larger distance from a stimulus, or in anticipation of a certain stimulus event.

In addition, changes in power and phase can be analyzed in combination and across frequency bands such as power–power coupling between two frequency bands or power–phase coupling (see Jensen & Colgin, 2007, for a brief overview). These analyses become particularly interesting, as they have been suggested as underlying mechanisms for multisensory integration (Senkowski, Schneider, Foxe, & Engel, 2008) as well as linking between different brain regions in feature-binding (Tallon-Baudry & Bertrand, 1999). As I will focus only on power changes in the analysis of my data, I will come back to these suggested roles of oscillations in the discussion of my thesis.

2.3.1 Time–Frequency Analysis

To analyze the oscillatory data, I used a wavelet analysis with Morlet's wavelets comparable to the method described by Tallon-Baudry and Bertrand (1999) and as implemented in the Field-

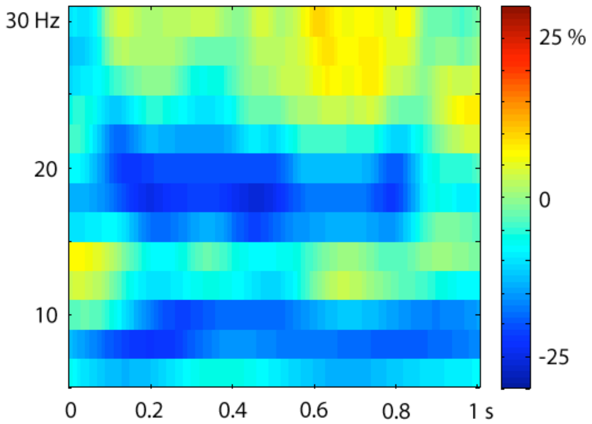

Figure 2.2: *Representation of Time–Frequency data.* As an example, a time–frequency power representation is shown over the course of 1 second (shown on the x-axis). On the y-axis, the different frequencies, ranging from 5–30 Hz, are shown. The color code denotes changes in power with respect to a baseline. While red colors code an enhancement in power, blue colors indicate a decrease. Hence, in the plot a power suppression in the alpha- and beta-range is depicted.

Trip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). For each trial, the EEG signal is convolved with a number of different wavelets, each representing a specific frequency. Crucially, a wavelet is characterized by a fixed ratio between temporal and frequency information. This implies that the duration of the wavelet is different for each frequency, as the same number of cycles extends over a longer duration for lower frequencies compared to higher frequencies. Thereby, a sine wave is generated for each frequency included in the analysis, which varies in amplitude (i.e. power) over time. A typical time-frequency power representation codes time on the x-axis, frequency on the y-axis, and power by a color code (see Figure 2.2 as an example).

2.3.2 Alpha Rhythm

Classically, the alpha-rhythm has been considered to signal a brain area at rest; occipital alpha is increased when a participant closes his eyes (and therefore does not process any visual information), and decreases in visual tasks (Palva & Palva, 2007). Recently, alpha has been ascribed an inhibiting function, emphasizing its role in actively suppressing activity in a specific brain area (Klimesch, Sauseng, & Hanslmayr, 2007; Jensen & Mazaheri, 2010).

First, in the context of the present studies, the role of alpha in mediating attention is particularly important. Changes in attention result in changes in alpha power (Marrufo, Vaquero, Cardoso, & Gómez, 2001; Ward, 2003). On the one hand, *bottom-up* visual attention can lead to a decrease in alpha power (Marrufo et al., 2001); on the other hand *top-down* attention can result in an increase in alpha power, signaling an inhibition of potentially distracting processes (Jensen, Gelfand, Kounios, & Lisman, 2002). Salient emotional signals that capture attention are therefore expected to result in a decrease in alpha power. Indeed, emotional words have been shown to decrease alpha-power compared to neutral words (Alfimova & Uvarova, 2008).

A second relevant aspect of oscillations in the alpha-range is the mu-rhythm. Already in 1954, changes in alpha-power were reported while participants watched different movie clips (Gastaut & Bert, 1954). Subsequent investigations have shown that suppression in the mu-rhythm is connected to the execution of movements as well as the perception of other people's actions and behavior (e.g. Cochin, Barthelemy, Lejeune, Roux, & Martineau, 1998; Muthukumaraswamy, Johnson, & McNair, 2004; Ulloa & Pineda, 2007; Perry, Troje, & Bentin, 2010). The mu rhythm can be distinguished from the alpha rhythm not only by its ascribed function, but also by the observed topographical distribution. While classical alpha, assumed to originate in occipital areas, can be observed primarily at posterior electrodes, the mu rhythm is generated in the rolandic area and hence observed more centrally (Perry, Troje, & Bentin, 2010). Besides mu rhythm in the alpha range, functionally similar oscillations (i.e. originating in rolandic areas and responding to action perception and execution) can also be observed at higher frequencies in the beta-range (see below).

In the studies presented here, I used short video clips depicting various motor actions. Hence, a suppression in the alpha-range is of particular interest not only due to its role in attentional mechanisms, but also as a potential indicator of action processing.

2.3.3 Beta Rhythm

Compared to oscillations in other frequency ranges, "[...] the functional role of beta-band oscillations at present seems to be least understood" (Engel & Fries, 2010, p.156). They have often been linked to various forms of motor processing. A suppression in the beta rhythm can typically be observed during the planning, preparation and execution of movement (Babiloni et al., 2002; Alegre et al., 2004; Grabska-Barwinska & Zygierewicz, 2006), while the beta rhythm is enhanced after the completion of a movement and in particular during the maintenance of a specific position, such as holding something (Kilner et al., 1999; Babiloni et al., 2002). Crucially, beta-power is also suppressed during the observation of biological motion (Babiloni et al., 2002; Muthukumaraswamy, Johnson, Gaetz, & Cheyne, 2006; Muthukumaraswamy & Singh, 2008). Interestingly, this decrease seems to be even larger during the perception of object-related movements in contrast to aimless, empty grip movements (Muthukumaraswamy et al., 2004).

Furthermore, the suppression in beta seems to be specifically linked to the perception of non-linguistic biological motion, as speech-unrelated mouth movements elicit a stronger decrease in beta-power compared to speech movements (Muthukumaraswamy et al., 2006). Alike to the alpha/mu-rhythm described above, the beta-rhythm is modulated by the perception and execution of action and is assumed to originate in rolandic areas. It therefore has also been termed “high mu”-rhythm (Pineda, 2005).

Besides its role in action perception and execution, an increase in beta-oscillations has also been implied in multisensory binding (A. von Stein, Rappelsberger, Sarnthein, & Petsche, 1999; Senkowski, Molholm, Gomez-Ramirez, & Foxe, 2006). For instance, an increased beta power correlates with a decrease in reaction time in an audiovisual decision task (Senkowski et al., 2006).

In sum, oscillations in the beta range are a particularly interesting candidate parameter for the present studies, as they have been implied in biological motion perception as well as multisensory integration.

2.4 Functional Magnetic Resonance Imaging (fMRI)

Electroencephalography offers a high temporal resolution, and thus a minute tracking of changes in brain activity. However, as mentioned above, the spatial resolution is poor. Functional magnetic resonance imaging (fMRI) offers a higher spatial resolution, typically between 2 and 3 mm, at the cost of a lower temporal resolution. The complementary use of both methods hence offers a good approach to obtain information about spatial as well as temporal characteristics of neural processing.

Compared to EEG, fMRI is a much younger technique. While first MR scans where acquired in the 1970s, functional MR scanning was only developed in the 1990s (Huettel, Song,

& McCarthy, 2004). Furthermore, in contrast to EEG, fMRI does not measure changes in brain activity directly. Rather, it measures changes in the oxygenation of the blood supplying different brain areas. As hemoglobin and deoxyhemoglobin differ in their relaxation time, the ratio between the two and thereby the blood-oxygenation can be determined. Given the assumption that an active brain area requires more oxygen and thus leads to an increased level of oxygen, one can indirectly infer that this brain region must be activated in a given process. Hence, a typical fMRI scan shows changes in the blood-oxygenation-level dependent (BOLD) signal.

This fact also explains the low temporal resolution; an adaptation of the blood-oxygenation level does not occur within milliseconds but rather within seconds. Therefore, the hemodynamic response, characterized by an initial peak and a subsequent undershoot, extends over a time-window of 10 seconds or more (Huettel et al., 2004).

2.4.1 Analysis of fMRI Data

Numerous data preprocessing steps are necessary before activations in specific brain areas become visible. Depending on the specific image acquisition technique and the experimental design, these steps can be done in varying order, and I will therefore focus on the order used for analyzing the data in the present fMRI study.

A number of separate slices are acquired, in the present study in horizontal orientation. As only one slice can be acquired at the same time, however, a certain temporal lag arises between the first and the last slice. This implies, that activity is not sampled at the same time everywhere in the brain. It is therefore necessary to do a *slice time correction*, which accounts for this fact and adjusts the recorded values according to their acquisition time.

One potential problem in the acquisition of fMRI are movements of the participant. If the participant moves within the scanner and thereby within the magnetic field, the same point within a given slice will correspond to a different location in the brain at different time points. One way to account for movement after the data acquisition is the *realignment* of the slices. However, this form of motion correction can only account for a certain amount of movement. If the participant showed very strong movement during scanning, his data cannot be used for further analysis.

The next step in preprocessing is the *corregistration* to a high-resolution structural image. In fMRI images, the image quality is rather poor compared to structural images, which have very high fine-grained resolution. Hence, a corregistration between these two images allows for the analysis of finer anatomical details.

Subsequently, the images of each participants are *normalized* to a standard space, such as the space established by the Montreal Neurological Institute (MNI space). Such a normalization is necessary, as the precise anatomical structure of the brain can largely differ between individual participants. To allow for comparisons between different individuals, and also between different studies and institutes, it is essential to establish a common reference frame.

Finally, a *spatial filtering* is employed to reduced noise in the data. In the spatial domain, such a filtering is commonly done using a Gaussian filter that reduces high-frequency components.

In an event-related design as employed in the present study, after the preprocessing steps the responses to specified events can be modeled using for instance a general linear model (GLM). As the steps described above, this is done for each participants. Finally, the data from several participants can then be averaged for statistical comparison.

Part II

Empirical Studies

Chapter 3

Material and Pilot Studies

3.1 Material

To investigate the multisensory perception of emotional information I created a new set of audiovisual video stimuli. The use of this designated set of stimuli offers several advantages for the controlled investigation of emotion perception. All actors recorded in the stimulus material are native German speakers and grew up in Germany, hence a possible confound of cultural background was avoided. Auditory and visual material was recorded simultaneously to ensure perfect congruency not only between the expressed emotional state, but also between the body movements and the generated sound. Non-emotional (or neutral) stimuli were matched as close as possible to the emotional stimuli with regard to motion content, vocalization, and duration, thereby minimizing the physical differences between emotional and non-emotional stimulus material.

The stimulus set consists of 540 audiovisual video clips, of which 180 were designated to depict "anger", 180 "fear", and 180 were intended to be neutral control stimuli. The recognition of these three emotional states was verified in several behavioral pilot studies (see below). Four semi-professional actors were invited to portrait the different emotional states: two women (24 and 41 years old) and two men (30 and 48 years old). Each actor wore a black t-shirt and blue jeans and was instructed to stand in front of a light-gray screen on a fixed spot in a neutral position. They were then asked to express the designated emotion by body, facial, and vocal expressions. After a short expression of the emotional state, they were instructed to return to the original, neutral position. As vocal expressions, "ah", "oh", and "mh" were used, as these interjections contain close to no semantic content and can therefore in a natural way be used to express anger as well as fear (unlike for instance “ih”/“eeh”, which is almost exclusively used in humans to express disgust).

Between 15 and 20 takes were recorded for each setting, that is, per emotion, interjection, and actor (i.e. 15 to 20 takes of anger expressed by actor A with “ah” as an interjection). In order to obtain a large variety of different emotion expressions the actors were instructed to express the emotion in different ways they thought suited that emotion best. For the non-emotional stimuli, actors were asked to produce a movement that contains as little emotion as possible, such as grooming gestures (scratching one’s head, adjusting a piece of clothing), stretching, or arm movement with no specific content. In addition, they were instructed to take a step in any direction, in order to increase the similarity to the emotional stimuli, which mostly contained a step back (fear) or forward (anger). Videos were recorded using a video camcorder (SONY HDVHC70 (HCV 1080i/Mini DV)) in HDV1080i quality, and sounds were recorded with a Zoom Handy Recorder H4.

Of the recorded stimulus material, I chose 15 items based on the video quality to be preprocessed and included in the pilot studies (see below). Preprocessing was done using Final Cut Pro 5 (Apple Inc.) for the video material and MATLAB 7.7.0 (The MathWorks Inc, Natick, MA, USA) for the sound files. The video material was cut, luminance was adjusted to be approximately equal for all videos, and the actor was centered in the middle of the picture. At the beginning of each video clip, a still picture of the respective actor standing in a neutral position was added and displayed for a duration of 520 ms. The sound files were digitized at 16-bit/ 44.1 kHz, cut, and normalized to –32 dB full scale.

Several physical stimulus parameters were analyzed for the stimulus material (see Table 3.1). Duration, pitch, and spectral center of gravity were computed using Praat (Boersma, 2001), and vocalization onset and motion content were computed using MATLAB. The vocalization onset in each recording was determined in a scripted way using changes in root mean square (RMS) amplitude of the video’s sound track; resulting onset values (in ms) were validated by auditory inspection. To quantify the motion content, I applied a procedure based on Pichon et al. (2008). Each frame of the video was converted to a gray-scale image using a luminance coding from 0 to 255. The difference in luminance between to consecutive frames was calculated. The sum of all pixels that revealed a difference larger than 10 (to account for noise in the recording)

Table 3.1: *Physical Stimulus Parameters.* Listed are physical stimulus parameters separately for the different emotions. Standard deviations are given in brackets. * = See text for the procedure used to quantify the motion content.

	anger	fear	neutral
Pitch in Hz	398.53 (76.44)	407.93 (52.29)	173.03 (36.28)
Spectral center of gravity in Hz	1118.05 (273.01)	1019.99 (313.40)	599.79 (244.93)
Sound onset in s.	1.03 (0.34)	0.86 (0.15)	1.62 (0.44)
Stimulus length in s.	4.48 (1.21)	4.66 (0.81)	3.96 (0.87)
Motion (total) in change per frame*	38.29 (4.46)	42.64 (4.82)	38.61 (4.92)
Motion (before sound onset) in change per frame*	36.65 (8.60)	40.28 (7.77)	39.57 (5.99)

was calculated and divided by the total number of pixels. This average change per frame was calculated for the entire video as well as for the period before the sound onset.

3.2 Pilot Study 1: Emotion Recognition from Auditory, Visual, and Audiovisual Stimuli

In order to verify that the emotions were recognized as the intended emotional state, a rating study was conducted including all 540 video clips. A total of 71 participants was invited and split randomly into three groups. An auditory group (N = 23, 15 women, mean age: 24.4, SD = 3.5), in which participants only heard the sound files, a visual group (N = 25, 13 women, mean age: 25.9, SD = 2.9), in which participants saw the mute video clips, and an audiovisual group (N = 23, 14 women, mean age: 23.9, SD = 2.2), in which participants were presented with video clips along with the matching sound. All participants were asked to classify the emotion in a 3-alternative forced-choice task (anger, fear, neutral) as well as rate the arousal on a 9–point SAM–scale (self-assessment mannequin, Lang, 1980), ranging from “not aroused at all” to “extremely aroused”. Participants were instructed to judge the emotion they thought the person in the video/the sound file was experiencing (i.e. not their own emotional state). The stimuli were presented on an overhead screen and via loud speakers and participants were tested in small groups. They gave their responses on paper by checking boxes indicating their arousal rating as well as the emotion they thought was expressed.

Emotions were recognized with an accuracy above 90 % in all modalities (with the exception of “neutral” in the visual condition, which was recognized at 89.80 %, see Table 3.2). As expected, arousal ratings were higher for emotional compared to neutral stimuli (larger than 5 for anger and fear stimuli, and smaller than 3 for neutral stimuli) (see Table 3.2).

As for each of the following main studies only a subset of these stimuli were used, accuracy values as well as statistical comparisons between these values for each subset are reported in the section describing that study.

3.3 Pilot Study 2: Emotion Recognition from Bodies

In the fMRI study and the EEG-study 2, faces in the video material were blurred in order to avoid any influence of the facial expression on the processing of body language. This material was created by using the software Motion (Apple Inc.). To ensure emotions were also recognized when no facial expression was present, a second pilot study was conducted using these videos (participants: N=16, 8 women, mean age: 23.87, SD = 3.32). In the pilot study, 360 items were included that were later used in EEG-study 2, and a subset of which was used in the fMRI study. This time, participants were tested individually; videos were presented on a computer screen and participants gave their response by pressing marked buttons on a prepared keyboard (Figure 3.1). Again, they were asked to judge the emotion in a forced-choice task, as well as rate the level of arousal the person in the video was experiencing. Videos were only presented mute, thus corresponding to the visual condition above.

Table 3.2: *Accuracy and Arousal in Pilot Study 1.* Accuracy values in % and arousal rated on a 9–point SAM–scale (1: “not aroused at all”, 9: “extremely aroused”) are listed separately for each emotion and each modality (audio, visual, audiovisual). Standard deviations are given in brackets.

	Accuracy		
	anger	fear	neutral
audio	91.45 (6.05)	92.53 (4.39)	94.23 (8.46)
visual	97.58 (2.72)	95.19 (4.99)	89.80 (11.72)
audiovisual	98.53 (1.66)	98.51 (1.47)	98.43 (2.16)
	Arousal		
	anger	fear	neutral
audio	6.43 (0.74)	6.06 (0.76)	1.96 (0.89)
visual	5.78 (0.78)	5.81 (0.88)	2.29 (0.93)
audiovisual	6.26 (0.95)	6.21 (1.04)	1.94 (0.70)

To analyze the results two repeated-measures ANOVAs were conducted; one on arousal ratings and one on accuracy values (see Table 3.3 for accuracy and arousal values). Both emotions received higher arousal ratings than neutral stimuli (anger vs. neutral: $F(1,15) = 72.83, p < .0001$, fear vs. neutral: $F(1,15) = 75.60, p < .0001$). Furthermore, emotional stimuli were recognized better than neutral stimuli (anger vs. neutral: $F(1,15) = 4.77, p = .045$, fear vs. neutral: $F(1,15) = 5.93, p = .028$). Nevertheless, neutral stimuli were also recognized with a probability of 93.9 %. No difference in either arousal or accuracy was observed between anger and fear stimuli.

Hence, the same pattern regarding arousal as well as accuracy can be observed for stimuli with as well as without facial expressions.

Figure 3.1: *Keyboard Layout for Arousal Ratings in Pilot Study 2.* Emotion categorizations were given likewise with keys labeled “Wut”, “Angst”, and “Neutral”.

Table 3.3: *Accuracy and Arousal in Pilot Study 2.* Accuracy values in % and arousal rated on a 9–point SAM–scale (1: “not aroused at all”, 9: “extremely aroused”) are listed separately for each emotion. Standard deviations are given in brackets.

	anger	fear	neutral
Accuracy	91.1 (0.7)	97.4 (0.7)	93.9 (1.5)
Arousal	5.71 (0.30)	5.59 (0.32)	2.10 (0.25)

3.4 Pilot Study 3: Emotion Recognition under Noise

In order to investigate inverse effectiveness (see Introduction) in EEG-study 2, the signal quality of the sound material was manipulated. Using MATLAB, the sound information was embedded in white noise with 5 different signal-to-noise ratios (SNRs): 0 dB, –6 dB, –12 dB, –18 dB, and –24 dB. An ideal signal quality for investigating perception under difficult conditions would result in a reduction in accuracy, but still allow for recognition above chance level. To determine the SNR best fitting these criteria, a third pilot study was conducted using 360 items, which were also used in EEG-study 2.

Fifteen participants (7 women, mean age: 25.73, SD = 4.01) were invited, who listened to all 360 stimuli at a computer via headphones. Stimuli were randomly assigned to one of the five SNRs, ensuring that an equal number of stimuli was presented at each SNR-level. Furthermore, it was ensured that, across participants, each stimulus was presented an equal number of times at each noise level. The task was identical to pilot study 2; the participants were asked to determine the emotion in a forced-choice task, rate the arousal on a 9–point SAM–scale, and gave their responses via a specially prepared keyboard.

To analyze the data, accuracy was compared at different noise levels using repeated measures ANOVA with the factors emotion (anger, fear, neutral) and noise level (0 dB, –6 dB, –12 dB, –18 dB, –24 dB). Note that values are Greenhouse-Geisser corrected if necessary (Greenhouse & Geisser, 1959).

Based on a significant interaction between the factors emotion and noise level ($F(3.61, 50.52) = 27.344, p < .001$), the accuracy in emotion recognition was computed separately for each noise level (see Table 3.4 for accuracy and arousal values). No difference in recognition accuracy between the emotions was observed for 0 dB, –6 dB, and –12 dB, while they differed significantly at –18 dB ($F(1.84, 25.82) = 22.302, p < .001$) and –24 dB ($F(1.52, 21.30) = 88.640, p < .001$). Furthermore, angry stimuli were recognized significantly worse at –6 dB ($t(14) = 2.33, p < .05$) compared to 0 dB, while no difference was observed for fearful and neutral stimuli. At –12 dB, however, stimuli in all emotion categories were recognized worse than at 0 dB (angry: $t(14)4.24, p < .001$; fear: $t(14) = 2.32, p < .05$; neutral: $t(14) = 2.88, p < .05$). Based on these results, –12 dB was chosen as the “high noise” level in EEG-study 2; a clear drop in performance was observed compared to the 0 dB/ “low noise” condition, yet recognition accuracy did not differ between the emotions at the –12 dB level. Furthermore, emotions were still recognized as intended with an average of 81.1 % (SD = 3.09), and a clear difference in arousal can be seen (Table 3.4) between emotional and neutral stimuli (i.e. emotional larger 5, neutral smaller 3).

As can be seen from the accuracy results (Table 3.4), a clear drop in performance occurred for –18 and –24 dB for emotional, but not for neutral stimuli. In fact, while the performance for angry stimuli at –24 dB even dropped below chance level, recognition accuracy for neutral stimuli did not drop below 78 % even at –24 dB. This at first glance puzzling observation can be explained by the response strategies of the participants. After the experiment, almost all participants reported to have pressed “neutral” in all cases in which they did not understand anything. Hence, neutral stimuli achieved a high accuracy rate on the cost of a high number of false positives, showing a clear response bias for neutral stimuli. Therefore, the actual performance of the participants is better reflected in measures taking into account the ratio between hits and false alarms, such as the perceptual sensitivity *d'* (MacMillan & Creelman, 2005). In EEG-study 2 and the fMRI-study, analysis of behavioral data will therefore focus on *d'* rather than on simple accuracy values.

Table 3.4: *Accuracy and Arousal in Pilot Study 3.* Accuracy values in % and arousal rated on a 9–point SAM–scale (1: "not aroused at all", 9: "extremely aroused") are listed separately for each emotion and each noise level. Standard deviations are given in brackets. The noise levels "0 dB" and "–12 dB" were included in EEG Study 2 as the low and high noise condition, respectively.

	Accuracy		
	anger	fear	neutral
0 dB	**91.79 (7.54)**	**86.59 (10.97)**	**87.93 (16.39)**
–6 dB	84.44 (11.96)	82.22 (11.56)	86.09 (18.64)
–12 dB	**73.44 (16.55)**	**79.17 (13.87)**	**81.57 (20.05)**
–18 dB	45.56 (18.59)	69.91 (20.68)	82.88 (12.34)
–24 dB	10.97 (7.98)	23.00 (12.37)	78.12 (18.44)
	Arousal		
	anger	fear	neutral
0 dB	**6.88 (0.69)**	**6.37 (0.95)**	**2.19 (1.11)**
–6 dB	6.25 (0.91)	5.87 (1.06)	2.02 (1.09)
–12 dB	**5.69 (1.30)**	**5.41 (1.09)**	**2.02 (0.86)**
–18 dB	4.33 (1.49)	4.82 (1.39)	1.97 (0.62)
–24 dB	2.09 (1.24)	2.51 (1.28)	1.87 (0.95)

Chapter 4

The Temporal Dynamics of Processing Emotions from Vocal, Facial, and Bodily Expressions[*]

S. Jessen[1,2] & S.A. Kotz[1]

[1]: Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1A, 04103 Leipzig, Germany

[2]: Cluster of Excellence "Languages of Emotion", Free University of Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany

Face-to-face communication works multimodally. Not only do we employ vocal and facial expressions; body language provides valuable information as well. Here we focused on multimodal perception of emotion expressions, monitoring the temporal unfolding of the interaction of different modalities in the electroencephalogram (EEG). In the auditory condition, participants listened to emotional interjections such as "ah", while they saw mute video clips containing emotional body language in the visual condition. In the audiovisual condition participants saw video clips with matching interjections. In all three conditions, the emotions "anger" and "fear", as well as non-emotional stimuli were used. The N100 amplitude was strongly reduced in the audiovisual compared to the auditory condition, suggesting a significant impact of visual information on early auditory processing. Furthermore, anger and fear expressions were distinct in the auditory but not the audiovisual condition. Complementing these event-related

---

[*] Jessen, S., & Kotz, S. A. (2011). The Temporal Dynamics of Processing Emotions from Vocal, Facial, and Bodily Expressions. Neuroimage, 58(2), 665-674.

potential (ERP) findings, we report strong similarities in the alpha- and beta-band in the visual and the audiovisual conditions, suggesting a strong visual processing component in the perception of audiovisual stimuli. Overall, our results show an early interaction of modalities in emotional face-to-face communication using complex and highly natural stimuli.

4.1 Introduction

We perceive our environment via multiple senses. Usually these senses provide congruent information, which leads to facilitated processing (e.g. Giard & Peronnet, 1999; Frassinetti, Bolognini, & Làdavas, 2002; Suied, Bonneel, & Viaud-Delmon, 2009). A situation in which such multimodal perception becomes particularly obvious and relevant is face-to-face communication, where information is conveyed via voice, facial expression as well as body language. One crucial information we can obtain this way is the current emotional state of a communicative partner.

Numerous studies have investigated the brain mechanisms underlying multimodal emotion perception from facial and vocal expressions (e.g. Dolan et al., 2001; Kucharska-Pietura, Phillips, Gernand, & David, 2003; Ethofer, Pourtois, & Wildgruber, 2006; Kreifelts et al., 2007), as well as the time-course of emotion processing (de Gelder et al., 1999; Pourtois et al., 2000; Pourtois, Debatisse, Despland, & de Gelder, 2002; Brosch et al., 2009; Paulmann et al., 2009). Electrophysiological data utilizing mismatching emotional information demonstrate that first interaction effects can be seen as early as 100 ms (N100) after stimulus onset (Pourtois et al., 2000). Considering studies that employ matching stimuli, a more adequate test of perception under natural circumstances, emotional differences and interaction effects between modalities can reliably be seen within 200 ms (P200) post-stimulus onset in the event-related potential (ERP) (Paulmann et al., 2009).

Also at later processing stages, ERPs are associated with emotional processing, for instance in the late positive complex (LPC) (Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000; Hajcak & Nieuwenhuis, 2006; Kissler, Herbert, Winkler, & Junghofer, 2009; Foti, Olvet, Klein, & Hajcak, 2010). The LPC has been linked to increased and sustained attention to salient stimuli (Kanske & Kotz, 2007; Foti et al., 2010).

Studies considering emotional body language have shown that similar to facial expressions, body language can be a reliable source of emotional information (Grèzes et al., 2007; van Heijnsbergen et al., 2007; Pichon et al., 2008; de Gelder, 2009; Pichon et al., 2009). Furthermore, body language is integrated with facial expressions (Meeren et al., 2005; van den Stock et al., 2007) as well as with emotional prosody (van den Stock et al., 2007, 2008). To investigate visual emotion perception it is therefore essential to include body language as well as facial expressions.

Similar to visual emotion sources, auditory sources encompass several different expression facets. In the current study, we focus on the perception of interjections (Bostanov & Kotchoubey, 2004; Belin, Fillion-Bilodeau, & Gosselin, 2008; Dietrich et al., 2008), which are spontaneous and reflex-like utterances containing prosody with minimal semantic content. As shown in previous studies, emotional prosody allows to clearly discriminate between different emotion expressions starting at the P200 (Ashley, Vuilleumier, & Swick, 2004; Paulmann et al., 2009; Sauter & Eimer, 2010). By using these very primitive utterances, we can investigate very early aspects of emotion processing, before any cognitive evaluation takes place.

Combining body language, facial expressions, and prosody therefore is a necessary and promising step in the investigation of emotional communication in an ecologically valid way.

Natural emotion perception is not only multimodal but also dynamic (Blake & Shiffrar, 2007). We do not perceive emotions from static photographs, but from a constantly changing stream of visual information, which impacts our emotion perception (Kilts et al., 2003; Sato et al., 2004; Carretié et al., 2009). When investigating multimodal emotion perception it is rather important that both auditory and visual information is dynamic as static visual information may produce an artificial mismatch between modalities. We avoided this mismatch by using dynamic visual stimuli.

More recently the analysis of ERPs has been complemented by the analysis of oscillatory brain activity. One great advantage of analyzing induced oscillatory activity is the fact that processes need not be precisely time-locked to an event. This is particularly relevant when utilizing natural dynamic visual stimuli that vary in the development of the movement, thus not providing a clear emotion recognition point. We were interested in comparing the processing of unimodal visual and audiovisual stimuli, and analyzing induced oscillatory activity allows us to do so.

The two frequency bands that have been most often associated with the processing of biological motion are the alpha- and the beta-band (8–13 Hz and 15–30 Hz, respectively). A suppression in these frequency bands has been robustly reported for motor imagery (e.g. Neuper, Scherer, Wriessnegger, & Pfurtscheller, 2009; Nakagawa et al., 2011), as well as for the perception of biological motion (e.g. Cochin et al., 1998; Fawcett, Hillebrand, & Singh, 2007; Muthukumaraswamy & Singh, 2008; Perry, Bentin, et al., 2010; Perry, Troje, & Bentin, 2010). Additionally, it has been suggested that beta-band synchronicity is modulated by the necessity to adapt to a changing environment (Engel & Fries, 2010). A strong desynchronization in the alpha-band, on the other hand, has been linked to increased attention and/or stimulus saliency (e.g. Marrufo et al., 2001; Ward, 2003). Both, adaptation to a changing environment as well as an increase in attention, are characteristic for emotional situations. In line with this assumption, several studies have reported changes in the alpha- or beta- frequency band related to the perception (e.g. Alfimova & Uvarova, 2008) or experience of emotions (e.g. Glauser & Scherer, 2008).

Based on these previous findings, we expected to find a stronger desynchronization in the alpha- and beta-band for emotional stimuli containing biological motion.

Furthermore, the computation of the oscillatory activity allows comparing the sum of all changes in the unimodal conditions to changes in the multimodal condition (Senkowski, Gomez-Ramirez, et al., 2007), as is suggested by the superadditivity criterion (e.g. Calvert et al., 2001).

Combining both analysis techniques, ERPs as well as time-frequency analysis, therefore allows us to investigate multimodal emotion processing in a more comprehensive way than either method alone could. While our focus in analyzing ERPs lies on exploring the influence visual information has on auditory processing, we investigate the auditory influence on visual processing in the oscillatory data.

4.1.1 Present Study

In the present study we focus on two emotional (angry and fearful) as well as a neutral state. We chose these emotions as they achieved comparable ratings with respect to valence and arousal, thus reducing a possible confound in the results. In order to provide participants with a complete, naturalistic visual input, we presented short video clips containing facial expressions as well as body language. These video clips were presented either without sound (visual condition) or with an emotionally congruent interjection (audiovisual condition). Furthermore, we also presented the emotional interjections without videos (audio condition).

4.1.2 Hypotheses

We expect facilitated information processing in the multimodal compared to unimodal conditions. This should affect the amplitude size of the N100. Furthermore, we hypothesize facilitated processing of emotional compared to neutral stimuli, in the N100, the P200, and the LPC components. We also expect to find a stronger suppression in the alpha- and beta-frequency range as emotional stimuli capture attention due to increased saliency. Additionally, we expect to find an interaction between modalities and the emotional quality of the stimulus. This should be seen in the N100 and P200 components as well as in different oscillatory activity.

4.2 Methods

4.2.1 Participants

Twenty-three native speakers of German (12 female) participated in the study. They were between 19 and 32 years old (mean age 24.7, SD = 3.1). All were right-handed and had normal or corrected-to-normal vision. They received financial compensation for their participation. The participants gave written informed consent and the experiment was approved by the ethics committee at the University of Leipzig.

4.2.2 Stimulus Material

4.2.2.1 Recording

We used video clips and recordings of interjections expressing fear, anger, or no emotion (neutral). In order to create the stimulus material, two actors (male, 48 years, and female, 41 years) were invited and asked to perform body language in a way they thought suitable for the respective emotion to be expressed. Furthermore, they had to simultaneously produce an interjection to express the emotion. For the recording, the actors were standing in front of a light gray screen on an indicated spot. All actors wore blue jeans and a black t-shirt. They were instructed to stand in a neutral posture, then perform the body movement, and in the end return to the indicated spot, assuming again a neutral body posture. For the current study, video clips displaying angry, fearful, and neutral body language were used as well as the interjections “ah”, “oh”, and “mh” in all emotional conditions. As the actors were encouraged to produce a variety of different movements to express the emotions, the precise body language used varies between the different stimuli. Movements often used to depict anger include clenching fists, making a step towards the camera, and raising the arms, while fear was often expressed by a backwards movement, bending over, and moving the arms towards the body. Neutral movements included grooming gestures and common speech accompanying gestures. To increase the similarity to the emotional videos, the actors were instructed to take a step in any direction also in the neutral videos. Numerous takes were recorded, of which the top 15 in each condition were chosen by the experimenters to be validated in a rating study (see below).

4.2.2.2 Preprocessing

The videos were recorded with a video camcorder (SONY HDV-HC70 (HCV 1080i/ Mini DV)) in HDV1080i quality, while sounds were recorded using a Zoom Handy Recorder H4. The sound files were digitized at 16-bit/44.1kHz, and normalized to -32dB using MATLAB 7.7.0 (The MathWorks Inc, Natick, MA, USA). The video files were cut and processed using Final Cut Pro 5 (Apple Inc.). Each video clip started with a still picture of the actor standing in a neutral posture, displayed for 520 ms. Furthermore, for each actor, a still picture of a neutral body posture was extracted from one of the videos to be used in the auditory condition.

4.2.2.3 Rating Study

All material was validated in a rating study to ensure that the intended emotion was well recognized. Auditory, visual, or audiovisual stimuli were presented to different groups of participants (audio: N=23, visual: N=25, audiovisual: N=23) (see Figure 4.1), who were asked to determine the expressed emotion in a forced-choice-task, as well as to rate the arousal the person in the video was experiencing on a 9-point SAM-scale (Lang, 1980). Emotions were recognized with an accuracy above 90% in all groups. The results of the rating study for the stimuli used in the EEG-experiment are summarized in Table 4.1.

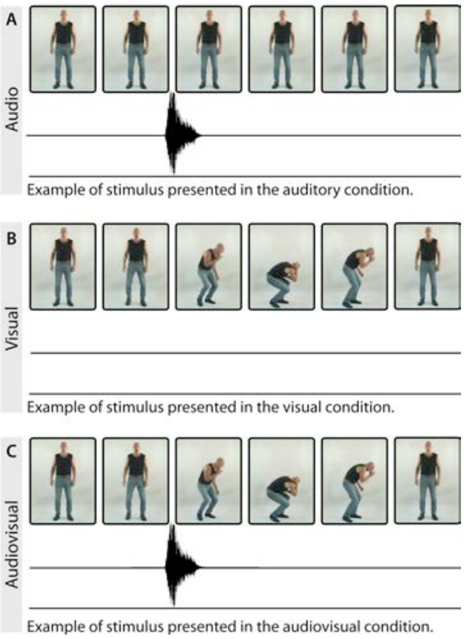

Figure 4.1: *Stimulus Material.* Example of the stimulus material used in the three different modality conditions. Depicted is the emotion "fear" displayed by a male actor.

4.2.3 Stimulus Selection

Of the validated stimuli, 12 items for each condition (interjections: "ah", "oh", "mh"; emotions: anger, fear, neutral) achieving the highest accuracy values were chosen for the current EEG-study, amounting to a total of 216 stimuli.

These stimuli were split into 3 sets; an audio, a visual, and an audiovisual set. For the stimuli assigned to the audio group, only the sound file was presented, accompanied by a still of the respective actor standing in a neutral posture. For the stimuli from the visual set, the video clip was presented without sound, and for the stimuli assigned to the audiovisual group, the video clips were presented with matching sound. For each participant, items were randomly assigned to the audio, visual, and audiovisual set, while ensuring the distribution of emotions, interjections, and actors was the same for all three sets.

Table 4.1: *Results of the Rating Study for the Stimuli used in the EEG Study.* The results for the different emotions are separated by modality (A=audio, V = visual, AV = audiovisual). Accuracy is given in %, with standard deviations in brackets. Arousal was measured using a 9-Point-SAM-Scale (Lang1980), with 1 being "not aroused at all" and 9 being "very aroused". Mean values (out of 9) are given with standard deviations in brackets.

	Accuracy		
	anger	fear	neutral
audio	96.12 (6.62)	98.50 (2.51)	95.72 (6.23)
visual	98.77 (2.38)	98.20 (2.30)	92.07 (7.18)
audiovisual	99.31 (1.71)	98.90 (1.95)	98.66 (2.68)
	Arousal		
	anger	fear	neutral
audio	7.03 (1.02)	6.72(0.0.93)	2.10 (0.44)
visual	6.21 (0.68)	6.24 (0.83)	2.52 (0.45)
audiovisual	6.76 (0.74)	6.75 (0.90)	1.95 (0.29)

4.2.4 Acoustic and Visual Stimulus Parameters

Visual as well as auditory information displaying different emotions by their very nature differ also in physical parameters.

In order to minimize the influence of loudness, all auditory stimuli were normalized to –32 dB. Furthermore, we computed the mean pitch for the auditory stimuli as well as the spectral center of gravity, the total duration of the video, and the delay between video- and sound-onset.

To control for differences in the visual stimulus material, we computed the amount of movement contained in the videos following a procedure described by Pichon et al. (2008). Using MATLAB 7.7.0 (The MathWorks Inc, Natick, MA, USA), for each video all frames where read in individually and converted to a gray-scale image, so that luminance was coded ranging from 0 to 255. Then the difference in luminance between two consecutive frames was calculated. For all pixels with a change larger then 10 (to compensate for the noise of the video recording), the difference was summed up and divided by the number of pixels to compute the average change. This was done once for the length of the entire video, and once for the part of the video before the sound onset. See supplementary material for details on the physical stimulus parameters as well as the results of the statistical comparisons between these parameters.

4.2.5 Procedure of EEG Study

After EEG preparation, the participants were seated in a dimly lit, sound-shielded chamber at a distance of approximately 120 cm to the monitor, on which the videos were presented with a size of 22 cm diagonal. The sounds were presented via headphones (SONY Stereo Headphones, MDR-XD100) at the same loudness level for all participants. After a training session consisting of 6 test trials, the actual experiment started.

Two tasks were assigned to the participants. In an explicit task, participants judged which emotion the actor was expressing in a forced-choice task with the options “Wut” (anger), “Angst” (fear), and “Neutral” (neutral). In the implicit task, participants judged the length of the stimulus, again responding in a forced-choice mode with either “kurz” (short), “mittel” (medium), or “lang” (long). Half of the participants responded with their left hand, while the other half was asked to respond with their right hand. Responses were given via a 4-button response box, adapted to either the left or the right hand. The button-assignment was balanced across participants, with “Neutral”/ “mittel” always being assigned to the middle-finger-button. Only the 3 buttons assigned to ring-, middle-, and index- finger were used. The 216 trials of the experiment were split into mini-blocks of 18 trials each, leading to a total of 12 blocks. The two tasks alternated between blocks, so that the implicit task was given for an entire block, followed by a block with an explicit task. Therefore, 108 trials were realized in the explicit task and the other 108 trials in the implicit task. The order of the tasks was counterbalanced across all participants. The distribution of modality (audio, visual, audiovisual), emotion (anger, fear, neutral), interjection (“ah”, “oh”, “mh”), and actor (actor1, actor2) was the same in both tasks. Between the blocks, participants could take a short self-defined break, and start the next block by pressing the middle-finger-button on the response box. Each block started with a start screen, stating the task to be performed in the upcoming block. After each trial, a question appeared on the screen, stating again the task (“Welche Emotion?” [“Which emotion?”]/ “Wie lang?”[“How long?”]) and the button assignment. The participant was given 3000 ms to answer. After that period or as soon as the participant responded, the question disappeared and a blank screen was presented for 2000 ms. After that time, a fixation cross was presented in the center of the screen for another 1000 ms before the next trial started. The experiment was implemented using the Presentation software (Neurobehavioral Systems, Inc.).

4.2.6 EEG Recording

Sixty-four Ag-AgCl electrodes were recorded according to the modified 10-20 system (Sharbrough et al., 1991), using the sternum as ground. The electrodes were mounted in an elastic cap (Electro- Cap International, Eaton, OH, USA). Impedances were kept below 5 kΩ, and the signal was bandpass-filtered online between DC and 140 Hz. As reference, the left mastoid was used. Brain Vision Recorder was used as recording software.

4.2.7 Analysis

4.2.7.1 Event-Related Potentials

The data was re-referenced offline to the linked mastoids, and all trials containing EOG artifacts above 30 μV were rejected automatically. The data were not baseline corrected, as the triggers were set at the auditory onset and hence several seconds after the onset of the visual stimulus, and therefore differences in the baseline might have arisen due to visual processing. Instead, we filtered the data with a bandpass filter of 0.5–30 Hz (based on Kissler et al., 2009; Hurtado, Haye, González, Manes, & Ibanez, 2009). Trials were averaged over a length of 0 to 2000 ms, starting at the sound onset. For the emotion recognition task, only trials that were answered correctly were included in the analysis.

For the analysis of the N100–P200-complex, the central midline as well as its surrounding electrodes were analyzed (F3, FZ, F4, FC3, FCZ, FC4, C3, CZ, C4, CP3, CPZ, CP4, P3, PZ, P4), as these ERP components are predominantly observed at central electrodes. For the LPC, only posterior electrodes were analyzed (OZ, O1, O2, PO3, PO4, POZ, P5, P6, PZ, P3, P4, CPZ) (based on Kissler et al., 2009). We selected three time-windows for analysis, 70–150 ms for the N100, 120–250 ms for the P200 (based on Stekelenburg & Vroomen, 2007), and 450–700 ms for the LPC (adapted from Huang & Luo, 2006). For these time-windows, we computed the mean amplitude. Besides differences in amplitude, we also compared the N100-latencies for the different conditions. To this end, we determined the time-point of the minimum in a time-window of 70–150 ms after sound onset.

For all four comparisons (N100-amplitude, N100-latency, P200-amplitude, LPC-amplitude), we computed repeated-measures ANOVAs with the within-subject factors modality (audio, visual, audiovisual), emotion (angry, fear, neutral), and task (explicit, implicit). Whenever necessary, a Greenhouse-Geisser correction was applied (Greenhouse & Geisser, 1959).

4.2.7.2 Time–Frequency Analysis

The data was re-referenced offline to the average reference, and again, all trials containing EOG artifacts above 30 μV were rejected automatically. A time-frequency-analysis was performed on a time-window of 0 to 1000 ms respective to the sound onset using the MATLAB toolbox FieldTrip (Oostenveld et al., 2011). Morlet's wavelets were used to estimate a time-frequency representation (Tallon-Baudry & Bertrand, 1999) with a time-frequency relation of m=7, offering a good relation between time- and frequency-resolution. We focused on the change in spectral power relative to a pre-stimulus baseline of 500 ms in the range of 8–13 Hz (alpha-band) and 15–25 Hz (beta-band). For the statistical comparisons, we focused on the time-range of 250–1000 ms after sound onset to avoid analyzing changes directly related to the physical properties of the sound onset. We computed a cluster-based permutation test with 1000 randomizations (Maris & Oostenveld, 2007), where a cluster was defined along the dimensions time, frequency, and electrode position. As a constraint, clusters had to extend across at least 3 adjacent electrodes. A type-1-error probability of less than .05 was ensured for all clusters. In planned comparisons, we computed Student's t-tests between (1) the different emotion-conditions for each modality condition separately and (2) for the different modality conditions irrespective of emotional content, and report the sum of the t-values of the electrodes and time-frequency-bins included in the cluster. Results of these planned comparisons were corrected for multiple comparisons using the Bonferroni-Holm-method (Holm, 1979), resulting in adjusted alpha-thresholds of 0.05, 0.025, and 0.017.

We also compared the sum of the changes in power spectrum for the unimodal to that of the multimodal condition. To this end the EEG responses of randomly drawn visual and auditory trials of the same condition (that is, same emotion and task-condition) were summed (Senkowski, Gomez-Ramirez, et al., 2007). We then proceeded with the time–frequency-analysis as described in the previous paragraph.

4.3 Results

4.3.1 Behavioral Results

As the task was designed to engage the participant's attention to the stimulus in general (implicit task) or the emotional content in particular (explicit task), we did not focus on the behavioral results. See Supplementary Material for accuracy rates in the emotion recognition task.

4.3.2 ERP Results

In the following, we will only mention the key findings, which we will also address in the discussion. A full list of all statistical comparisons can be found in the supplementary material.

4.3.2.1 N100 amplitude

A clear amplitude reduction for audiovisual compared to unimodal auditory stimuli was found in the N100 amplitude ($F(1,22) = 30.40, p < .0001$) (Figure 4.2).

When considering audiovisual and auditory stimuli separately, in both conditions a smaller amplitude can be observed for emotional compared to neutral stimuli (Figure 4.3). This holds true for the comparison between angry and neutral stimuli (auditory: $F(1,22) = 6.85, p < .05$, audiovisual: $F(1,22) = 18.34, p < .001$) as well as for the comparison between fearful and neutral stimuli (auditory: $F(1,22) = 29.45, p < .001$, audiovisual: $F(1,22) = 29.59, p < .001$). In the auditory condition we also observed an amplitude reduction for fearful compared to angry stimuli ($F(1,22) = 7.39, p < .05$), which was not found in the audiovisual condition.

4.3.3 N100 latency

For both, auditory and audiovisual stimuli, we report shorter N100 latencies for angry than for either fearful (in the auditory condition: $F(1,22) = 4.65, p < .05$, in the audiovisual condition:

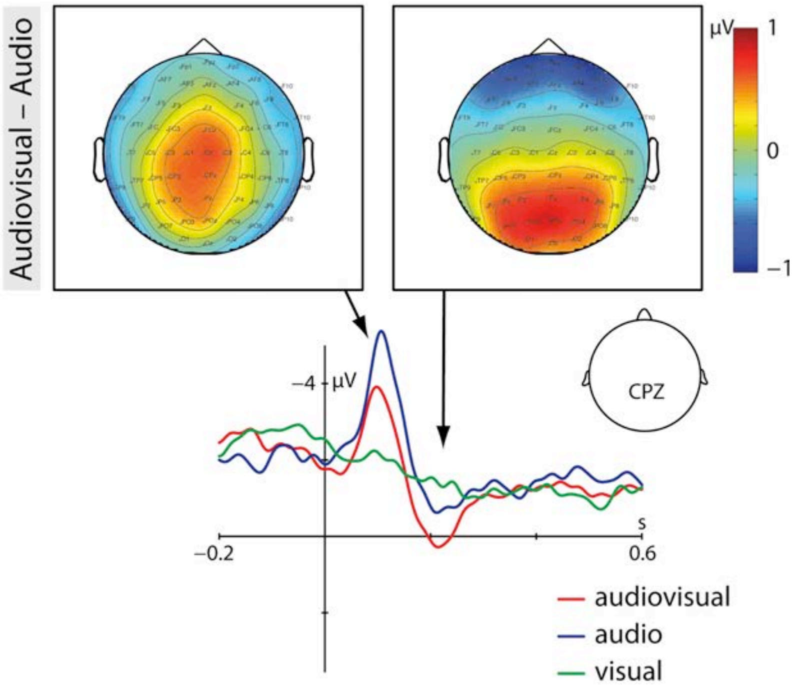

Figure 4.2: *ERP Response in the Different Modality Conditions.* In the top row, the difference between the audio and the audiovisual condition is depicted; the left head map shows an average for the N100 time-window analysis and the right head map displays the average for the P200 time-window analysis. The bottom row is showing the ERP response at the CPZ-electrode (green = visual, blue = audio, red = audiovisual). While in the N100 all three conditions differ significantly (p<.05), in the P200, the audiovisual condition elicited a larger amplitude than either the audio or the visual condition.

$F(1,22) = 5.25, p < .05$) or neutral (in the auditory condition: $F(1,22) = 4.65, p < .05$, in the audiovisual condition: $F(1,22) = 7.77, p < .05$) stimuli.

4.3.3.1 P200 amplitude

In the P200 amplitude, we found a similar pattern as in the N100 amplitude. Angry as well as fearful stimuli differed from neutral ones in both, the auditory (angry vs. neutral: $F(1,22) = 21.77, p < .001$, fearful vs. neutral: $F(1,22) = 4.95, p < .05$) as well as the audiovisual condition (angry vs. neutral: $F(1,22) = 14.75, p < .001$, fearful vs. neutral: $F(1,22) = 7.11, p < .05$). This time, however, smaller amplitudes were observed for neutral than for emotional

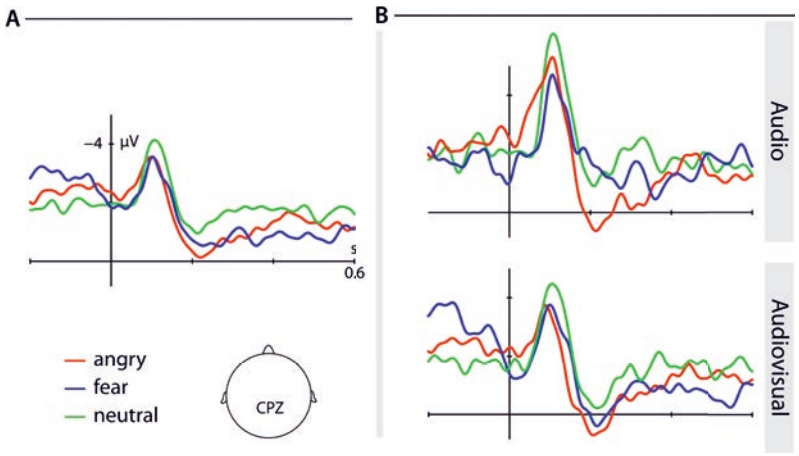

Figure 4.3: *ERP Response to the Different Emotion Condition.* In (A), the ERP response to different emotions is depicted (green = neutral, red = angry, blue = fear). In (B), the ERP response to the emotions is shown separately for the audio and for the audiovisual condition. While in both conditions a significant difference in the N100 and P200 is seen in both emotional expressions (compared to neutral ones), in the auditory condition we also report a difference between angry and fearful stimuli.

stimuli. Furthermore, we again observed a difference between angry and fearful stimuli in the auditory ($F(1,22) = 5.88, p < .05$) but not in the audiovisual condition.

Overall, a larger amplitude was observed for audiovisual than for auditory stimuli ($F(1,22) = 28.18, p < .001$).

4.3.3.2 LPC amplitude

Larger LPC amplitudes were observed for fearful stimuli in comparison to both, angry and neutral stimuli. An interaction with modality revealed that this was the case for audiovisual (fearful vs. neutral: $F(1,22) = 35.43, p < .0001$, fearful vs. angry: $F(1,22) = 21.90, p = .0001$) and visual (fearful vs. neutral: $F(1,22) = 28.69, p < .0001$, fearful vs. angry: $F(1,22) = 15.82, p < .001$), but not for auditory stimuli.

4.3.4 Time–Frequency Analysis

For the alpha- as well as the beta-band we contrasted the uni- with the multimodal conditions, and, separately for the different modality conditions, the three emotions expressed.

We compared the unimodal audio as well as the unimodal visual to the audiovisual condition in the alpha-band. While a large cluster was found in the right posterior hemisphere, showing stronger alpha-suppression in the audiovisual than in the audio-condition over the entire alpha-range and time-window ($p < .01, T_{sum} = 11425$), no difference was found between unimodal visual and audiovisual stimuli (see Figure 4.4). Comparing the sum of the unimodal conditions to the multimodal condition, we also found a significant cluster in left frontal regions between 8 and 10 Hz at a time-window of 300 to 700 ms ($p < .05, T_{sum} = 1914$), indicating stronger alpha-suppression in the multimodal condition than resulting from the pure sum of the two unimodal conditions.

When comparing the different emotional conditions split up by modalities in the alpha-band, no differences were observed in the audio or audiovisual condition but only in the visual condition. Here, the comparison between both, angry ($p < .001, T_{sum} = 8364$) and fearful ($p < .001, T_{sum} = 16723$), to the neutral condition showed significant clusters. In both cases, a broad cluster extending bilaterally at posterior electrodes was found between 300 and 700 ms, predominantly in the frequency range between 10 and 13 Hz (see Figure 4.5). A stronger alpha-suppression was seen for the emotional condition.

In the beta-band, for both, the audio and the visual condition, we observed clusters of significant differences compared to the multimodal condition. Compared to the auditory condition, a stronger beta-suppression over the entire time-window was found for the multimodal condition, extending over the entire scalp with a focus in posterior regions primarily between 16 and 23 Hz ($p < .001, T_{sum} = 49839$) (see Figure 4.4). For the visual compared to the multimodal condition, a very focal cluster was observed in the left hemisphere in the higher beta-band (19 Hz and above) extending from 250 to 450 ms ($p < .01, T_{sum} = 4619$). When comparing the sum of the audio and visual condition to the audiovisual condition, we observed a cluster very similar to that in the contrast between audio alone and audiovisual. While being largest at posterior electrodes, the cluster is extending over the entire head, covering the entire beta-range and time-window, and showing a stronger beta-suppression for the multimodal than for the summed unimodal conditions ($p < .01, T_{sum} = 29465$).

When considering only the auditory condition, a very focal right-anterior cluster was found in the comparison between fearful and angry stimuli, showing a stronger beta-suppression for

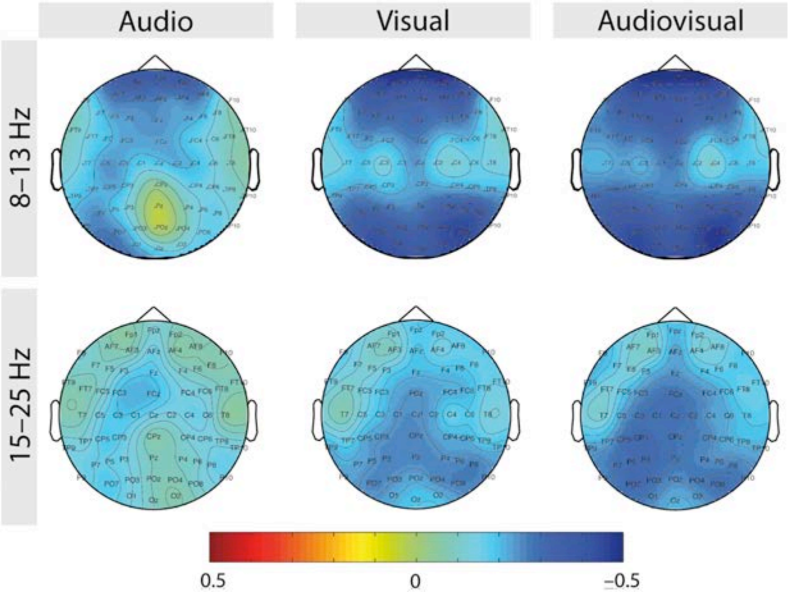

Figure 4.4: *Proportional Change in Power Spectrum for the Different Modalities Compared to the Respective Pre-stimulus Baseline.* Values are averaged across the frequency band displayed in a time-window of 250–1000 ms after sound onset.

angry stimuli between 300 and 500 ms in a low-beta-range of 15 to 22 Hz ($p < .01, T_{sum} = 1522$).

In the visual as well as in the audiovisual conditions, differences were found between both, fearful and angry stimuli, compared to the neutral stimuli. In all cases, a stronger beta-suppression is observed for the emotional than for the neutral stimuli. For visual stimuli, two significant clusters in the right hemisphere were found for the comparison between angry and neutral stimuli, one between 250 and 500 ms in a range of 16.5 to 24.5 Hz ($p < .001, T_{sum} = 5181$) and one between 650 and 900 ms in a range of 18.5 to 23.5 Hz ($p < .01, T_{sum} = 2362$). Comparing fearful and neutral stimuli, a cluster was found bilaterally in posterior regions between 15 and 23 Hz, extending from 400 to 1000 ms ($p < .001, T_{sum} = 19107$).

In the audiovisual condition, a very focal cluster was found in right posterior regions between 250 and 400 ms at 15 to 21.5 Hz, differentiating between angry and neutral stimuli ($p < .001, T_{sum} = 3724$). Comparing fearful and neutral stimuli, a similar focal cluster was

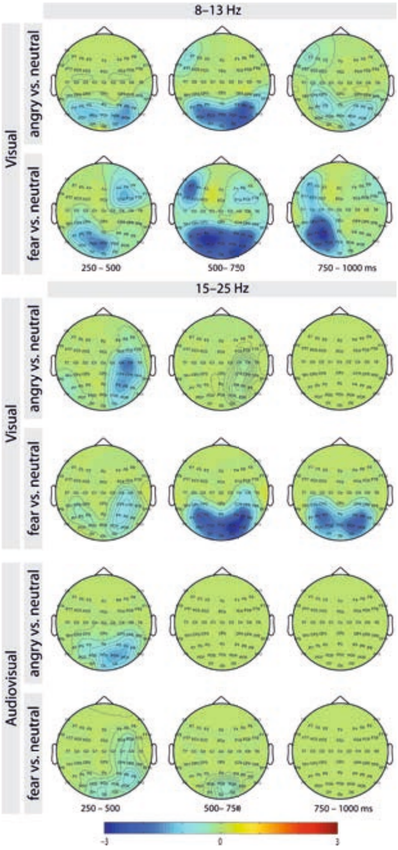

Figure 4.5: *Significantly Different Clusters between Emotional and Neutral Stimuli for the Different Modalities (all p<.05).* Plotted are t-values, averaged over the respective frequency band and in all three conditions.

found in the right hemisphere between 300 and 450 ms, primarily between 15 and 18.5 Hz ($p < .05, T_{sum} = 2416$).

Overall, we found a large difference between auditory and audiovisual stimuli in the alpha- and the band-band. Furthermore, differences between emotional and neutral stimuli were observed primarily in the visual and, to a somewhat smaller degree, in the audiovisual condition.

4.4 Discussion

Using complex stimuli including facial expression, body language, and vocal information, we investigated multisensory emotion perception in a setting closely approximating natural circumstances.

Participants were presented with emotional and neutral stimuli visually, auditorily, and audiovisually. In analyzing the EEG signal, we computed ERPs as well as the power spectra in the alpha- and beta-frequency bands.

As both, the visual and the auditory stimuli, contain highly reflex-like, non-speech information, we could observe very early differences in processing between emotional and neutral stimuli, and even among the expressed emotions. Furthermore, the use of two complementary analysis techniques allowed us to assess multimodal interaction processes as well as modulations by emotional content in the visual and the auditory domain.

In the following, we will first consider the effects of multimodal presentation irrespective of emotionality, then take a closer look at the emotion effects, and in the end briefly discuss the interaction between the two factors.

We investigated the processing of auditory information with or without prior predictive visual information in a paradigm similar to that of Stekelenburg and Vroomen (2007) and replicated an N100-suppression in the audiovisual condition (Stekelenburg & Vroomen, 2007; Vroomen & Stekelenburg, 2010), providing evidence for very early cross-modal influence also in emotion perception.

Extending these findings we also analyzed the differences in oscillatory activity in order to obtain more information about the impact of visual information. While the ERP-signal in the audiovisual condition clearly differs from the corresponding ERP-signal in the visual condition, no difference is found in the alpha-band, and only small differences in the beta-band.

Thus, while the time-locked ERP response seems to be strongly influenced by the onset of the auditory information, the ongoing oscillatory activity seems to be dominated by visual information, largely irrespective of the incoming auditory signal. Hence, the overall perception of audiovisual information seems to be mainly driven by visual information, while sounds per se may only provide additional, supplementary information.

Next to the comparison of the unimodal and the multimodal conditions, we also contrasted the sum of the unimodal to the multimodal condition, testing for superadditivity, which is assumed to indicate integration processes between modalities (e.g. Senkowski, Gomez-Ramirez, et al., 2007). Indeed we observed a stronger alpha- as well as beta-suppression in the multimodal condition than would be expected by summing up the unimodal conditions. This provides first evidence for an integration of the two modalities in the multimodal case.

However, the reliability of this criterion has recently been criticized (e.g. Laurienti et al., 2005; Beauchamp, 2005b; Ethofer, Anders, Erb, Droll, et al., 2006), and other criteria, such as inverse effectiveness (e.g. Stevenson & James, 2009; Holle et al., 2010) or the investigation of incongruency (e.g. Vroomen, Driver, & Gelder, 2001) have been suggested to provide a more comprehensive interpretation of integration processes. Therefore, our finding provides a starting point for the investigation of a possible integration process, and this topic needs to be clearly further addressed.

Contrasting emotional and non-emotional information, we observed differences in all modalities, in the ERP analysis in the audio and the audiovisual condition, and in the time–frequency-analysis mostly in the visual condition.

We observed a smaller N100-amplitude for emotional compared to neutral stimuli, which suggests facilitated processing of emotional compared to neutral stimuli (Paulmann et al., 2009). This effect is further strengthened by the latency reduction of the N100-peak for angry stimuli. Though clearly influenced by factors such as attention (Doallo, Cadaveira, & Holguín, 2007) and predictability (Dekio-Hotta et al., 2009), the N100 is often interpreted as a sensory component, and only few studies report modulations of the N100 amplitude in response to varying emotional content (e.g. Doallo et al., 2007; G. G. Scott et al., 2009). One reason why we observed differences in the N100 may be that prior visual information influences the response to auditory information. However, when considering the interaction between the factors modality and emotion, it becomes apparent that a differentiation between emotional and neutral stimuli is not only seen in the audiovisual condition, but also in the auditory condition, in which no prior emotional visual information was present. In fact, in the auditory condition, angry and fearful stimuli differ, an effect not seen in the audiovisual condition. Hence, it seems unlikely that any type of differentiation in the N100-amplitude can be solely attributed to the preceding visual information. However, one may argue that even in the purely auditory condition visual information was present, namely the still picture of a neutral body. This would provide congruent information in the case of neutral auditory information and incongruent information in the case of emotional auditory information. Indeed, Pourtois et al. (2000) observed a decrease in N100 amplitude for mismatching compared to matching audiovisual emotional information. However, as angry and fearful voices are equally mismatching compared to the neutral static information, this hypothesis cannot account for the observed difference between angry and fearful auditory stimuli in the audio condition.

Another plausible reason for the current finding is the use of interjections rather than highly complex pseudolanguage (Paulmann et al., 2009). Contrasting our study to studies using emotional vocalizations (Sauter & Eimer, 2010), our study differs in that we used a very limited number of emotions. Hence, when using primitive auditory stimuli in a simple context, differences in processing can be observed a lot early than in previous studies using more elaborated auditory stimuli and settings.

Unlike the N100 difference, emotion effects in the P200 range are often reported with larger amplitudes for emotional compared to neutral stimuli (Paulmann et al., 2009; Sauter & Eimer, 2010), and in some cases even among different emotions (Ashley et al., 2004). Our study replicates this finding. Furthermore, we report an interaction of emotion and modality. The P200-differences were more pronounced in the audio condition than in the audiovisual condition, in that, similar to the N100-effect, a difference between angry and fearful stimuli was found in the audio condition that was not present in the audiovisual condition. This suggests that in the auditory condition, in which no prior visual information is available and hence no prediction of the expressed emotion can take place, a more fine-grained distinction in processing is necessary and, in fact, takes place.

Another reason why the observed ERP-differences are unlikely to arise from purely sensory differences is that we observed differences in the LPC, a rather late component, which has been linked to evaluative processing and sustained attention rather than to sensory differences (e.g. Cuthbert et al., 2000). A larger amplitude for emotional compared to neutral stimuli has been reported for studies investigating the processing of emotional pictures (Cuthbert et al., 2000; Hajcak & Nieuwenhuis, 2006) and faces (Foti et al., 2010). Again, fearful stimuli clearly separate in processing from neutral as well as angry stimuli. Unlike the other ERP-components, we also observe a difference in the visual condition, as the LPC, which is associated with sustained attention to emotion stimuli, can extend over several hundred milliseconds (Moser, Hajcak, Bukay, & Simons, 2006).

Taking these findings together, differential processing of emotional and neutral stimuli can be found very early (N100, P200) as well as later (LPC) in the evolution of dynamic stimuli, suggesting an influence of emotional content on auditory processing at various processing stages (Schirmer & Kotz, 2006).

These findings are in line with recent studies investigating the neural structures underlying the processing of emotional voices. According to one of the recent models (Schirmer & Kotz, 2006) on vocal emotion expressions, emotional prosody is processed in three stages, namely sensory processing (primary auditory areas), integration of acoustic cues (superior temporal gyrus), and cognitive evaluation (inferior frontal gyrus (IFG) and orbitofrontal cortex (OFC)). Evidence for the influence of emotional content on early as well as late processing steps has been reported in several functional magnetic resonance imaging (fMRI) studies. An enhanced activation for emotional compared to neutral stimuli has been reported in early auditory areas (e.g. Ethofer, Ville, Scherer, & Vuilleumier, 2009; Grandjean et al., 2005) as well as in regions associated with higher level cognitive processing, such as the IFG/OFC (Ethofer, Kreifelts, et al., 2009) and the superior frontal gyrus (Fecteau et al., 2007).

Furthermore, several studies have sought to identify brain regions involved in the combination of emotional information from different modalities. One area often reported is the superior temporal sulcus (STS) (Kreifelts et al., 2009; Robins et al., 2009), but also areas such as the amygdala and the OFC respond to both, visual and auditory, emotional information (Sander et al., 2005; Wiethoff, Wildgruber, Grodd, & Ethofer, 2009; Pichon et al., 2009). However, as most of these studies used rather simple and static stimuli, the question arises which brain regions are involved in the processing of complex emotional stimuli.

Finally, we also investigated the influence of emotional content on the visual processing. In the alpha-band range, a stronger suppression can be observed for emotional than for neutral stimuli only in the visual condition. Alpha-suppression over occipital regions has been associated with an increase in visual attention (e.g. Marrufo et al., 2001; Ward, 2003), and hence it seems plausible that emotional, and thus highly salient stimuli, lead to increased attentional load.

Similar to the alpha -suppression, in the visual as well as in the audiovisual condition, we also observed a stronger beta-suppression for emotional compared to neutral stimuli. This finding may indicate an adaptation process in response to the perceived emotional information, as synchronization in the beta-band has been associated with the maintenance of the current state (Engel & Fries, 2010), while desynchronization is often found in action planning (Tzagarakis, Ince, Leuthold, & Pellizzer, 2010).

This interpretation is further supported by a study by Perry, Troje, and Bentin (2010), who report an enhanced alpha- as well as beta-suppression at occipital sites for the processing of socially relevant information from body language.

Overall, when considering how emotion perception is influenced by the multimodal presentation of information, it becomes apparent that there is stronger differentiation of emotions in the unimodal auditory as well as visual condition. This seems counterintuitive at first, as a facilitatory effect was expected in the case of multimodal perception. On second glance, however, this lack of differentiation in the multimodal case is precisely the type of evidence needed for facilitated processing and a strong interaction. In the audiovisual condition, the onset as well as the content of the auditory signal can be predicted by the preceding visual signal, as evident by the amplitude decrease in the N100–P200-complex, as well as by the strong emotional differences in the oscillation data in the visual condition. Hence, the auditory signal provides only limited new information, while in the unimodal condition, the auditory onset is unpredictable and the only source of information. Therefore, much more elaborate processing is required, which can be seen in the ERP-components as larger amplitudes as well as stronger differentiations between the emotion conditions.

The role of predictive visual information in multimodal perception has been investigated (Stekelenburg & Vroomen, 2007), but its role in the perception of dynamic and complex emotional stimuli has hardly been studied. While our results show that the processing of new auditory information can be greatly facilitated by preceding visual information, the exact parameters for such a facilitation are still unknown. Questions to be addressed in future studies thus include the role of a precise timing between the visual and the auditory signal, the reliability of the visual signal, and the degree of specificity for certain emotions.

4.4.1 Conclusion

In summary, complex visual information, consisting of body language and facial expression, strongly influences the processing of congruent auditory information under natural processing conditions. We were able to show that this interaction is influenced by emotional content at a very early point in processing, differentiating emotional from neutral stimuli after 100 ms. Hence, an interaction between facial and vocal, but also between numerous modalities seems to play an essential role in the perception of others' emotions. Furthermore, our results show that analyzing ERPs as well as time–frequency-data can provide valuable insights into different aspects of multimodal processing that could not have been obtained by employing only one analysis technique.

Acknowledgments

The authors would like to thank Cornelia Schmidt for help with data acquisition, Kerstin Flake for help with the graphical displays, and Jonas Obleser for technical advice.

4.5 Supplementary Material

See following pages for supplementary tables to EEG-Study 1.

Table 4.2: *Physical Parameters of the Stimulus Material.* Mean values with standard deviations in brackets. * = see main text for an explanation of the computation of motion contained in the video.

	anger	fear	neutral
Pitch in Hz	427.19 (61.21)	412.95 (51.47)	171.53 (40.61)
Spectral center of gravity in Hz	1180.56 (283.34)	1091.22 (318.26)	613.57 (181.81)
Sound onset in s.	1.21 (0.36)	1.14 (0.40)	1.53 (0.51)
Stimulus length in s.	4.94 (1.47)	5.06 (0.93)	3.73 (0.95)
Motion (total) in change per frame*	34.71 (3.83)	37.63 (3.72)	34.09 (4.57)
Motion (before sound onset) in change per frame*	39.15 (8.86)	38.70 (8.83)	37.81 (5.91)

Table 4.3: *Statistical Results for the Comparisons Between the Different Physical Stimulus Parameters.* n.s. = no significant difference. ***= p<.001.

	anger vs. fear	anger vs. neutral	fear vs. neutral
Pitch in Hz	n.s.	***	***
Spectral center of gravity in Hz	n.s.	***	***
Sound onset in s.	n.s.	***	***
Stimulus length in s.	n.s.	***	***
Motion (total) in change per frame*	***	n.s.	***
Motion (before sound onset) in change per frame*	n.s.	n.s.	n.s.

Table 4.4: *Results of the Emotion-Determination Task Used in the Experiment.* The results for the different emotions are separated by modality (A = audio, V = visual, AV = audiovisual). Accuracy is given in %, with standard deviation in brackets.

	Accuracy		
	anger	fear	neutral
audio	95.93 (1.55)	98.91 (0.60)	98.91 (0.60)
visual	89.49 (3.15)	97.83 (1.08)	99.28 (0.72)
audiovisual	98.91 (0.60)	97.83 (0.94)	98.91(0.60)

Table 4.5: *Results of the Statistical Comparisons of the ERP Results I.* Repeated measures ANOVAs were conducted with the factors emotion(3), modality(3), and task(2). Listed are all contrasts with significant results for the N100 amplitude and the N100 latency.

		N100 amplitude	N100 latency
Modality		$F(2,44) = 29.52, p < .001$	–
	audio vs. visual	$F(1,22) = 54.51, p < .001$	
	audiovisual vs. visual	$F(1,22) = 8.93, p < .01$	
	audio vs. audiovisual	$F(1,22) = 30.40, p < .001$	
Emotion		$F(2,44) = 31.11, p < .001$	
	anger vs. neutral	$F(1,22) = 25.37, p < .001$	
	fear vs. neutral	$F(1,22) = 77.50, p < .001$	
	anger vs. fear	$F(1,22) = 4.45, p < .05$	
Modality*Emotion		$F(4,88) = 5.46, p < .001$	$F(4,88) = 2.79, p < .05$
	Audio	$F(2,44) = 14.39, p < .001$	$F(2,44) = 3.43, p < .05$
	anger vs. neutral	$F(1,22) = 6.85, p < .05$	$F(1,22) = 4.65, p < .05$
	fear vs. neutral	$F(1,22) = 29.45, p < .001$	–
	anger vs. fear	$F(1,22) = 7.38, p < .05$	$F(1,22) = 4.65, p < .05$
	Audiovisual	$F(2,44) = 15.02, p < .001$	$F(2,44) = 4.78, p < .05$
	anger vs. neutral	$F(1,22) = 18.34, p < .001$	$F(1,22) = 7.77, p < .05$
	fear vs. neutral	$F(1,22) = 29.59, p < .001$	–
	anger vs. fear	–	$F(1,22) = 5.25, p < .05$
	Visual	–	–
Emotion*Task		$F(2,44) = 3.82, p < .05$	–
	Explicit	$F(1,22) = 20.68, p < .001$	
	anger vs. neutral	$F(1,22) = 10.94, p < .01$	
	fear vs. neutral	$F(1,22) = 52.92, p < .001$	
	anger vs. fear	$F(1,22) = 7.63, p < .05$	
	Implicit	$F(2,44) = 10.49, p < .001$	
	anger vs. neutral	$F(1,22) = 17.37, p < .001$	
	fear vs. neutral	$F(1,22) = 15.63, p < .001$	
	anger vs. fear	–	

Table 4.6: *Results of the Statistical Comparisons of the ERP Results II.* Repeated measures ANOVAs were conducted with the factors emotion(3), modality(3), and task(2). Listed are all contrasts with significant results for the P200 amplitude and the LPC amplitude.

		P200 amplitude	LPC amplitude
Modality		$F(2,44) = 17.47, p < .001$	$F(2,44) = 4.20, p < .05$
	audio vs. visual	–	$F(1,22) = 7.26, p < .05$
	audiovisual vs. visual	$F(1,22) = 20.33, p < .001$	$F(1,22) = 5.38, p < .05$
	audio vs. audiovisual	$F(1,22) = 28.18, p < .001$	
Emotion		$F(2,44) = 17.47, p < .001$	$F(2,44) = 15.15, p < .001$
	anger vs. neutral	$F(1,22) = 28.81, p < .001$	–
	fear vs. neutral	$F(1,22) = 15.36, p < .001$	$F(1,22) = 39.09, p < .001$
	anger vs. fear	–	$F(1,22) = 17.02, p < .001$
Modality*Emotion		$F(4,88) = 4.12, p < .01$	$F(4,88) = 10.31, p < .001$
	Audio	$F(2,44) = 10.83, p < .001$	–
	anger vs. neutral	$F(1,22) = 21.77, p < .001$	
	fear vs. neutral	$F(1,22) = 4.95, p < .05$	
	anger vs. fear	$F(1,22) = 5.88, p < .05$	
	Audiovisual	$F(2,44) = 7.47, p < .01$	$F(2,44) = 19.07, p < .001$
	anger vs. neutral	$F(1,22) = 14.75, p < .001$	–
	fear vs. neutral	$F(1,22) = 7.11, p < .05$	$F(1,22) = 35.43, p < .001$
	anger vs. fear	–	$F(1,22) = 21.90, p < .001$
	Visual	$F(2,44) = 4.73, p < .05$	$F(2,44) = 15.81, p < .001$
	anger vs. neutral	–	–
	fear vs. neutral	$F(1,22) = 5.37, p < .05$	$F(1,22) = 28.69, p < .001$
	anger vs. fear	$F(1,22) = 9.11, p < .01$	$F(1,22) = 15.82, p < .001$
Emotion*Task		–	–

Chapter 5

Differential Audiovisual Interactions for Emotional and Non-emotional Information *

S. Jessen[1,2] & S.A. Kotz[1]

[1]: Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1A, 04103 Leipzig, Germany

[2]: Cluster of Excellence "Languages of Emotion", Free University of Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany

Humans easily perceive and integrate body and voice expressions. This process is also influenced by emotional saliency. However, to date the brain's joint processing of emotional information from body and voice expressions is not yet well understood. We conducted a functional magnetic resonance imaging (fMRI) study in which participants were presented with dynamic audiovisual stimuli of emotional body expressions and accompanying emotional vocalizations. Activations implied in emotion processing were modulated in the respective unisensory control conditions: the inferior frontal gyrus, the anterior cingulate cortex, and the amygdala were activated more for emotional than neutral vocal expressions; the fusiform gyrus, the extrastriate body area, and the hippocampus were activated more for emotional than neutral body expressions. A conjunction analysis yielded a multisensory network consisting of the bilateral superior temporal gyrus (STG), the thalamus, and the cerebellum. With respect to joint

---

*based on Jessen, S. & Kotz, S. A. (submitted to Cerebral Cortex). Differential Audiovisual Interaction for Emotional and Nonemotional Information.

effects of emotional and multisensory processing, however, the dynamic audiovisual emotion stimuli elicited increased activation in “downstream” areas of the visual central pathway (right fusiform gyrus), but led to decreased activations along the auditory central pathway (bilateral STS/STG). The current data speak for multiple pathways of multisensory processing dependent both on the emotional saliency and the relative importance of audiovisual information. Multisensory emotional processing seems characterized by an influence of predictive visual information on auditory processing, rather than an a balanced interaction between both modalities.

5.1 Introduction

Emotion perception is an essential part of everyday communication. We can readily determine whether someone is angry or happy and adapt our behavior accordingly. To do so, we can utilize various channels of information, such as facial, vocal, and bodily expressions. While numerous studies have investigated the interplay between facial and vocal emotion expressions, the interaction between body and vocal information has largely been neglected. Nevertheless, both, body and vocal expressions, offer reliable information regarding someone’s emotional state (Atkinson et al., 2004). Additionally, body expressions are particularly important for the perception of emotions at large distances and the link between emotional experience and an intended action (de Gelder, 2009). In the present study we therefore investigated the neural mechanisms underlying the interplay between body and vocal expressions, focusing in particular on ecologically valid, dynamic visual information.

The perception of emotional information from body as well as vocal expressions in isolation has been addressed in several previous studies. De Gelder and colleagues (Hadjikhani & de Gelder, 2003; de Gelder et al., 2004; Grèzes et al., 2007; Pichon et al., 2008, 2009) have delineated various brain regions involved in the perception of emotional body expressions, including visual body areas (e.g. fusiform gyrus, extrastriate body area) and areas implied in more general, modality-independent emotion processing (e.g. amygdala, insula) (Pichon et al., 2009). They suggest an interplay between several brain networks combining fast reflex-like responses as well as more evaluative processing involving sensorimotor and frontal areas (de Gelder, 2006b).

While emotional body expressions are quite clearly defined, emotional vocal expressions subsumes a rather heterogenous field of potential stimuli. One approach is to investigate the perception of emotional vocalizations by combining semantically neutral sentences or pseudo-language with different affective pronunciations (Kotz et al., 2003; Ethofer, Anders, Erb, Herbert, et al., 2006; Ethofer, Pourtois, & Wildgruber, 2006; Kreifelts et al., 2007). While such stimuli have a rather strong linguistic aspect, at the other end of the continuum vocalizations such as laughter or crying have been used (Fecteau, Armony, Joanette, & Belin, 2005; Fecteau et al., 2007). An intermediate approach is the use of interjections such as “ah” and “oh” (Dietrich et al., 2008). These very simple utterances contain close to no semantic information, naturally express different emotional states, and yet allow for certain acoustic control of stimulus properties. Comparable to emotional body expressions, emotional vocal expressions elicit increased BOLD responses in voice specific areas (e.g. superior temporal sulcus and gyrus, STS/STG) as well as cortical (e.g. ventral anterior cingulate cortex, ventral ACC) and subcortical areas (e.g. amygdala, insula) associated with the more elaborate processing of emotional information (Grandjean et al., 2005; Beaucousin et al., 2007; Dietrich et al., 2008).

The interaction between auditory and visual emotional information has been investigated in previous studies, focusing on facial expressions and spoken words. Typically, multisensory emotions elicit increased activation in areas involved in multisensory perception in general, such as the STG and the thalamus (Ethofer, Pourtois, & Wildgruber, 2006; Kreifelts et al., 2007; Park et al., 2011). These activations are not only observed in multisensory *emotion* perception; the same areas are also involved in the integration of auditory and visual information in general (Sekiyama, Kanno, Miura, & Sugita, 2003; Werner & Noppeney, 2010b). Hence, emotional content seems to result primarily in an activation increase in areas associated with multisensory processing in the first place.

However, most previous studies on multisensory emotion perception relied on stimulus material that differs drastically from real life expressions. Many studies use static visual information (Dolan et al., 2001; Pourtois, Gelder, Bol, & Crommelinck, 2005; Ethofer, Pourtois, & Wildgruber, 2006; Müller et al., 2011; Park et al., 2011), which is unnatural and has shown to elicit strong processing differences when compared to dynamic visual information (LaBar et al., 2003; Sato et al., 2004; Yoshikawa & Sato, 2006; Trautmann et al., 2009). Furthermore, predictive visual information (i.e. visual information containing accurate information about the onset and the content of the following auditory information) preceding the auditory onset is essential for an optimal integration between the two modalities (Stekelenburg & Vroomen, 2007; Vroomen & Stekelenburg, 2010). Others used semantically neutral words spoken with affective prosody (Ethofer, Pourtois, & Wildgruber, 2006; Kreifelts et al., 2007; Robins et al., 2009), creating a potential conflict between semantic and prosodic information, as suggested in work investigating an interaction between prosodic and semantic content (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b).

The main aim of the present study was therefore to investigate the interaction between auditory and visual emotional information by controlling for these possibly confounding factors. Hence, we used dynamic videos containing emotional body expressions as visual stimuli. By using comparably long stimuli (on average above 4 s) and a naturally occurring delay between the auditory and the visual onset (on average larger than 0.6 s) we were able to investigate the influence of preceding and ongoing visual information on auditory processing. Additionally, we used interjections, such as “ah” and “oh”, which allow investigating voice perception with only a minimal semantic influence (Dietrich et al., 2008).

Previous studies have investigated the influence of predictive visual information on auditory processing in cross-modal priming paradigms. While priming tasks differ strongly from the setting employed in the present study (i.e. the use of two subsequent but separate stimuli as compared to one video, in which sound naturally follows visual information) they may nevertheless yield valuable insights on the influence of prior information has on the succeeding one. For example, sounds incongruent with a preceding picture (compared to a congruent setting) result in enhanced activation of the middle temporal gyrus (MTG)/STG and several frontal areas, such as the inferior frontal sulcus and the medial prefrontal cortex (Noppeney, Josephs, Hocking, Price, & Friston, 2008). Likewise, semantic priming with affective stimuli results in a decreased activation in the bilateral STG for congruent pairs (H. Liu, Hu, Peng, Yang, & Li, 2010). Hence, predictability of stimuli seems to result in a decreased activation compared to a violation of prediction.

Using only congruent stimuli, we were interested in the influence of emotional content on the multisensory interaction of dynamic information. Previous studies suggest that an interaction between auditory and visual information may be stronger for threatening compared to neutral stimuli (Maier, Neuhoff, Logothetis, & Ghazanfar, 2004; Cappe, Rouiller, & Barone, 2009). If indeed congruent audiovisual information leads to a decrease in STG activity, and audiovisual interaction is enhanced for emotional stimuli, this would predict reduced STG activation for emotional compared to neutral audiovisual processing.

Regarding the audiovisual processing of emotional information, two different and opposing outcomes are therefore possible. (i) A combination of visual and auditory information resulting in enhanced activation for emotional compared to neutral stimuli in particular in the STG. (ii) Temporally leading visual processing attenuates the auditory response, in particular for emotional stimuli, resulting in a reduced activation for emotional compared to neutral stimuli in the STG.

The unimodal perception of emotional information is expected to result in decreased activations in networks previously suggested for the respective modality, that is the visual body areas and amygdala for body expressions, and STS/STS and anterior cingulate cortex (ACC) for vocal expressions.

5.2 Methods

5.2.1 Participants

Eighteen native speakers of German (8 female) participated in the experiment. Their mean age was 25.9 (standard deviation (SD) = 3.9). All were right-handed, had normal or corrected-to-normal vision, and reported no hearing impairment. They gave written informed consent and were compensated financially for their participation. The study was approved by the local ethics committee.

5.2.2 Stimuli

To investigate the multisensory perception of emotional information from complex material, we created a stimulus set consisting of short video clips showing one of four semi-professional actors (two female) portraying anger, fear, and a non-emotional (neutral) state in vocal and body expressions. Each actor was standing on an indicated spot in front of a gray screen and was instructed to express the intended emotion in a way he/she thought best fitted. Synchronized with the body motion, he/she was instructed to express the emotion verbally by using “ah”, “oh”, or “mh”. See Figure 5.1 for an example of the stimulus material and Jessen and Kotz

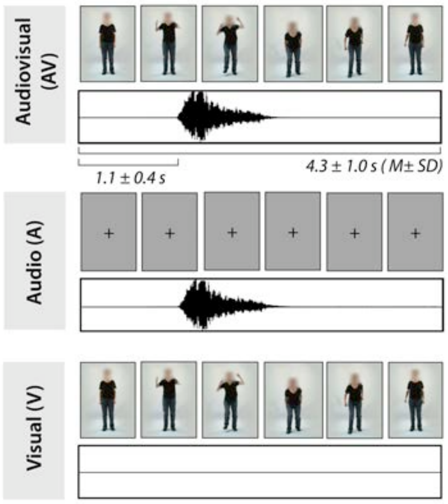

Figure 5.1: *Stimulus Material.* Shown are 6 frames of a video clip showing a female actor depicting anger. In the audio condition, only the sound information was presented along with a fixation cross. In the visual condition, a mute video clip was shown, and in the audiovisual condition, the video clip was paired with the matching interjection.

(2011) for a further description of the stimulus set. The voice onset naturally occurred with average delay of 627 ms (SD = 431) with respect to the body motion onset. As each video started with a 500 ms still frame of the respective actor standing in a neutral position, this resulted in a delay of 1127 ms (SD = 431) to the video onset.

In order to avoid an influence of facial expression on the perception of body language, the faces of the actors were blurred using the software Motion 3.0.2 (Apple Inc.), rendering the faces unrecognizable to the participant. A pilot study with 16 participants (who did not take part in the present study) confirmed that the intended emotions were recognized from bodies only with an accuracy of 96.2 % (SE (standard error) = 0.8).

5.2.3 Design

In the present study, we used 216 stimuli of the stimulus set described above. For each participant, an equal number of stimuli was pseudo randomly assigned to an auditory, a visual, or an audiovisual condition, resulting in 72 items in each modality. Furthermore, we ensured that in each modality, an equal number of stimuli depicted "anger", "fear", and "neutral", resulting in 24 items per modality and emotion. Actor (female 1, female 2, male 1, male 2) and interjections (“ah”,“oh”,“mh”) were counterbalanced across the emotion conditions.

In the auditory condition, only the sound file was presented, while a fixation cross was displayed in the center of the screen. In the visual condition, a mute video clip was shown, and in the audiovisual condition, a video clip was presented along with the matching sound file. Additionally, we included 24 events, in which a fixation cross was presented for 8 seconds as a baseline condition. After the presentation of each stimulus, the participants had to categorize the emotion displayed in a forced-choice-task using a four-button response box (only the central three buttons were used in this experiment). The participants were instructed to give a response within one second, after which an inter-trial-interval was triggered by the next scanner pulse. The duration of the inter-trial-interval was jittered and varied between 1500 and 3000 ms during which a fixation cross was presented.

Half of the participants responded with their right hand, while the other half responded with the left hand. Furthermore, the button assignment was counterbalanced across participants, with “neutral” always being assigned to the middle button and “anger” and “fear” to the left and right button in half of the participants and vice versa in the other half.

The average duration of one stimulus was 4.29 s (SD = 1.02) and the total scanning time approximately 36 min. Trials were presented in random order. The experiment was implemented using the Presentation software package (Neurobehavioral Systems, Inc.).

Participants completed a short training outside the scanner, in which seven stimuli additional stimuli were presented.

5.2.4 Data Acquisition

Imaging data was acquired at 3.0 T using a Bruker MedSpec 30/100 system (Bruker Medizintechnik GmbH, Ettlingen,Germany). Visual stimuli were presented via a projector (SANYO PLC-XP50L) on a screen outside the scanner, which the participants were able to see via a mirror. Auditory stimuli were presented via specialized headphones. We acquired 32 horizontal slices every 2 seconds parallel to the plane intersecting the anterior and posterior commissure (AC-PC plane) using an in plane resolution of 3 x 3 mm and a slice thickness of 3 mm (gap 0.5 mm). Slices were collected interleaved using a T2* weighted gradient echo-planar imaging sequence (repetition time (TR) = 2 s, echo time (TE) = 25 ms, field of view (FOV) = 192 mm, flip angle = 90¡, image matrix = 64 x 64, acquisition bandwidth 100 kHz). Additionally, high resolution T1-weighted MR scans that had been acquired before were used for normalization during preprocessing.

5.2.5 Data Analysis – Behavioral Data

In order to assess the participants' ability to discriminate angry, fearful, and neutral stimuli in the different modalities, we computed the perceptual sensitivity, *d'* (MacMillan & Creelman, 2005), comparing separately anger and fear, anger and neutral, as well as fear and neutral. On these values, we computed a repeated–measures ANOVA with the factors modality (audio, visual, audiovisual) and emotion comparison (anger–fear, fear–neutral, anger–neutral). Greenhouse-Geisser corrected values are reported were necessary (Greenhouse & Geisser, 1959). Student's t-tests were computed for the analysis of the individual contrasts.

5.2.6 Data Analysis – Imaging Data

Data analysis was performed using SPM 8 (Wellcome Department of Imaging Neuroscience, London, UK) and Matlab 7.7.0 (The MathWorksInc, Natick, MA, USA).

After performing slicetime correction, realignment, and unwarping, images were realigned using the middle slice (31st slice) as a reference slice. Images were then co-registered to high resolution images, segmented, and normalized to the Montreal Neurological Institute (MNI) standard space (Evans et al., 1992). Spatial smoothing was applied using an 8-mm Gaussian full-width-half-maximum kernel (Friston et al., 1995).

At the first level, the sound onset was selected as the event onset in the auditory and audiovisual conditions. In the visual condition, the time point of sound onset in the matching sound file was chosen as the event onset. Events were modeled using a stick function convolved with the hemodynamic response function. A 128 Hz highpass filter was used to remove slow signal drifts.

At the second level, we performed a series of one-tailed t-test. To investigate the processing of emotional information in the different modalities, changes in BOLD signal in response to emotional (angry and fearful) stimuli were compared to neutral ones separately for each modality. Additionally, we also compared angry and fearful stimuli separately to neutral ones. Clusters of activation below a threshold of $p < .001$ (uncorrected) are reported if the cluster size exceeds $k = 54$ voxels resulting in a significance level of $p < .05$ corrected for multiple comparison at the whole brain level (based on Slotnick, Moo, Segal, and Hart (2003)).

Furthermore, we computed a conjunction null analysis (Nichols, 2005) to investigate the interaction between auditory and visual stimuli ($AV > V \cap AV > A$). We computed this analysis once collapsed across all emotion conditions, and once separately for emotional (angry and fearful) and neutral stimuli. Again, we report clusters of activation below a threshold of $p < .001$ (uncorrected). We report all clusters exceeding $k = 10$, though only cluster with a size of $k > 54$ reach a significance level of $p < .05$ corrected for multiple comparison on a whole brain level.

5.3 Results

5.3.1 Behavioral Results

Stimuli in all conditions were recognized with an accuracy higher than 60 %, the condition with the lowest recognition accuracy being fearful stimuli in the auditory condition (60.8 %). A significant effect of modality in the analysis of the d' values ($F(1.95, 31.24) = 12.11, p < .0001$) revealed that stimuli were discriminated worse in the auditory than the visual ($t(16) = -7.89, p < .0001$) as well as in the audiovisual condition ($t(16) = -15.39, p < .0001$) (see Figure 5.2). Furthermore, visual stimuli showed lower discrimination accuracy compared to audiovisual stimuli ($t(16) = -4.25, p < .001$). While participants showed no difference in discrimination ability between the different emotional states for visual and audiovisual stimuli, clear differences were found in the auditory condition ($F(2.06, 32.90) = 18.34, p < .0001$). Participants were worst at discriminating anger and fear stimuli, better at discriminating fear from neutral stimuli, and showed the best performance in the discrimination of anger from neutral stimuli (anger–fear vs. fear–neutral: $t(16) = -2.98, p < .01$, fear–neutral vs. anger–neutral: $t(16) = -2.81, p < .01$).

5.3.2 Imaging Results

5.3.2.1 Emotion Contrasts for Vocal Expressions

The perception of vocal emotion expressions compared to neutral ones resulted in larger activation increases in the left posterior ACC/ preSMA, the left IFG, as well as the right cerebellum

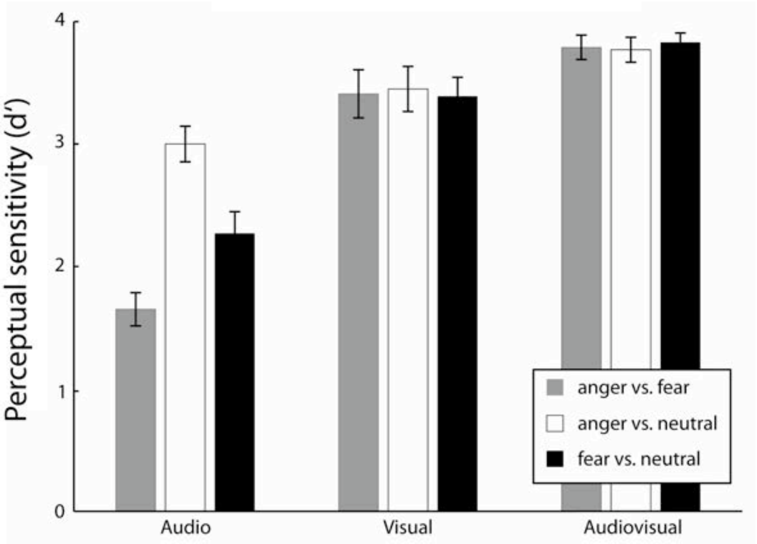

Figure 5.2: *Perceptual Sensitivity in the Different Modalities.* The discrimination ability measured as d' was computed for the comparison between all emotions, that is anger vs. fear, fear vs. neutral, and anger vs. neutral. Participants performed significantly worse in the auditory condition, in particular for the discrimination between anger and fear.

(crus I) (see Figure 5.3 and Table 5.1). With the exception of the cerebellum, the same regions were also activated in the separate contrasts of anger compared to neutral and fearful compared to neutral vocal expressions. In contrast, smaller activations for emotional compared to neutral vocal expressions were observed in the right fusiform gyrus. This difference seems to arise primarily from the contrast between fear and neutral expressions, in which a decrease for emotion expressions was also observed in the left inferior occipital lobe and the right supramarginal gyrus. We found no difference in activation for neutral compared to anger expressions.

5.3.2.2 Emotion Contrasts for Body Expressions

Contrasting visual emotional and neutral body expressions (Figure 5.3 and Table 5.2) enhanced activations were observed in various occipital regions, the right fusiform gyrus, as well as the left medial cerebellum. Similar activations in occipital areas were also found in the separate contrasts between anger and fear expressions compared to neutral ones. Anger stimuli addi-

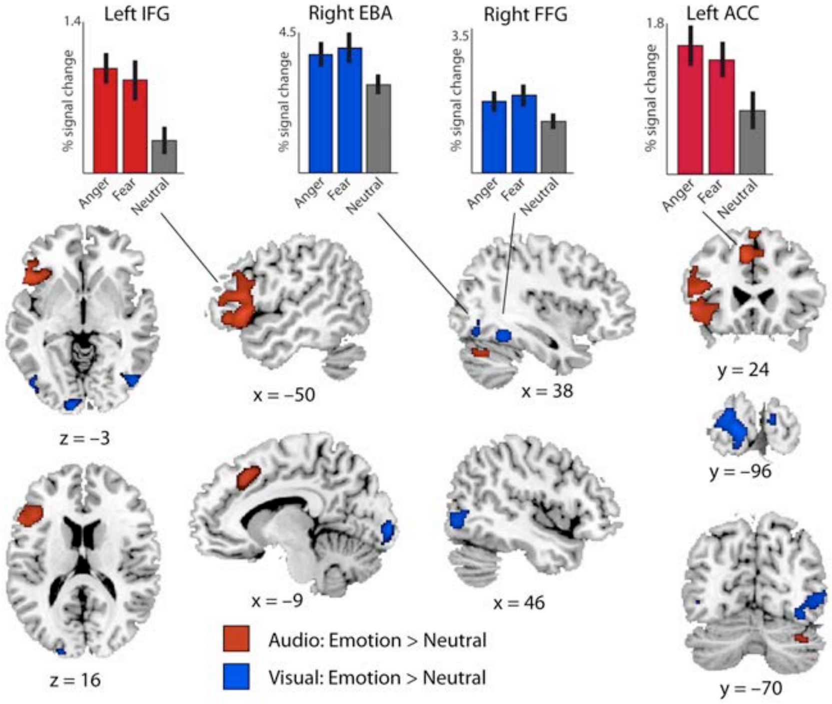

Figure 5.3: *Unimodal Activations for Emotion Processing.* For both, vocal and body expressions, a number of brain areas showed an enhanced activation for emotional in contrast to neutral stimuli. Activations are depicted at a threshold of p<.001 (uncorrected), and both emotions are compared against neutral stimuli. In the auditory condition (red), we observed an enhancement in the right inferior frontal gyrus, and the anterior cingulate cortex. For visual stimuli (blue), an increase in activation was found in the fusiform gyrus as well as various occipital areas. The bar graphs show % signal change in an sphere with a radius of 3 mm around the peak voxel of each cluster. IFG = inferior frontal gyrus, EBA = extrastriate body area, FFG = fusiform gyrus, ACC = anterior cingulate cortex.

tionally elicited enhanced activation in the right amygdala. We found no significant decrease in activation for emotional compared to neutral body expressions.

5.3.2.3 Emotion Contrasts for Audiovisual Stimuli

In the audiovisual condition, we observed enhanced activations for emotional compared to neutral stimuli only in the right fusiform gyrus. In emotion specific contrasts, stronger activations for anger compared to neutral stimuli were found in the left hippocampus, while no increased

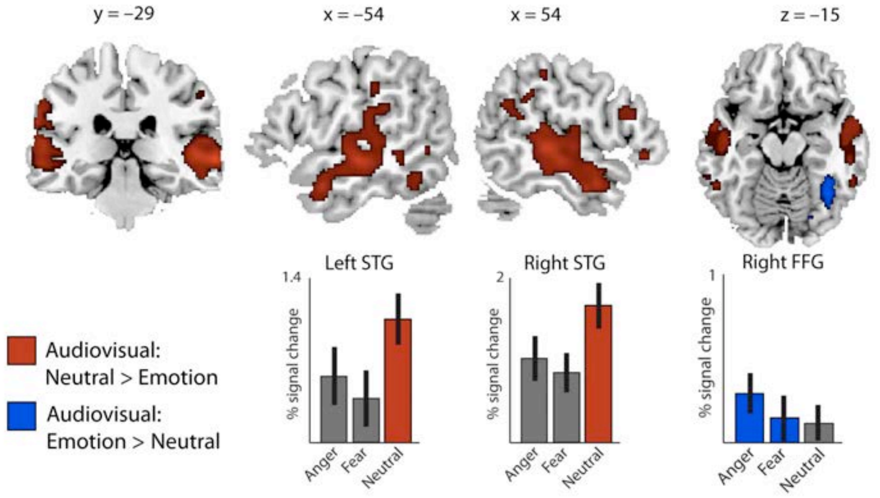

Figure 5.4: *Activation Pattern for Audiovisual Stimuli.* Collapsed across anger and fear stimuli, a broad deactivation was found in the bilateral STG/STG compared to neutral stimuli. In contrast, a stronger BOLD response for emotional compared to neutral stimuli is observed in the right fusiform gyrus. Activations are shown at a threshold of p<.001. The bar graphs show % signal change in an sphere with a radius of 3 mm around the peak voxel of each cluster. STG/STS = superior temporal gyrus/sulcus, FFG = fusiform gyrus.

activation was found for fear compared to neutral stimuli. A broad decrease in activation was observed for emotional compared to neutral audiovisual stimuli in the bilateral STS/STG (see Figure 5.4 and Table 5.3). Additionally, decreased activation was found in the right inferior orbitofrontal cortex (OFC) and the right inferior frontal pars triangularis. When contrasting neutral stimuli to anger and fear stimuli separately, a similar decrease was observed in the bilateral STS/STG and the right inferior frontal pars triangularis. Additional deactivations were observed in the right supramarginal gyrus, the left inferior temporal lobe, and the right superior frontal lobe when contrasting neutral and anger stimuli. When contrasting neutral and fear stimuli, additional deactivations were furthermore found only in the right inferior OFC.

5.3.2.4 Conjunction Analysis

Results of the conjunction analysis ($AV > V \cap AV > A$) for both, emotional and neutral stimuli combined, reveals enhanced activation for the audiovisual compared to either unimodal

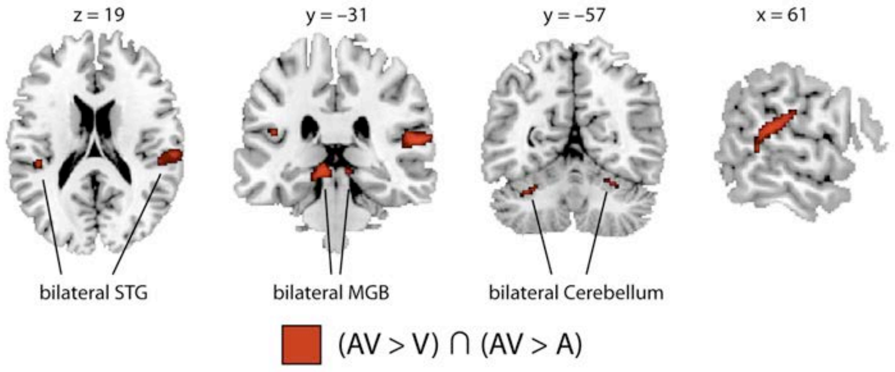

Figure 5.5: *Conjunction Analysis* $AV > V \cap AV > A$ The conjunction analysis was computed across all conditions, combining anger, fear, and neutral stimuli. An increased activation for audiovisual stimuli compared to either auditory and visual stimuli was found in the STG, the MGB, and the cerebellum. Activations are shown at a threshold of p<.001. The bar graphs show % signal change in an sphere with a radius of 3 mm around the peak voxel of each cluster. STG = superior temporal gyrus, MGB = medial geniculate body.

condition in the bilateral STG, bilateral thalamus, and bilateral cerebellum (6) (Figure 5.5 and Table 5.4). However, only the activations in the right STG and the left thalamus had a cluster size larger than $k = 54$. The conjunction analysis for emotional stimuli revealed larger activation for audiovisual than for auditory or visual stimuli only in the right STG. For audiovisual neutral stimuli, we observed enhanced activations in the bilateral STG, bilateral thalamus, and the bilateral cerebellum, as well as the right inferior frontal operculum, and the right precentral gyrus.

5.4 Discussion

To investigate the neural basis underlying multisensory emotion perception from body and vocal expressions, we presented participants with short video clips depicting angry, fearful, or non-emotional body expressions along with matching voice expressions.

In the unimodal auditory and visual conditions we observed activations in regions typically associated with emotion processing in the respective modalities, such as the ACC and the inferior frontal gyrus (IFG) in the auditory modality, and the fusiform gyrus, the amygdala and various occipital regions in the visual modality. However, simultaneous perception of information from body and vocal expressions revealed a rather different picture. While enhanced activation was found for emotional information in the right fusiform gyrus, a strong decrease in activation was observed in the bilateral STS/STG, as well as the right inferior OFC and the right IFG.

A similar picture emerged in the conjunction analysis investigating the integration of body and vocal information. In the neutral condition, we found bilateral activations in the STG, the thalamus, as well as the cerebellum, a network that has been suggested previously as a key player in multisensory integration. For emotional stimuli, however, we only observed activation in the right STG.

In the following, we will first discuss the findings observed in the unimodal emotion perception. We will then discuss the pattern of results observed in the audiovisual condition with respect to multisensory emotion perception.

5.4.1 Emotional Bodies and Voices Activate Separate Networks

For both, fear and anger body expressions, an enhanced activation was observed in several occipital areas. This confirms reported results on the perception of emotional body language (de Gelder et al., 2004; Grèzes et al., 2007; Pichon et al., 2008, 2009). Anger stimuli also increased activation in the amygdala, while enhanced activation in the fusiform gyrus was observed for fear compared to neutral stimuli. Both areas have been observed in numerous studies contrasting the perception of emotional with neutral body expressions (Hadjikhani & de Gelder, 2003; de Gelder et al., 2004; Grèzes et al., 2007; Pichon et al., 2008, 2009).

The amygdala has been implied in processing highly salient and arousing stimuli across modalities, mediating fast and adaptive responses (Adolphs et al., 1999; LeDoux, 2000; Adolphs, 2002a; Gläscher & Adolphs, 2003; Sander, Grafman, & Zalla, 2003; Sergerie, Chochol, & Armony, 2008). In contrast, the fusiform gyrus has been linked to the processing of face and body information (Peelen & Downing, 2005; Kanwisher, McDermott, & Chun, 1997), and previous studies have reported a modulation of such processing by affective content (Vuilleumier et al., 2001; Peelen & Downing, 2007; Sabatinelli et al., 2011; Kret, Pichon, Grèzes, & de Gelder, 2011).

Another area specifically implicated in the processing of body expressions is the extrastriate body area (EBA, Downing, Jiang, Shuman, & Kanwisher, 2001; Downing et al., 2006; Peelen & Downing, 2007)) As we did not use a localizer in the present study, we were not able to precisely determine the EBA at an individual level. However, note that the peak coordinate observed for enhanced activation in response to emotional compared to neutral body expressions in the right MTG [50 -74 -2] closely coincides with the peak coordinates reported for the EBA by Downing et al. (2001) [51 -71 1] and Kret et al. (2011) [52 -70 -2], suggesting that the observed area may indeed be the EBA. An increase in activation for emotional compared to neutral stimuli is in line with recent studies showing that threatening content modulates the activation observed in the EBA (Peelen et al., 2007; Kret et al., 2011; Atkinson, Vuong, & Smithson, 2011).

Our findings fit well with previous results and suggest an extensive network of modality specific as well as higher cognitive areas involved in the perception of emotional body expressions.

In the auditory condition, enhanced activation for both, anger and fear stimuli, was observed in the left dorsal ACC as well as the left IFG. Activation in the ACC is commonly observed for emotional compared to neutral voice perception (Buchanan et al., 2000; Sander et al., 2005; Dietrich et al., 2008) as are enhanced activations in the IFG (Kotz et al., 2003; Wildgruber et al., 2005; Sander et al., 2005; Ethofer, Pourtois, & Wildgruber, 2006; Ethofer, Kreifelts, et al., 2009).

Based on previous studies on emotion perception (Kanske & Kotz, 2011; Etkin, Egner, & Kalisch, 2011) we expected an increase in activation primarily in the ventral ACC. The dorsal ACC on the contrary has rather been implied in the reappraisal of emotional, particularly negative, information (Etkin et al., 2011) and the preparation of adaptive responses to unexpected outcomes (Shackman et al., 2011; Alexander & Brown, 2011). Furthermore, increased activation in the dorsal ACC has also been observed for increased task difficulty involving emotional stimuli (Davis et al., 2005) as well as for conflict resolution (Kanske & Kotz, 2011; Mayer et al., 2011). While the observed activation in the dorsal ACC therefore conflicts with our initial hypothesis, it is plausible when considering the behavioral performance. Participants overall showed a low discrimination ability in the auditory condition, however, they performed worst at discriminating anger and fear stimuli. Therefore, uncertainty was highest for these stimuli, as was the number of incorrect trials. Hence, the observed activation increase for emotional stimuli can presumably be attributed to increased uncertainty about the perceived stimulus.

The observed activation therefore suggests evaluative processing of auditory emotional information in order to generate an adequate behavioral response, that is, to determine the perceived emotional state.

Overall, activations in both unimodal conditions revealed activations in separate networks that are classically associated with the processing of emotional information in the respective modality.

5.4.2 Is Emotional and Neutral Audiovisual Information Processed Differently?

We presented two conflicting hypotheses for the audiovisual processing of emotional stimuli. The first posits enhanced activation of typical multisensory areas such as the STG for emotional stimuli. The other posits the opposite pattern, namely decreased STG-activation for audiovisual emotional stimuli due to an attenuation of auditory processing by preceding visual information. Our data clearly support the second hypothesis. We observed a strong activation decrease for emotional compared to neutral stimuli in the bilateral STS/STG. Note that this was only the case in the audiovisual condition were visual information preceded the sound onset.

This suggests a strong influence of the preceding visual information on the incoming auditory signal; auditory responses in the STS/STG are attenuated in the case of emotional stimuli. As no differential activation in the STG/STS was observed in the unimodal auditory condition, this effect apparently does not arise from physical differences in the auditory signal. Rather, visual processing seems to play a dominant role. This assumption is further supported by an increased activation in the right fusiform gyrus for emotional compared to neutral stimuli. The same activation increase can also be observed in the unimodal visual condition thus suggesting the impact of visual information on audiovisual processing.

Within one modality, such an attenuation of neural responses to predicted signals has been widely reported, in the auditory domain in the MMN-literature (Näätänen, Paavilainen, Rinne, & Alho, 2007) and in the visual domain mainly under the term "predictive coding" (for review, see Summerfield & Egner, 2009). In audiovisual processing, such a reduction in auditory processing in response to predictive visual information has been reported in magneto- and electroencephalographic studies (Stekelenburg & Vroomen, 2007; Arnal, Morillon, Kell, &

Giraud, 2009; Vroomen & Stekelenburg, 2010; Jessen & Kotz, 2011), who describe a reduction in auditory M100/N100 amplitude following prior visual information. Connectivity analyses show that in the case of visual predictability, the connectivity between left STS and auditory cortex increases (Arnal et al., 2009). Conversely, the connectivity between motion-sensitive areas and middle STS increases with decreasing predictability of the visual signal. Hence, the STS seems to play a key role for the influence of visual information on auditory processing. Similar results arise from crossmodal priming; incongruent targets elicit a larger activation in the STG than congruent targets (Noppeney et al., 2008). Thus, STG-activity is reduced when the prediction based on visual information is fulfilled.

The fact that such an interaction may be enhanced for emotional content is supported by recent findings suggesting enhanced integration of looming, and thereby threatening, audiovisual stimuli, over other, non-threatening, stimuli (Maier et al., 2004; Cappe, Thut, Romei, & Murray, 2009). Hence, emotional content seems to enhance the role of predictive visual information, as indicated by the increased activity in the right fusiform gyrus. Auditory information in contrast looses importance, resulting in decreased activation.

The fact that we observed such a shift in activation while previous studies on audiovisual emotion perception did not can be explained by the specific features of our stimulus material. We believe that one crucial prerequisite for attenuated auditory processing by prior visual information is visual predictability, as suggested in several EEG-studies (Stekelenburg & Vroomen, 2007; Arnal et al., 2009; Vroomen & Stekelenburg, 2010). Visual predictability cannot be provided by static photographs (Pourtois et al., 2005; Ethofer, Pourtois, & Wildgruber, 2006; Müller et al., 2011), and even studies using video clips used much shorter stimulus material (Kreifelts et al., 2007; Robins et al., 2009). Therefore, the temporal delay between visual and auditory onset potentially was not sufficient to build up a clear prediction regarding the auditory information in those studies.

While our results provide an interesting starting point, future studies will be necessary to further delineate the predictive role of emotional content in dynamic visual information. Are the observed effects specific to negative/threatening emotional information, or can similar results be obtained in the case of positive emotions? Furthermore, the relation between body and face information needs to be studied in greater detail. As previous studies on audiovisual emotion recognition focused on facial expressions (Pourtois et al., 2005; Ethofer, Pourtois, &

Wildgruber, 2006; Kreifelts et al., 2007; Robins et al., 2009; Müller et al., 2011), it has to be tested whether the observed effect is specific to body expressions or can also generalize to facial information.

5.4.3 A Network for Audiovisual Interaction in Complex Stimuli?

The conjunction analysis $(AV > A) \cap (AV > V)$ revealed a network consisting of bilateral STG, thalamus, and cerebellum. Both, STG and thalamus, have been consistently reported in studies investigating the integration of auditory and visual information (Ethofer, Pourtois, & Wildgruber, 2006; Kreifelts et al., 2007; Park et al., 2011).

The thalamus is assumed to be one of the first, subcortical loci of multisensory integration (Cappe, Rouiller, & Barone, 2009). Not only does the thalamus receive feedback input via numerous cortico-thalamic connections, it has also been proposed as a "relay-station" for connections between cortical areas (Sherman & Guillery, 2002). Based on tracer studies in the macaque, it has been suggested that the thalamus may receive information from unimodal cortical areas, integrate information, and then feed this information to designated multisensory cortical areas (Cappe, Morel, Barone, & Rouiller, 2009).

A similar interpretation regarding the coactivation of thalamus and STG has been put forward by Kreifelts et al. (2007), who suggest a direct connection between the two areas in the processing of audiovisual emotional information. Furthermore, the STS is thought to be involved in the binding of auditory and visual features originating from the same source (Beauchamp et al., 2004).

While activations in the STG have been commonly observed for emotional stimuli, the STG has also been implied in the audiovisual integration of other types of stimuli, such as speech (Calvert et al., 2000; van Atteveldt, Formisano, Goebel, & Blomert, 2004) or objects (Beauchamp et al., 2004).

In addition to activations in the STG and thalamus, we observed audiovisual interaction in the cerebellum, which has been found in various previous studies on multisensory integration (e.g. Bushara et al., 2001; Skipper, Nusbaum, & Small, 2005). Its precise role in audiovisual perception, however, remains unclear. One link to audiovisual perception of dynamic stimuli seems to be the processing of biological motion. Studies have shown an involvement of the cerebellum not only in the production of movements, but also in the visual perception of body motion and action understanding (Sokolov, Gharabaghi, Tatagiba, & Pavlova, 2010). Furthermore a bidirectional connection between the STS and the cerebellum has been suggested (Sokolov et al., 2011).

In addition, studies have shown activations in the cerebellum to both, visual and auditory motion perception (Baumann & Mattingley, 2010), and differences in activation in response to congruent or incongruent audiovisual perception of motor actions (Petrini et al., 2011).

Hence, it seems plausible that the cerebellum plays a role in the audiovisual perception of stimuli depicting motor actions such as body expressions. Our study differs from studies investigating audiovisual emotion perception using facial expressions, and thereby only very little motor actions. Most of these studies did not report cerebellar activation (Dolan et al., 2001; Ethofer, Pourtois, & Wildgruber, 2006; Kreifelts et al., 2007). Interestingly, studies contrasting static with dynamic facial or bodily expressions report cerebellar activations (Kilts et al., 2003; Grèzes et al., 2007), which provides further evidence for the role of the cerebellum in processing dynamic body information.

The separate conjunction analyses for emotional and neutral stimuli show that the network described above – STG, thalamus, cerebellum – is clearly activated in the neutral condition, while in the emotion conditions, the only activation is observed in the right STG. These results further support our assumption that the visual modality dominates the perception of audiovisual emotional information. Emotional processing is characterized by the influence of predictive visual information on auditory processing, rather than a balanced interaction between both modalities.

5.4.4 Conclusion

We investigated the perception of emotional information from dynamic body and vocal expressions. Unimodal emotion perception and multisensory perception of neutral stimuli activate the networks typically described, including earlier sensory as well as later evaluative brain areas. The perception of audiovisual emotional compared to nonemotional information was characterized by strong decreased activation in the bilateral STG/STS and increased activation in the right fusiform gyrus. Hence, emotion processing seems to be driven primarily by visual information preceding the sound onset of emotional vocal expressions, which in turn results in attenuated processing of auditory emotional information. As revealed in the conjunction analysis, neutral audiovisual interactions take place in the bilateral STG, thalamus, and cerebellum. This suggests a strong interaction between both, auditory and visual information. Emotional audiovisual interactions in contrast only elicited an increased activation in the right STG. Hence, emotional content seems to shift processing from a classical interaction, as observed for neutral stimuli, to a processing driven by visual information.

Acknowledgments

The authors would like to thank Anke Kummer and Simone Wipper for help with the data acquisition and Joeran Lepsien, Kathrin Rothermich, Anna Mestres-Misse, and Jonas Obleser for help with the data analysis.

Table 5.1: *Enhanced Activations in Response to Auditory Stimuli.* Listed are all significant clusters at a threshold of p<.001 (uncorr.) with an extent of k>54, thus resulting in a significance level of p<.05 across the whole brain correcting for multiple comparison. In the contrasts “emotion > neutral” and “neutral >emotion”, neutral auditory stimuli are compared to both, anger and fear stimuli. In the remaining contrasts, anger and fear stimuli are compared to emotional ones separately. l. = left, r. = right, pdACC = posterior dorsal anterior cingulate cortex, preSMA = pre supplementary motor area, IFG = inferior frontal gyrus.

	MNI peak voxel	Z score	Cluster size
Emotion > Neutral			
l. pdACC/preSMA	–6 24 42	4.35	292
l. IFG	–46 16 4	4.54	1162
r. cerebellum (crus I)	36 –74 –30	4.44	72
Anger > Neutral			
l. pdACC/preSMA	–8 24 40	4.44	380
l. IFG	–46 18 12	4.26	1267
Fear > Neutral			
l. IFG	–52 24 4	4.55	601
l. inferior parietal lobe	–44 –56 48	4.18	121
l. pdACC/preSMA	–6 26 42	3.85	153
Neutral > Emotion			
r. fusiform gyrus	32 –46 –10	3.69	65
Neutral > Fear			
l. inferior occipital cortex	–20 –90 –10	3.97	100
r. fusiform gyrus	32 –48 –10	3.95	144
r. supramarginal gyrus	60 –20 36	3.89	164

Table 5.2: *Enhanced Activations in Response to Visual Stimuli.* Listed are all significant clusters at a threshold of p<.001 (uncorr.) with an extent of k>54, thus resulting in a significance level of p<.05 across the whole brain correcting for multiple comparison. In the contrast "emotion > neutral", neutral visual stimuli are compared to both, anger and fear stimuli. In the remaining contrasts, anger and fear stimuli are compared to emotional ones separately. r. = right, l. = left.

	MNI peak voxel	Z score	Cluster size
Emotion > Neutral			
l. middle occipital cortex	–12 –96 2	4.25	386
r. middle temporal gyrus	50 –74 –2	3.86	193
r. fusiform gyrus	38 –48 –16	3.78	99
l. calcarine cortex	–2 –82 –14	3.50	68
Anger > Neutral			
l. middle occipital cortex	–16 –91 –3	3.87	223
r. amygdala	20 –6 –10	3.61	59
Fear > Neutral			
r. middle temporal gyrus	48 –74 –4	4.16	820
l. calcarine cortex	–2 –80 –10	3.86	240
	–8 –96 2	4.00	96
l. middle temporal gyrus	–48 –76 –4	3.71	90
l. fusiform gyrus	–26 –62 –14	3.61	101
l. middle occipital cortex	–24 –92 8	3.59	69

Table 5.3: *Enhanced Activations in Response to Audiovisual Stimuli.* Listed are all significant clusters at a threshold of p<.001 (uncorr.) with an extent of k>54, thus resulting in a significance level of p<.05 across the whole brain correcting for multiple comparison. In the contrasts "emotion > neutral" and "neutral > emotion", neutral visual stimuli are compared to both, anger and fear stimuli. In the remaining contrasts, anger and fear stimuli are compared to emotional ones separately. r. = right, l. = left, STS = superior temporal sulcus, STG = superior temporal gyrus, OFC = orbitofrontal cortex, IFG = inferior frontal gyrus.

	MNI peak voxel	Z score	Cluster size
Emotion > Neutral			
r. fusiform gyrus	36 –56 –16	3.94	180
Anger > Neutral			
l. hippocampus	–26 –32 –2	3.88	116
Neutral > Emotion			
l. STS/STG	–58 –14 –6	5.47	2063
r. STS/STG	64 –30 –2	5.27	2726
r. inferior OFC	48 34 –6	3.99	101
r. IFG	44 20 20	3.74	274
Neutral > Anger			
r. STS/STG	62 –26 –4	4.97	1650
l. STS/STG	–58 –14 –6	4.16	323
r. supramarginal gyrus	60 –38 44	4.22	100
l. inferior temporal lobe	–54 –50 -20	3.66	120
r. superior frontal lobe	20 16 44	3.54	100
r. IFG	44 20 22	3.40	69
Neutral > Fear			
r. anterior STG	48 –30 2	5.08	1832
l. STS/STG	–60 –28 2	4.63	1442
r.posterior STS/STG	64 –46 18	3.92	86
r. inferior OFC	46 34 –6	3.64	99
r. IFG	44 20 18	3.35	79

Table 5.4: *Activations in Conjunction Analysis* ($AV > V \cap AV > A$) Listed are all clusters at a threshold of p<.001 (uncorr.) with an extent of k>10 that showed a larger activation for audiovisual stimuli than for either unimodal condition in the conjunction analysis ($AV > V \cap AV > A$). In the conjunction analysis "all", emotional as well as neutral stimuli are included, while for "emotion" only anger and fear stimuli and for "neutral" only neutral stimuli were considered. * = k>54, resulting in a significance level of p<.05 across the whole brain correcting for multiple comparison. r. = right, l. = left, STG = superior temporal gyrus, OFC = orbitofrontal cortex, IFG = inferior frontal gyrus.

	MNI peak voxel	Z score	Cluster size
All			
r. STG	58 –32 14	4.61	202*
l. STG	–44 –32 20	3.21	12
l. thalamus (MGB)	–8 –30 –6	4.46	82*
r. thalamus (MGB)	8 –30 –6	3.47	11
l. cerebellar vermis (6)	–30 –54 –30	3.79	31
	–22 –64 –24	3.35	12
r. cerebellar vermis (6)	32 –52 –28	3.43	21
Emotion			
r. STG	58 –32 –14	3.84	55*
Neutral			
r. STG	58 –32 14	4.72	415*
l. STG	–62 –56 4	3.47	40
l. cerebellar vermis (6)	–32 –54 –30	4.02	186*
r. cerebellum (crus I)	36 –70 –28	3.35	22
l. thalamus	–10 –28 –6	3.67	82*
r. thalamus	28 –16 0	3.55	64*
	18 –16 -6	2.28	13
	10 –26 –6	3.38	12
r. IFG	46 14 28	3.56	98*
r. dorsal premotor cortex	42 4 50	3.48	94*
r. superior temporal lobe	36 2 –18	3.43	13

Chapter 6

How Bodies and Voices Interact in Early Emotion Perception [*]

S. Jessen[1,2], J. Obleser[1] & S.A. Kotz[1]

[1]: Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1A, 04103
Leipzig, Germany

[2]: Cluster of Excellence "Languages of Emotion", Free University of Berlin, Habelschwerdter
Allee 45, 14195 Berlin, Germany

Successful social communication draws strongly on the correct interpretation of others' body and vocal expressions. Both can provide emotional information and often occur simultaneously. Yet their interplay has hardly been studied. Using electroencephalography, we investigated the temporal development underlying their neural interaction in auditory and visual perception. In particular, we tested whether this interaction qualifies as true integration following multisensory integration principles such as inverse effectiveness. Emotional vocalizations were embedded in either low or high levels of noise and presented with or without video clips of matching emotional body expressions. In both, high and low noise conditions, a reduction in auditory N100 amplitude was observed for audiovisual stimuli. However, only under high noise, the N100 peaked earlier in the audiovisual than the auditory condition, suggesting facilitatory effects as predicted by the inverse effectiveness principle. Similarly, we observed earlier

---

[*] Jessen, S., Obleser, J., & Kotz, S. A. (2012). How Bodies and Voices Interact in Early Emotion Perception. PLoS ONE, 7(4), e36070.

N100 peaks in response to emotional compared to neutral audiovisual stimuli. This was not the case in the unimodal auditory condition. Furthermore, suppression of beta-band oscillations (15–25 Hz) primarily reflecting biological motion perception was modulated 200–400 ms after the vocalization. While larger differences in suppression between audiovisual and audio stimuli in high compared to low noise levels were found for emotional stimuli, no such difference was observed for neutral stimuli. This observation is in accordance with the inverse effectiveness principle and suggests a modulation of integration by emotional content. Overall, results show that ecologically valid, complex stimuli such as joined body and vocal expressions are effectively integrated very early in processing.

6.1 Introduction

Body expressions and vocalizations play an important role in communication. We can readily determine from either modality someone's gender (Pollick, Kay, Heim, & Stringer, 2005), emotion (Atkinson et al., 2004; Paulmann & Kotz, 2008a), or how familiar a person is (Loula, Prasad, Harber, & Shiffrar, 2005; Latinus & Belin, 2011). Crucially, these sources of information are not independent: both are rooted in biological motion. While body expressions per definition are biological motion, vocalizations are generated by our vocal tract and strongly influenced by body posture, making them a product of biological motion. Furthermore, both provide closely time–locked and congruent information.

Not surprisingly, early interactions of body and vocal expressions are reliably observed (van den Stock et al., 2008; Jessen & Kotz, 2011). However, do these interactions truly reflect integration (B. Stein & Meredith, 1993)? Various approaches have been suggested to address this question. One popular method, especially in the investigation of audiovisual emotion perception, is the use of mismatch paradigms, in which violation responses can be observed when the two modalities provide conflicting information (Ethofer, Anders, Erb, Droll, et al., 2006; Campanella & Belin, 2007). While these studies suggest an integration of facial and vocal information, they can only indirectly infer integration from incongruent responses. Additionally, in an ecologically valid context, congruency between modalities is far more common than incongruency.

A second approach to quantify multisensory integration is the comparison of an audiovisual condition to the sum of the two unimodal conditions, potentially resulting in either sub- or superadditivity (Calvert et al., 2001; Senkowski, Gomez-Ramirez, et al., 2007). However, the reliability of this criterion has recently been criticized (Beauchamp, 2005b; Laurienti et al., 2005). An alternative way to test multisensory integration is the inverse effectiveness (IE) paradigm, which states that interaction effects should be largest if a stimulus is unimodally least effective (B. Stein & Meredith, 1993). Originally developed in neurophysiological recording (B. Stein & Meredith, 1993; Ghazanfar, Maier, Hoffman, & Logothetis, 2005; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007), this paradigm has been extended to functional magnetic resonance imaging (fMRI) studies (Holle et al., 2010; Werner & Noppeney, 2010b). Recently, Senkowski and colleagues (Senkowski et al., 2011) were able to demonstrate inverse effectiveness also in event-related potentials (ERPs). Using very simple visual and auditory stimuli, they report enhanced multisensory interactions for low intensity stimuli within 60 ms after stimulus onset. By using complex, naturalistic stimuli we seek to extend these findings to an ecologically valid communicative context.

Such a communicative context comprises facial, bodily, and vocal expressions, however, most previous studies investigating multisensory communication have focused on the interplay between facial and vocal (Pourtois et al., 2002; Ethofer, Anders, Erb, Droll, et al., 2006) or facial and bodily expressions (Meeren et al., 2005). While we were able to demonstrate an early interaction between facial, body, and vocal expressions in a previous study (Jessen & Kotz, 2011), it cannot be ruled out that the reported results were mainly driven by the processing of facial expressions. To focus specifically on the body–voice interaction, in the current study we blurred facial expressions contained in the stimulus material to avoid any influence of facial information. In contrast to previous studies using incongruent body and vocal expressions (van den Stock et al., 2007, 2008), we further investigate a possible integration in a congruent setting.

Hence, we addressed two main questions. (i) Can an early interaction between voice and body be observed in ecologically valid stimulus set-ups? If so, does this interaction result in facilitated processing, as has been previously described for audiovisual interactions (Paulmann et al., 2009; Thomas & Shiffrar, 2010)? (ii) Does such an early interaction follow the inverse effectiveness principle, suggesting multisensory integration (Werner & Noppeney, 2010b)?

While this interaction is important for many communicative settings, an especially striking one is emotional communication. Correctly identifying others' emotions is of high relevance, and perceiving emotions via multiple modalities provides a strong processing benefit (de Gelder & Vroomen, 2000; Kreifelts et al., 2007). Emotion perception therefore offers an ideal example to examine the interaction between body expressions and vocalizations.

We recorded the electroencephalogram (EEG) to investigate early emotional and sensory processes. To assess the impact of visual on auditory processing, we focused on the auditory N100, an ERP component robustly reflecting this impact (Stekelenburg & Vroomen, 2007) and indicating facilitated processing in shorter peak latencies (van Wassenhove et al., 2005) and reduced peak amplitudes (van Wassenhove et al., 2005; Stekelenburg & Vroomen, 2007). As the sound onset in the unmanipulated, natural videos does not coincide with the video onset, computing ERPs in relation to the video onset was not of interest to the present study. In addition, the information contained in the video material is unfolding over the course of several seconds with no fixed event to allow for the computation of ERPs measuring visual processing. To circumvent these problems and assess non-phase-locked interactions between visual and auditory processing, we analyzed oscillatory changes in EEG power. In particular, a suppression of the beta-band (15–25 Hz) (Ulloa & Pineda, 2007; Perry, Troje, & Bentin, 2010) is thought to reflect the processing of biological motion such as body movements. We therefore flanked the N100 ERP analysis with a time–frequency-analysis. To further support the assumption that the observed beta-band effects indeed reflect the processing of biological motion, we conducted a source localization of beta-band changes based on the results of the time–frequency-analysis.

6.2 Materials and Methods

6.2.1 Participants

Twenty-four native German speakers (12 female), who had not taken part in the study reported in Jessen and Kotz (2011), participated in the current study (mean age 24.7 years, SD (standard deviation) = 2.9 years). All were right-handed, had normal or corrected-to-normal vision, and did not report any hearing impairments. They were compensated financially for their participation, and gave written informed consent prior to the experiment. The study was approved by the ethics committee at the University of Leipzig.

6.2.2 Stimuli and Design

We used video clips containing emotional body expressions as well as recordings of emotional interjections (“ah”, “oh”, and “mh”) (Dietrich et al., 2008) in three different affective states: “anger”, “fear”, and “neutral” (see Figure 6.1 for an example). All three states were video recorded from four different semi-professional actors; 2 women (24 and 41 years old) and 2 men (30 and 48 years old) and each emotion was expressed vocally by the interjections. We recorded several takes of each emotion. Based on a prior rating study (see below) we selected 10 items for each condition, amounting to 360 stimuli in total. Further details about the recording of the stimulus material can be found elsewhere (Jessen & Kotz, 2011). In order to investigate the perception of body expressions while controlling for facial expressions, the faces of the actors were blurred using the software Motion 3.0.2 (Apple Inc.), rendering the facial expressions unrecognizable. In a pilot study 16 participants were asked to categorize the emotion expressed by the actors in these videos; the intended emotion was recognized with a mean accuracy of 96.2 % (SE (standard error) = 0.8). All auditory stimuli were embedded in white noise using Matlab 7.7.0 (The MathWorksInc, Natick, MA, USA). In a second pilot study, the auditory stimuli were presented to 15 participants at different signal-to-noise ratios (SNRs): 0dB, –6 dB, –12 dB, –18 dB, and –24 dB. The participants were asked to judge the expressed emotion. The SNR of –12 dB resulted in a decrease in performance but emotions were still reliably recognized above chance (average recognition accuracy of 81.1 %, SE = 2.2). This ratio was chosen as the high-noise condition in the EEG experiment. As the low-noise condition, a signal-to-noise ratio of 0 dB was used (average recognition accuracy of 88.2 %, SD = 1.6). We decided to use a low-noise rather than a no-noise condition in order to minimize the physical difference between the two noise levels.

Stimuli were presented in five conditions (Figure 6.1). In two auditory conditions, interjections were presented at an SNR of 0 dB ($\text{audio}_{lowNoise}$) or –12 dB ($\text{audio}_{highNoise}$) while a static picture of the respective actor was shown on the screen, and thus no visual dynamic emotional information was conveyed. In the two audiovisual conditions, interjections were presented at the same SNRs, but this time accompanied by the matching video ($\text{audiovisual}_{lowNoise}$

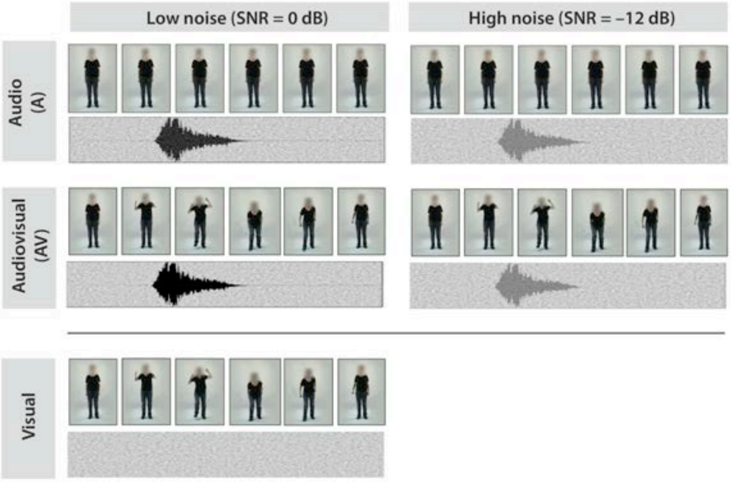

Figure 6.1: *Example of the Stimulus Material.* Six frames of a video clip of a female actress depicting "anger" are shown in the five conditions used in the experiment. To test for inverse effectiveness, we compared the difference between audiovisual and auditory stimuli under high and low noise conditions ($(\text{AV-A})_{lowNoise}$ vs. $(\text{AV-A})_{highNoise}$). Additionally, we recorded responses in a purely visual condition (bottom row).

and audiovisual$_{highNoise}$, respectively). Finally, in the visual condition, videos were presented accompanied by a sound file containing white noise but no interjection. In each of these conditions, 24 angry, 24 fearful, and 24 neutral stimuli were presented, amounting to a total of 360 stimuli. The distribution of actors and interjections was counterbalanced across all conditions.

6.2.3 Data Acquisition

Participants sat comfortably in a dimly lit sound-shielded chamber at a distance of approximately 120 cm from the monitor. Videos presented at a visual angle of 8.86°. Auditory stimuli were presented via headphones (SONY Stereo Headphones, MDR-XD100) at the same level of loudness for all participants. Before the experiment, the participants completed a short training that consisted of 10 trials and contained items not used in the actual experiment. A trial started with the presentation of a fixation cross (1 second), followed by the presentation of the stimulus which lasted on average 4.37 seconds (SD = 1.03). After each stimulus, the participants judged the previously seen emotion in a forced-choice task with the options "anger", "fear", and "neutral" using the three left-most keys of a 4-button response box. Half of the participants responded with the index-, middle-, and ring finger of the right hand, while the other half used their left hand. Furthermore, the response assignment was also counter-balanced across all participants. After 3 seconds, or as soon as the participant responded, an inter-trial-interval of 2 seconds started, during which the participants saw a grey screen. The participants were instructed to avoid eye blinks during the trial and use the inter-trial-interval to blink. After 20 trials participants took a short, self-paced break, leading to a total of 18 blocks. The experiment was programmed using the Presentation software package (Neurobehavioral Systems, Inc.).

We recorded from 64 Ag-AgCl electrodes according to the modified 10-20 system (Sharbrough et al., 1991). The sternum served as ground, the left mastoid as reference. Electrodes were mounted in an elastic cap (Electro-Cap International, Eaton, OH, USA), impedances were kept at less than 5 $k\Omega$ and the signal was bandpass-filtered only between DC and 140 Hz and recorded using the BrainVision Recorder software (Brain Products GmbH, Munich, Germany).

6.2.4 Data Analysis

As the task was mainly designed to direct the participants' attention to the stimulus material, we do not focus on the behavioral results. Nevertheless, we computed *d'* as the discrimination ability between each pair of emotions (anger vs. fear, anger vs. neutral, fear vs. neutral), separately for the five modality conditions (MacMillan & Creelman, 2005). On the resulting values, a repeated–measures ANOVA with the factors emotion–contrast (anger vs. fear, anger vs. neutral, fear vs. neutral) and modality (audio$_{lowNoise}$ audio$_{highNoise}$, audiovisual$_{lowNoise}$, audiovisual$_{highNoise}$, visual) was computed.

For computing ERPs, the data was re-referenced offline to linked mastoids, and filtered with a 1–30 Hz bandpass filter. Trials containing EOG artifacts larger than 30 μV were rejected automatically, and the resulting data were inspected visually to remove additional artifacts. Two participants had to be excluded from further analysis due to excessive artifacts. We excluded trials in which the emotional content was not correctly identified, and averaged the remaining trials over a length of –200 to 2000 ms in relation to the sound onset. As accuracy rates varied between the different conditions, the number of trials averaged differed. In order to ensure the results were not confounded by a varying signal-to-noise ratio resulting from the varying number of trials (due to artifact rejection and accuracy differences), we computed the same ERP analyses described below with only 60 % of the trials per condition, and thus with the same number of trials. We obtained the same results as with the full number of trials, and hence in the following report results with all correct trials per condition.

We conducted two separate analyses of the ERPs, one considering the peak latency in a time-window of 100–230 ms after sound onset, and one considering the amplitude at the peak of the component. Based on the visual inspection of the topographic distribution of the ERP responses we defined a fronto-central region of interest (FC3, FCz, FC4, C3, C1, Cz, C2, C4). For both, amplitude and peak latency, we computed a repeated–measures ANOVA with the factors emotion (anger, fear, neutral), modality (audiovisual, audio), and noise-level (high, low), and Student's t-tests for a step-down analysis of interactions. The visual condition was not included in this specific analysis of the auditory evoked response, as no visual ERP was expected at this time point. A Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied to contrasts including more than two conditions. Furthermore, the results of the Student's t-tests were corrected for multiple comparisons using the Bonferroni-Holm-method (Holm, 1979), resulting in corrected alpha-thresholds of 0.05, 0.025, and 0.017 for an initial p-value of 0.05. Effect sizes for ANOVAs are given in ω^2, while effect sizes for t-tests are given in r (Rosenthal, 1991).

To analyze the data in the frequency domain, we re-referenced the data to the average of all electrodes and again excluded trials containing EOG artifacts larger than 30 μV. Furthermore, a bandpass filter was applied, ranging from 0.1–100 Hz. For each trial, we computed a time-frequency representation in a time-window of –1000 to 1000 ms relative to the sound onset using the Matlab toolbox FieldTrip (Oostenveld et al., 2011). We used Morlet's wavelets (Tallon-Baudry & Bertrand, 1999) with a time–frequency relation of m = 7, and calculated condition and subject-specific average time–frequency representations. Changes in spectral power were computed relative to a baseline of –500 to 0 ms before the video onset. Based on a previous study by Perry et al. (Perry, Troje, & Bentin, 2010) as well as our own previous work using comparable multimodal stimulation (Jessen & Kotz, 2011), we focused on the low beta-range (15–25 Hz). Based on the visual inspection of the data, we chose a time-window of 200–400 ms for further analysis. We computed the average power change across the respective time-window for two groups of electrodes, a central group (C3, C1, Cz, C2, C4) and an occipital group (O1, Oz, O4). While central electrodes have been commonly used to measure changes in mu- and beta-rhythm, occipital electrodes have been used to measure sensory processing (Perry, Troje, & Bentin, 2010). On these average power changes we performed the same statistical comparisons as described in the ERP section, that is, a repeated–measures ANOVA with the factors emotion (anger, fear, and neutral), modality (audio, audiovisual), and noise-level (high, low). In order to contrast the visual and the audiovisual conditions, we computed an additional repeated–measures ANOVA with the factors emotion (anger, fear, and neutral) and modality (visual, audiovisual$_{lowNoise}$ and audiovisual$_{highNoise}$). Again, we used Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) when appropriate, the Bonferroni-Holm-method to correct for multiple comparisons (Holm, 1979), and report effect sizes for ANOVAs in ω^2 and for t-tests in r (Rosenthal, 1991). To ensure the observed power-suppression is indeed induced, we furthermore computed the inter-trial coherence (Lachaux, Rodriguez, Martinerie, & Varela, 1999), irrespective of condition (see Figure 6.5).

In order to localize the source of the observed changes in beta-band power, we used dynamic imaging of coherent sources (DICS) (Gross et al., 2001), a localization algorithm particularly suited for localizing sources of oscillatory activity. We followed a protocol suggested in various previous studies (Medendorp et al., 2007; Haegens, Osipova, Oostenveld, & Jensen, 2010): We repeated the time–frequency-analysis using a Hanning multi-taper approach, this time focusing on 20 Hz with a 5 Hz smoothing in a time-range of 200–700 ms after sound onset, thereby selecting the time-window chosen above for the statistical analysis but extending it to have a sufficient length to obtain a reliable estimation and using a multitaper frequency transformation. The source localization was conducted using electrode locations for the modified 10-20-system and a standard MRI template. In a first analysis step, we contrasted the overall beta-activity (across all conditions) to the beta-activity in the baseline using a one-sampled t-test in order to identify brain structures most likely generating the overall beta-power. In a second analysis, we specifically aimed at localizing the condition differences in beta-suppression, and thus to compare the emotion conditions (anger and fear) to the neutral condition, collapsed across both audiovisual conditions irrespective of noise level, using a one-way within-subject ANOVA. While the source localization was computed using FieldTrip (Oostenveld et al., 2011), the statistical comparisons as well as the visualizations were calculated in SPM8 (Wellcome Department of Imaging Neuroscience, London, UK).

6.3 Results

6.3.1 Behavioral Results

Emotions were recognized above chance in all conditions (accuracy for the different modality conditions: audio$_{lowNoise}$: 92.42 % (SD = 5.53); audio$_{highNoise}$: 82.77 % (SD = 8.28); audiovisual$_{lowNoise}$: 98.48 % (SD = 2.99); audiovisual$_{highNoise}$: 98.11 % (SD = 2.59); visual: 94.19 % (SD = 9.77)). Auditory stimuli, irrespective of noise level, were differentiated worse than either audiovisual or visual stimuli ($F(2.13, 44.72) = 78.61, p < .0001, \omega^2 = .60$) (see Figure 6.2). Furthermore, participants showed lower perceptual sensitivity for auditory stimuli at high noise levels compared to low noise levels ($t(21) = 7.38, p < .0001, r = .84$). An interaction with the factor emotion ($F(2.77, 58.20) = 16.32, p < .0001, \omega^2 = .19$) shows that for high a well as low noise levels in the auditory condition, participants performed worst at distinguishing anger from fear (high noise: $t(21) = 4.96, p < .0001, r = .73$, low noise: $t(21) = 3.38, p < .01, r = .59$, both compared to the next-worst distinction). While in high noise levels, the discrimination is worse between anger and neutral compared to fear and neutral ($t(21) = 3.41, p < .01, r = .59$), this is reversed in low noise levels ($t(21) = 2.09, p < .05, r = .41$).

6.3.2 ERP Latency

We observed shorter N100 latencies for audiovisual compared to auditory stimuli ($F(1,21) = 5.62, p < .05, \omega^2 = .11$), as well as for low noise compared to high noise ($F(1,21) = 375.00, p < .0001, \omega^2 = .89$). As can be seen in Figure 6.3, an interaction between these two factors ($F(1,21) = 7.235, p < .05, \omega^2 = .07$) shows that these effects are mainly driven by a speed–up of the audiovisual compared to the audio condition at high noise levels ($t(21) = 3.25, p < .01, r = .56$), while there was no difference at low noise levels. Irrespective of the noise level, we found a latency reduction for emotional compared to neutral stimuli ($F(1.88, 39.39) = 6.27, p < .01, \omega^2 = .13$) (see Figure 6.4). As revealed by an interaction between modality and emotion ($F(1.65, 34.71) = 9.22, p < .01, \omega^2 = .114$), this reduction only occurred in the audiovisual condition (anger vs. neutral: $t(21) = 4.74, p < .001, r = .72$, fear vs. neutral: $t(21) = 5.16, p < .0001, r = .75$).

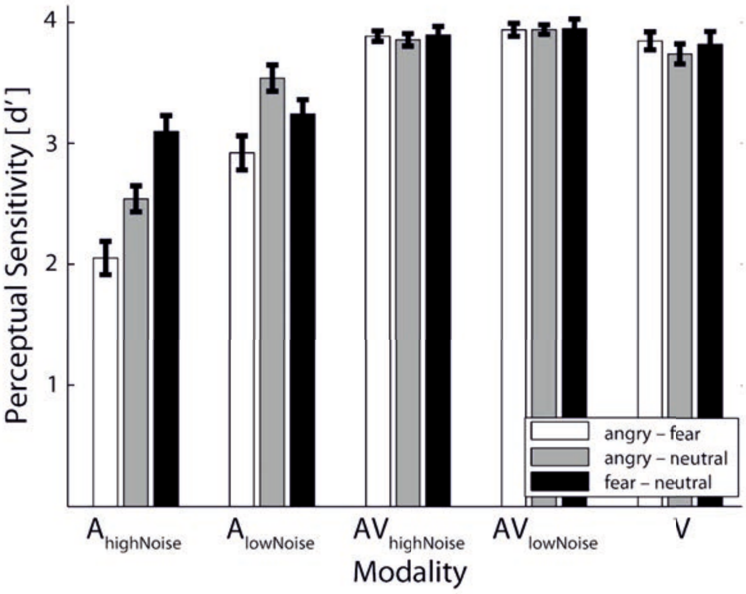

Figure 6.2: *Perceptual Sensitivity in the Different Modalities.* Perceptual sensitivity for distinguishing anger from fear (anger–fear), anger from neutral (anger–neutral), and fear from neutral (fear–neutral) is depicted separately for each modality condition. Both auditory conditions show a significantly lower sensitivity than either audiovisual or the visual condition. Furthermore, lower d' values are observed for $A_{highNoise}$ compared to $A_{lowNoise}$.

6.3.3 ERP Amplitude

Smaller N100 amplitudes were observed for audiovisual compared to auditory stimuli ($F(1,21) = 17.50, p < .0001, \omega^2 = .27$) as well as for high noise compared to low noise stimuli ($F(1,21) = 72.01, p < .0001, \omega^2 = .62$). Furthermore, we found a smaller N100 amplitude for fearful stimuli compared to both, angry ($t(21) = 4.21, p < .001, r = .67$) and neutral stimuli ($t(21) = 4.61, p < .001, r = .71$).

6.3.4 Time–Frequency Results

As can be seen in Figure 6.5, the analyzed time-window shows a comparably low inter-trial coherence (around 0.2; phase-locking ranging from 0 to 1) in the beta-band, demonstrating that power changes are indeed induced rather than phase-locked (Shahin, Picton, & Miller,

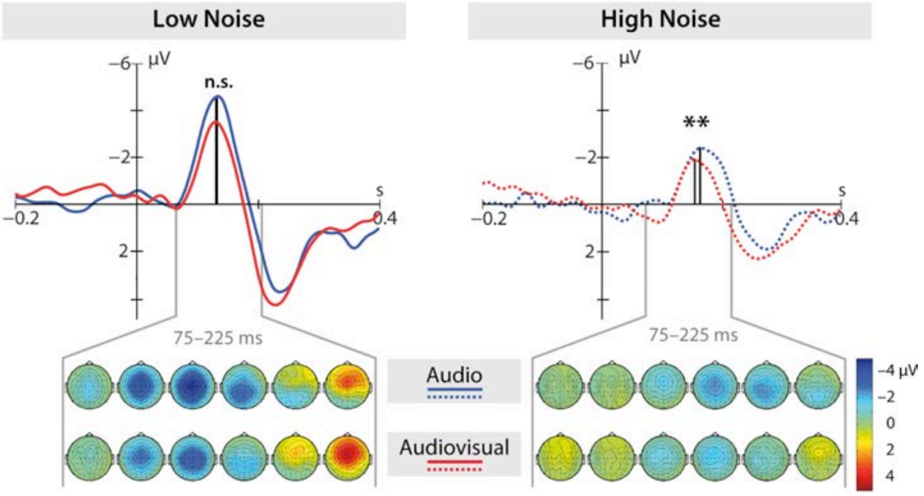

Figure 6.3: *Multisensory Interaction in High and Low Noise.* Event-related potentials (averaged over FC3, FCz, FC4, C3, C1, Cz, C2, C4) in response to the auditory onset are depicted in the low-noise condition (left panel, bold) and in the high-noise condition (right panel, dotted). In red, the response in audiovisual conditions can be seen, while the response to auditory conditions is shown in blue. Earlier N100 peak-latencies can be observed in the high noise condition but not in the low noise condition (n.s. = not significant, $** = p < .01$). In the lower part, topographical maps of the auditory condition (top row) and the audiovisual condition (bottom row) are depicted in steps of 25 ms, starting at 0.075 ms after auditory onset.

2009). At all frequency ranges and in all electrode groups we observed a stronger suppression for audiovisual compared to auditory stimuli (central: $F(1,21) = 31.28, p < .0001, \omega^2 = .40$; occipital: $F(1,21) = 44.38, p < .0001, \omega^2 = .49$) (Figure 6.6). Emotional processing did affect the suppression of oscillatory activity differentially dependent on the electrode location. Only at occipital electrodes ($F(2,41.92) = 9.49, p < .0001, \omega^2 = .20$) a stronger suppression was observed for fearful compared to neutral video clips ($t(21) = -4.41, p < .001, r = .69$), as well as for angry compared to neutral video clips ($t(21) = -2.47, p < .05, r = .47$). At both, central and occipital electrodes, we observed an interaction between the factors emotion and modality (central: $F(1.89,39.69) = 6.36, p < .01, \omega^2 = .08$; occipital: $F(1.81,38.01) = 15.08, p < .0001, \omega^2 = .18$), showing a stronger suppression for audiovisual compared to auditory stimuli only for angry (central: $t(21) = 6.35, p < .0001, r = .81$; occipital: $t(21) =$

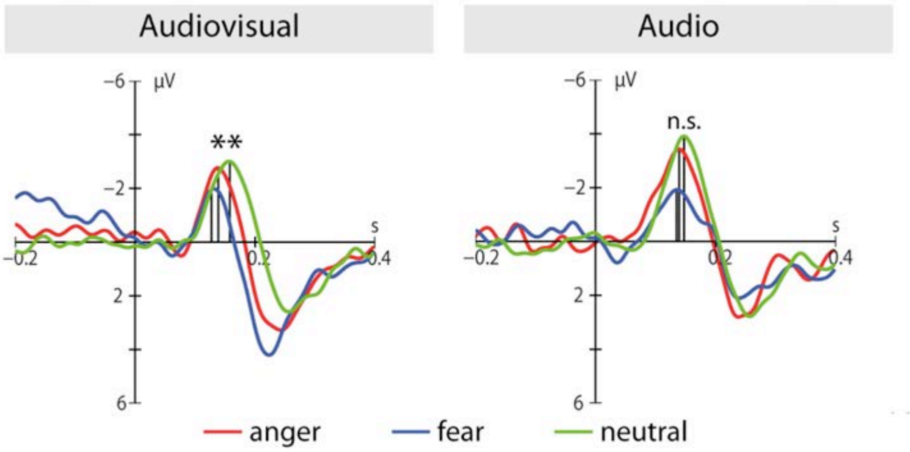

Figure 6.4: *Emotion Processing in the N100.* Event-related potentials (averaged over FC3, FCz, FC4, C3, C1, Cz, C2, C4) in response to the auditory onset in the audiovisual condition (left panel) and the auditory condition (right panel) are shown for the different emotions (red = anger, blue = fear, green = neutral). Only in the audiovisual condition, a reduction in N100 peak-latency is observed (n.s. = not significant, $** = p < .01$).

$7.60, p < .0001, r = .85$) and fearful stimuli (central: $t(21) = 3.80, p < .01, r = .63$; occipital: $t(21) = 4.57, p < .001, r = .70$). Finally, we observed a three-fold interaction between emotion, noise, and modality at central electrodes ($F(1.89, 39.76) = 3.96, p < .05, \omega^2 = .03$), as depicted in Figure 6.6. Only at high levels of noise ($F(1.63, 34.13) = 8.86, p < .01, \omega^2 = .10$), we observed a larger suppression for audiovisual compared to auditory stimuli for angry ($t(21) = 5.63, p < .0001, r = .77$) and fearful ($t(21) = 3.73, p < .01, r = .63$) stimuli.

In the second model we included the visual and both audiovisual conditions and found no modality differences. As in the previous comparison, emotional stimuli resulted in a larger suppression at occipital electrodes ($F(1.52, 31.82) = 22.00, p < .0001, \omega^2 = .32$; anger vs. neutral: $t(21) = -6.53, p < .0001, r = .81$, fearful vs. neutral: $t(21) = -4.47, p < .001, r = .69$).

When localizing the origin of the beta-suppression across all conditions (see Methods) we found the right premotor cortex as the primary source (Figure 7; Z = 6.42, MNI peak coordinate 24, -3, 46; BA 6). Second, the differential beta-suppression for emotional compared to neutral stimuli in the audiovisual conditions seems to originate from the more posterior

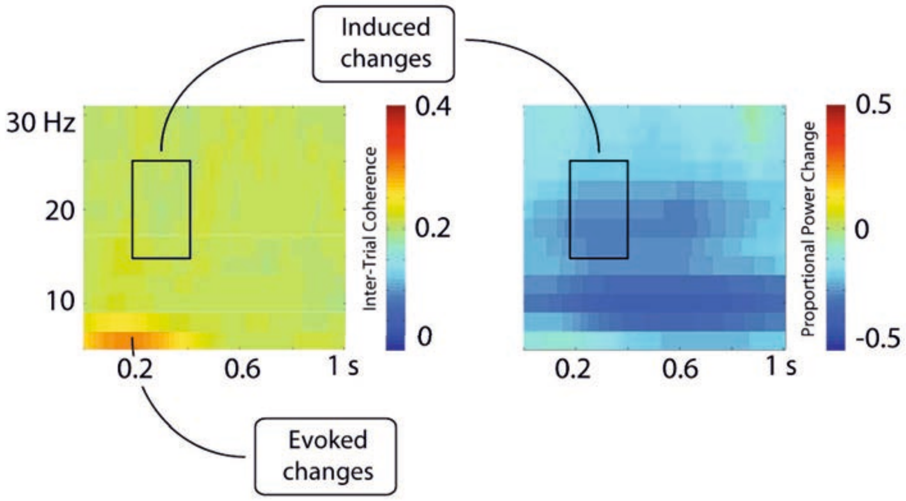

Figure 6.5: *Intertrial Coherence and Power Changes Across all Conditions.* In the left panel, the intertrial coherence irrespective of condition is depicted. While a strong coherence can be seen shortly after sound onset in the range below 10 Hz, this is not the case for the time-window used to investigate the beta-suppression, indicated by the black frame. In the right panel, power changes are shown, again irrespective of condition.

superior parietal cortex (Z = 4.19, MNI peak coordinate 14, -54, 78; BA5). The condition-wise power changes as obtained from this posterior parietal source (Figure 6.7) show that difference between emotional and neutral stimuli can be seen in both, high and low noise levels.

6.4 Discussion

Using ecologically valid stimuli we investigated the integration of body and vocal expressions in emotional communication. Body expressions show a clear influence on the processing of vocalizations leading to facilitated processing within less than 200 ms. Importantly, a modification of this influence by signal quality suggests integration of both modalities (rather than mere interaction) according to the IE principle (B. Stein & Meredith, 1993). Furthermore, emotional content elicits enhanced signal processing irrespective of the accompanying vocalization, as indicated by a stronger beta-suppression for visual as well as audiovisual stimuli.

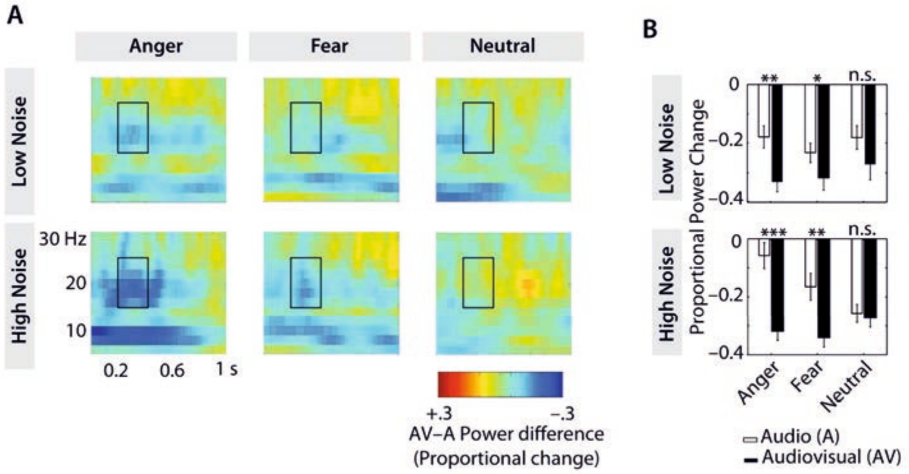

Figure 6.6: *Beta-Suppression in Different Conditions.* Power changes relative to baseline are shown for one second after auditory onset and in a frequency range of 5–31 Hz (6A). Each plot depicts the difference in spectrum between the audiovisual and the auditory condition (AV-A). In the top row, differences in the low noise condition are shown, while in the bottom row, differences in the high noise condition are depicted (in both cases from left to right: anger, fear, neutral). In 6B, mean values for the auditory as well as the audiovisual condition are depicted for the range marked by the black frame (200–400 ms, 15–25 Hz). While stronger suppressions for audiovisual compared to auditory conditions can be seen in both emotion conditions, these differences are larger in the high-noise condition than in the low-noise condition (n.s. = not significant, $* = p < .05, ** = p < .01, *** = p < .001$).

6.4.1 Interaction Between Voice and Body Expressions

On a behavioral level, we observed a clear improvement in performance, when visual and auditory information was present, in comparison to both purely auditory conditions. The participants showed the worst discrimination ability in the unisensory auditory condition with high levels of noise, replicating the results from our pilot study and demonstrating that intelligibility was manipulated as intended.

As predicted, we observed a clear interaction of auditory and visual information in the auditory N100 time-range. Previous studies have reported a facilitatory influence of concurrent visual information on early auditory perception (Stekelenburg & Vroomen, 2007). While we were able to show in a previous study (Jessen & Kotz, 2011) that such an influence also occurs

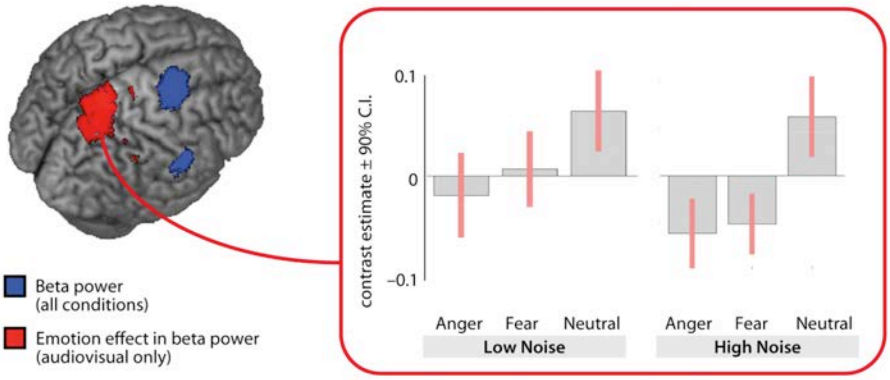

Figure 6.7: *Source Localization of Beta-Suppression.* In the left panel, localized overall beta-suppression is depicted in blue ($p < 10^{-6}$, family-wise error corrected). The difference between emotional and neutral audiovisual beta-power localizes to a more posterior area (shown in red, $p < .001$, uncorrected). This difference is also seen in the contrast estimates (± 90 % confidence intervals) shown in the right panel for both, high and low levels of noise.

in early multisensory communication, we now extend this finding by demonstrating that such facilitation can be observed even in the absence of facial information.

Processing advantages for emotional auditory information have been observed before at slightly later (P200) processing stages (Paulmann et al., 2009; Sauter & Eimer, 2010). However, our results imply that auditory emotion processing can be further sped–up by reliable (i.e. predictive) body motion information. Stekelenburg and Vroomen (Stekelenburg & Vroomen, 2007) suggest that this early facilitatory influence through visual information may be mediated by the simulation of actions in motor areas such as the premotor cortex (Callan et al., 2003), an assumption supported by the localization of the beta-suppression discussed below. Our results provide evidence that this influence not only allows for a precise prediction of the auditory onset, but that this prediction process is modulated by the emotional content of the signal. In contrast to a previous study (Jessen & Kotz, 2011), in which we reported stronger emotion effects in an auditory than an audiovisual condition, we report the reverse pattern in the present study. We observed shorter N100 peak latencies for emotional compared to neutral stimuli only in the audiovisual conditions. This difference in findings can be explained by the fact that in the present study, auditory information was always embedded in noise, leading to a decrease in behavioral performance, lower amplitudes, and longer latencies irrespective of the emotional content. As participants become worse at discriminating the emotions on a behavioral level, it seems plausible that the same pattern can be observed in the ERP data. Concomitantly, a lower signal quality, as in the auditory conditions, is assumed to enhance the role of complementary visual information in the audiovisual condition. This may explain the differentiation observed in the audiovisual condition in the current study.

6.4.2 Integration of Voice and Body Expressions

The second main goal of the present study was to investigate whether vocalizations and body expressions are truly integrated (in the sense of the IE principle) and if so, under which conditions. By manipulating the noise level of the auditory signal, we contrasted differences between auditory and audiovisual processing in conditions of different signal quality. In the auditory N100 time-window, we observed clear inverse effectiveness; only at a high but not at a low noise level, a multisensory combination of visual and auditory information causes a speed–up in the peak latency of the N100. In contrast, if the auditory signal quality is satisfying (i.e. low noise), additional information from body expressions offers little processing advantage. Hence, information from body expressions and vocalizations has to be integrated very early in order to differentially affect auditory processing dependent on the noise level.

While this effect is observed most vigorously in the N100 latency, observations in the time–frequency-domain (beta-band) point in the same direction. At central electrodes, a larger suppression was observed for emotional compared to neutral stimuli. This pattern was observed for high but not for low noise stimuli. Furthermore, this effect points in an interesting direction for future studies: Is audiovisual integration not only affected by exogenous factors such as noise but also modulated by endogenous or content factors of the signal? That is, are emotional stimuli – highly salient and possibly necessitating rapid action – integrated more efficiently and processed faster? This hypothesis receives support from several studies reporting enhanced multisensory integration for highly salient stimuli, such as looming signals, in animals (Maier et al., 2004; Maier, Chandrasekaran, & Ghazanfar, 2008) but also in humans (Cappe, Thut, et al., 2009; Romei, Murray, Cappe, & Thut, 2009). So far, EEG signatures of multisensory integration in communicative settings have mainly been investigated in the mismatching paradigm (de Gelder et al., 1999; Pourtois et al., 2002). We show that the IE principle, which has been primarily applied in fMRI and neurophysiological research (Callan et al., 2003; Ghazanfar et al., 2005; Lakatos et al., 2007; Stevenson & James, 2009; Holle et al., 2010; Werner & Noppeney, 2010b) offers an alternative approach which not only reliably indicates integration but at the same time maintains ecological validity of the stimulus material. IE in the EEG signal has so far been demonstrated only in one recent study by Senkowski et al. (Senkowski et al., 2011) who used very simplified stimulus material consisting of Gabor patches and sinusoidal tones. The present data show that IE can be observed in the EEG for more complex, naturalistic stimuli. While it remains debatable at the current state of research whether the IE principle, originating from neurophysiology, can be directly applied to more indirect, noninvasive measures such as EEG (or also fMRI) (B. Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009), we believe that it is a promising approach in need of further investigation. With the present study, we take a step in this direction.

6.4.3 Processing Emotional Body Expressions

Finally, we were interested in the processing of body expressions per se. While suppressed oscillatory power in the beta-range has been commonly observed in the perception of biological motion (Cochin et al., 1998; Ulloa & Pineda, 2007; Perry, Troje, & Bentin, 2010), we extend these findings to a modulatory effect of emotional content on the processing of body expressions. This observation is especially relevant as it has been suggested that beta as a correlate of sensory processing and attention, on the one hand, and the mu rhythm (in the same frequency-range) as a correlate of social processing, on the other hand, are distinct (Pineda, 2005). To differentiate between these functions, a separate analysis of occipital (sensory beta) and central electrodes (mu) has been proposed (Oberman, McCleery, Ramachandran, & Pineda, 2007; Perry, Troje, & Bentin, 2010). Our emotion effect for visual and audiovisual conditions was only observed at occipital electrodes, which could suggest that the observed differences may be mainly driven by physical differences in the stimulus material.

A stronger beta-suppression for audiovisual compared to auditory stimuli is in line with previous findings reporting stronger suppression of beta in the perception of biological motion (Muthukumaraswamy et al., 2006; Muthukumaraswamy & Singh, 2008). Localization of the overall beta-suppression – irrespective of the experimental condition – yielded source localization in the premotor cortex. Concomitantly, previous fMRI studies report activation in the premotor cortex for the perception of biological motion (Grèzes & Decety, 2001; Saygin, Wilson, Hagle Jr., Bates, & Sereno, 2004; Wheaton, Thompson, Syngeniotis, Abbott, & Puce, 2004). The premotor cortex has also been suggested to play a key role in the perception of emotional body expressions (de Gelder, 2006b). Thus, this localization of beta-suppression contrasts with a purely sensory interpretation of effects in the sensor space and rather suggests processes specific to the perception of biological motion. Hence, most likely, the observed beta-effects arise from a combination of sensory and biological-motion-specific processes. The exact contributions of these processes will have to be delineated in future studies. Another notable result is the localization of emotion-specific beta-band differences in the right posterior superior parietal cortex in audiovisual emotion perception. This localization is also consistent with the localization results of beta-suppression during biological motion perception in previous studies (Muthukumaraswamy et al., 2006; Muthukumaraswamy & Singh, 2008). It appears that multisensory emotional input is able to elicit differences in areas specifically involved in the processing of biological motion rather than in unisensory visual or auditory areas only. This is another piece of evidence for a modulatory effect of affective information on biological motion processing. Future studies employing higher spatial resolution (fMRI, but also magnetoencephalography to retain temporal resolution) and different experimental designs will be necessary to determine cortical and subcortical loci of an interaction between biological motion and emotional vocalizations more precisely. However, our study provides a valuable starting point by highlighting differences in data interpretation depending on the analysis technique (sensor level versus source space).

6.4.4 Conclusions

Humans rapidly integrate information from various sources, a crucial pre-requisite for successful social interaction. We are able to show that the perception of body expressions is influenced by emotional content of these expressions as well as by accompanying vocalizations. While previous studies have focused on the role of body expressions in multisensory communication in mismatch paradigms (van den Stock et al., 2007, 2008), we show that in congruent settings, interaction effects similar to face–voice interactions can observed. Furthermore, we provide evidence that the inverse effectiveness principle can be used to investigate multisensory integration in EEG data. It provides a valuable tool to study the neural processing of complex, ecologically valid information. Finally, we show that ERPs and neural oscillations, when used complimentarily in the investigation of multisensory integration, allow for a more comprehensive understanding of the interplay between different modalities.

Acknowledgments

The authors would like to thank Heike Boethel for help with the data acquisition, Molly Henry for help with the data analysis, and Kerstin Flake for help with the graphical displays.

Part III

Discussion

Chapter 7

Perception of Emotion in Psychiatric Disorders: On the Possible Role of Task, Dynamics, and Multimodality*

P. Garrido-Vásquez[1], S. Jessen[1,2] & S.A. Kotz[1]
– Both first authors contributed equally –

[1]: Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1A, 04103 Leipzig, Germany
[2]: Cluster of Excellence "Languages of Emotion", Free University of Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany

Experimental evidence suggests an impairment in emotion perception in numerous psychiatric disorders. The results to date are primarily based on research using static displays of emotional facial expressions. However, our natural environment is dynamic and multimodal, comprising input from various communication channels such as facial expressions, emotional prosody, and emotional semantics, to name but a few. Thus, one critical open question is whether alterations in emotion perception in psychiatric populations are confirmed when testing patients in dynamic and multimodal naturalistic settings. Furthermore, the impact task demands may exert on results also needs to be reconsidered. Focusing on schizophrenia and depression, we review

---

*Garrido-Vásquez, P., Jessen, S., & Kotz, S. A. (2011). Perception of Emotion in Psychiatric Disorders: On the Possible Role of Task, Dynamics, and Multimodality. Social Neuroscience, 6(5-6), 515-536. Author Contributions: The General Introduction was written by PG, SJ, and SK, the section on schizophrenia mainly by SJ, the section on depression mainly by PG, and the Final Discussion was written by PG, SJ, and SK.

evidence on how emotions are perceived from faces and voices in these disorders and examine how experimental task demands, stimulus dynamics, and modality may affect study results.

7.1 Introduction

Accurately perceiving others' emotions is crucial for successful interpersonal interactions and may be a key factor in social deficits observed in many neuropsychiatric disorders. During recent decades, a great deal of evidence on emotion perception in psychiatric populations has accumulated. Even though this research has considerably advanced our understanding of how emotions are perceived in people with pathological mental states, we argue that inconsistencies in the results need to be viewed in light of the experimental procedures applied. As we will show, these procedures can be improved by considering higher ecological validity of stimuli, for instance, by using multimodal stimuli (i.e. the combination of face and voice). While an emotional expression in one modality is often enough to determine a person's emotional state, these information sources normally occur simultaneously, providing seemingly redundant information. However, simultaneous perception of congruent emotional information is by no means redundant, but rather offers a number of benefits to the perceiver as will be outlined in this review. These benefits may be especially high when information processing resources are limited, as is often apparent in psychiatric disorders. Along these lines, dynamic rather than static stimuli provide a more ecologically valid approach to test emotion perception. Furthermore, implicit rather than explicit task instructions may affect how emotional expressions are perceived as well as reduce confounds due to cognitive task demands.

The present review aims to promote a more naturalistic approach to emotion perception in patient investigations. To this end, we first present a general introduction to underline the benefits offered by multimodal, dynamic stimulus presentations and implicit task instructions. Second, we introduce models of how emotions are perceived from speech and faces. We then exemplify by focusing on two mental disorders, schizophrenia and major depression, and summarize what has been done so far to understand emotion perception in these two disorders. Finally, we draw disorder-specific and general conclusions from the current state of the art in research. The motivation for focusing on these specific disorders is that they are assumed to tap into different stages of emotion processing: While schizophrenia has been shown to already affect early sensory processes (Johnston, Stojanov, Devir, & Schall, 2005), major depression is considered to alter later processing stages (Gotlib & Joormann, 2010). Likewise, modality, dynamics, and task manipulations may have different effects in these two psychiatric populations which we try to elucidate in the present review. Our literature review concerning schizophrenia and depression is limited to the perception of social emotional cues (i.e. vocal and facial expressions).

7.1.1 Multimodal Emotion Processing

Several studies have shown facilitated emotional information processing when information is simultaneously presented in several modalities (Giard & Peronnet, 1999; Kreifelts et al., 2007; Paulmann et al., 2009). Emotions are recognized faster (Giard & Peronnet, 1999) and more accurately (Giard & Peronnet, 1999; Kreifelts et al., 2007) when simultaneously conveyed by face and voice rather than when presented unimodally.

Furthermore, multimodal perception allows fine-tuning of emotional expression that cannot be achieved unimodally. The voice can have a strong influence on how we perceive the face and vice versa (de Gelder & Vroomen, 2000). If, for instance, an ambiguous facial expression is presented with a happy voice, participants tend to perceive the facial expression as happy, while if the same facial expression is accompanied by sad prosody, the facial expression is also perceived as sad (de Gelder & Vroomen, 2000).

In addition, multimodal perception can ensure that our assessment of someone's emotional state is clear in the case that one of the modalities fails us. The less reliable information from one modality becomes, the more we benefit from information provided in another modality (Werner & Noppeney, 2010b). Consider, for example, being in the dark when you can hardly see your conversational partner but very clearly hear his or her voice.

For these reasons, multisensory perception is an essential mechanism in emotion communication. Yet multisensory perception of emotions is largely understudied in the field of neuropsychiatric disorders. However, it seems crucial to study multisensory emotion perception in patients who suffer from emotional as well as social deficits. It is an open question whether multimodal perception offers the same benefit to these patients as it does to healthy people, or in contrast, may hamper emotion perception.

7.1.2 Models of Processing Emotion from Speech and Faces

One important information source in emotional communication is a person's voice. Three basic steps have been suggested for the processing of emotional speech (Schirmer & Kotz, 2006): a first sensory processing phase, followed by the integration of acoustic cues to form a salient percept, and finally cognitive processes operating on such percepts. In the first step, auditory input is processed in the primary auditory cortices, projecting to the bilateral superior temporal sulcus (STS). These primary features are then integrated in the superior temporal gyrus as well as sulcus, moving along to anterior portions of the STS. Based on these processes, different steps occur dependent on the context, the task, and the exact specification of the stimulus. In emotional speech, emotional prosodic content is accompanied by semantic content. Thus, both types of emotion information need to be compared and aligned. This particular processing step engages the left inferior frontal gyrus. On the other hand, if participants have to just process and label prosodic information, the right inferior frontal gyrus as well as the orbito-frontal cortex seem to play an essential role (Schirmer & Kotz, 2006). A similar model was proposed by Wildgruber, Ackermann, Kreifelts, and Ethofer (2006), positing three steps: the extraction of suprasegmental cues, the representation of suprasegmental sequences, and the explicit judgment of emotional information. These steps correspond essentially to the three steps introduced above, and also involve mostly the same brain regions.

Regarding the processing of emotions from facial expressions, Adolphs (2002b) proposes a 2-step model, dividing the processing into a perception and a recognition part. However, the second step can be further divided into a recognition and a conceptual knowledge aspect, where the recognized emotion is associated with previously known information about the person and the general context. In the first, sensory step, simple as well as highly salient features are processed. This encoding encompasses subcortical regions such as the superior colliculus and the pulvinar nucleus of the thalamus as well as primary sensory cortices, such as the striate cortex. Following this first step, more detailed emotion processing mechanisms that can be subsumed under the term recognition are employed. Relevant facial features are analyzed in visual association cortices. If motion information is contained, the middle and middle superior temporal areas are involved. This information is passed on to the fusiform and the superior temporal cortices where emotional content and social relevance is processed. The amygdala and the orbito-frontal cortex are involved in then guiding various aspects of further emotion processing. First, feedback is directed back to regions implicated in earlier processing stages, allowing fine-tuned processing of the ongoing input. Second, connections to various cortical regions, such as the somatosensory and the prefrontal cortex, as well as the hippocampus, allow the perceived emotions to be put into emotional context based on previous knowledge. Finally, a simulation of another person's emotional state is facilitated by connections to motor structures such as the basal ganglia and the frontal operculum, as well as brainstem nuclei, enabling us to "feel with" somebody (Adolphs, 2002b). Besides these elaborate emotion processing pathways, the pulvinar-amygdalar pathway provides a second, faster route for processing emotion information (Adolphs, 2002b; Vuilleumier & Pourtois, 2007). Here, coarse stimulus features, in particular, are used to quickly determine potentially dangerous situations in order to allow a fast and adequate response.

Overall, processing of auditory as well as visual information can thus be divided into three essential steps, namely early feature processing, integration of these features, and finally evaluation of the percept.

Different regions in the brain have been proposed to be involved in the integration of auditory and visual emotional information, for example, the right posterior STS (Ethofer, Pourtois, & Wildgruber, 2006), the right posterior insula (Ethofer, Pourtois, & Wildgruber, 2006), the amygdala (de Gelder et al., 2004; O'Doherty et al., 2004; Ethofer, Pourtois, & Wildgruber, 2006), and the orbito-frontal cortex (O'Doherty et al., 2004). However, no specific model exists which accounts for the integration of multisensory emotional information, and it remains unclear to what extent an integration process is emotion-specific or rather draws on mechanisms involved in multisensory perception, such as audiovisual speech perception, in general.

7.1.3 Dynamic versus Static Emotional Stimuli

While speech and facial expressions are both inherently dynamic, facial expressions are often tested in static displays. However, our environment is dynamic and the use of dynamic stimuli in emotion research is hence a much more naturalistic approach for studying emotion perception than from photos or drawings of facial expressions.

In fact, dynamic face stimuli are more easily recognized, especially in the case of non-prototypical emotion displays (Ambadar, Schooler, & Cohn, 2005; Bould & Morris, 2008), and emotional deviants are easier to detect in a dynamic visual search task (Horstmann &

Ansorge, 2009). The advantage of dynamic stimuli is further underlined in a steady-state visual evoked potential study (Mayes, Pipingas, Silberstein, & Johnston, 2009). This method, in which stimuli are presented together with a flicker while the electroencephalogram (EEG) is recorded, provides information on cortical activation patterns and their latencies, as well as on inhibitory and excitatory processes. Applying this method, the authors found that processing of dynamic facial stimuli proved to be faster and yielded more activity at temporal electrodes and less activity at frontal electrodes than static ones. This may indicate more efficient processing of dynamic stimuli (Mayes et al., 2009). Furthermore, neuroimaging studies report more extended activation patterns for dynamic than for static facial stimuli. This concerns activation of the middle temporal gyrus, inferior and superior temporal gyri, visual areas, most notably the middle occipital and fusiform gyri, and frontal regions, especially the inferior frontal gyrus (Kilts et al., 2003; LaBar et al., 2003; Sato et al., 2004; Trautmann et al., 2009). Increased middle occipital activation, however, is also observed when viewing dynamic mosaics (Sato et al., 2004) whereas the effects in temporal brain regions are connected to the social relevance of stimuli.

7.1.4 Explicit versus Implicit Task Instructions

In emotion perception experiments, task instructions can be generally divided into two categories: explicit and implicit. Explicit tasks are aimed at the emotional content of stimuli and comprise, among others, emotion categorization (assigning a verbal emotion label to a stimulus), matching of emotional stimuli in terms of the emotion they express, or valence ratings. There is also a wealth of implicit tasks that are not directed towards the emotionality of a stimulus. An example for an implicit measure is the dot-probe paradigm in which a neutral and an emotional stimulus are presented simultaneously at different spatial locations and one stimulus is subsequently replaced by a dot probe to which subjects have to react as quickly as possible. Response latencies in this task can reveal attentional biases, as reactions should be faster when the probe appears in a previously attended location. Other examples of implicit tasks are decisions on physical stimulus characteristics, such as color, or gender decisions.

Some studies have directly compared how identical emotional stimulus material is processed under different task instructions. For example, event-related potential (ERP) studies have shown that explicit and implicit task instructions lead to quantitative differences in face processing. Some researchers reported different scalp distributions and latencies of an N400-like component when comparing identity- and emotion-matching tasks, as well as larger reaction times and error rates in an explicit task (Münte et al., 1998; Bobes, Martin, Olivares, & Valdés-Sosa, 2000), supporting the view that information processing is facilitated under implicit task instructions.

In neuroimaging research, increased amygdala activations in explicit compared to implicit tasks (Gorno-Tempini et al., 2001; R. C. Gur et al., 2002) and also the opposite pattern (Critchley et al., 2000) have been reported. In fact, the amygdala seems to be more sensitive to the absence of a task (i.e. passive viewing) than to differences between implicit and explicit tasks (Costafreda, Brammer, David, & Fu, 2008; but see Fusar-Poli et al., 2009). Furthermore, an absence of task-modulated amygdala effects but heightened responses in the somatosensory cortex have been reported during explicit face processing (Winston, O'Doherty, & Dolan, 2003).

ERP studies on emotional prosody perception also suggest task-dependent differences. Wambacq, Shea-Miller, and Abubakr (2004) reported an early differentiation (P200) between neutral and negative emotional prosody only when emotion was not task-relevant, while the explicit condition yielded a similar but later effect around 360 ms after stimulus onset. In a cross-modal priming study, Schirmer, Kotz, and Friederici (2002) presented an emotional-prosodic stimulus followed by a visually presented word or non-word, with participants required to make a lexical decision. On incongruent trials, the authors found an N400 effect that differed between women and men as a function of the inter-stimulus-interval. In contrast, when the task was to determine the valence of a word, no sex effects were found (Schirmer, Kotz, & Friederici, 2005). Kotz and Paulmann (2007) reported comparable effects when assessing ERP responses to emotional-prosodic and combined prosodic-semantic expectancy violations under implicit and explicit processing instructions. However, ERP amplitudes were overall more positive-going and larger in the explicit task.

Brain activation patterns revealed by neuroimaging methods during emotional prosody processing also depend considerably on study designs, among them task instructions (Kotz, Meyer, & Paulmann, 2006). In fMRI studies, explicit versus implicit emotional prosody processing has been related to increased orbito-frontal and inferior frontal activations (Imaizumi et al., 1997; Wildgruber et al., 2004, 2005; Ethofer, Kreifelts, et al., 2009). Enhanced activations in the right superior temporal region have also been observed under explicit versus implicit task instructions when processing emotional prosody (Wildgruber et al., 2005; Ethofer, Kreifelts, et al., 2009).

To sum up, task demands appear to have a considerable impact on study results, both at the neural level and behaviorally. In many cases, reaction times and/or error rates are lower for implicit compared to explicit tasks, for faces and prosody processing (Münte et al., 1998; Critchley et al., 2000; Gorno-Tempini et al., 2001; Wildgruber et al., 2005; Kotz & Paulmann, 2007; Ethofer, Kreifelts, et al., 2009). However, depending on the nature of the task assigned, an implicit task might be more difficult than one that is explicit (R. C. Gur et al., 2002). Thus, a simple distinction between implicit and explicit task dimensions may be too coarse, but choosing a simple implicit task could significantly reduce task demands. This may be a critical issue when testing patients who suffer from executive dysfunction. High task demands, as often occurring in explicit tasks, can confound executive dysfunction with emotional processing deficits. This could lead to false assumptions about emotional processing deficits in patients (e.g. Adolphs, Schul, & Tranel, 1998). On the other hand, a very simple task, or even more so, the absence of a task, may lead to increased distraction in psychiatric populations and could also alter study results. These facts need to be considered when designing patient studies or interpreting their outcomes.

7.2 Schizophrenia

Schizophrenia is a complex neuropsychiatric disorder comprising numerous symptoms such as hallucinations, delusions, incoherent thought, and blunting of affect, which can occur in basically any combination. While disturbances in the dopaminergic system have been commonly associated with schizophrenia, other pathological changes in the brain seem to also play an important role in the neuronal basis of schizophrenia.

Among the numerous deficits associated with schizophrenia, one often observes abnormalities in early sensory processing and deficits in emotion perception. It has been frequently reported that patients with schizophrenia seem to be unable to filter out relevant sensory information from irrelevant information in a stream of complex sensory input. This phenomenon has been described as a deficit in sensory gating (Freedman et al., 1987), and is also reflected in the electrophysiological brain response. When healthy people hear two consecutive sounds, an early auditory evoked potential (P50) in response to the second sound is reduced. This is commonly interpreted as reflecting the above-mentioned gating mechanism. In individuals with schizophrenia, no such reduction can be observed (Bramon, Rabe-Hesketh, Sham, Murray, & Frangou, 2004). Hence, schizophrenic patients seem to process sensory information in a different way than healthy controls.

At the same time, patients with schizophrenia appear to have difficulties in the processing of emotional information (R. Chan, Li, Cheung, & Gong, 2010). However, emotional information is usually presented in one or the other sensory modality, raising the question to what degree emotion perception in patients with schizophrenia is influenced by a general sensory processing deficit, and what role different modalities play in emotion perception. Do the reported deficits in emotion perception originate from an early, sensory deficit, from a deficit in early emotion processing, or rather at a later stage in emotion processing? Furthermore, differences at several of these stages are conceivable, and may have a differential effect dependent on the precise experimental set-up. An overview of selected emotion perception studies in schizophrenia is provided in Table 7.1.

7.2.1 Schizophrenia and Explicit versus Implicit tasks

Various tasks, such as as emotion identification (e.g. Hempel, Hempel, Schönknecht, Stippich, & Schröder, 2003; Quintana, Wong, Ortiz-Portillo, Marder, & Mazziotta, 2003; Kucharska-Pietura, David, Masiak, & Phillips, 2005; Bach et al., 2009) and emotion matching (e.g. Hempel et al., 2003; Quintana et al., 2003; Martin, Baudouin, Tiberghien, & Franck, 2005; Salgado-Pineda, Fakra, Delaveau, Hariri, & Blin, 2010), have been employed to investigate explicit emotion perception in patients with schizophrenia. While these tasks themselves differ in that one requires explicit labeling while the other only requires a comparison between a number of stimuli, a recent meta-analysis by Kohler, Walker, Martin, Healey, and Moberg (2010) shows that no difference in impairment can be found using these tasks. This suggests that the observed deficits, also in emotion identification tasks, are likely to arise from emotional or at least stimulus processing deficits rather than an impairment in correctly labeling emotions. Regarding implicit emotion processing, gender decision tasks (Johnston et al., 2005; Williams et al., 2007; Lepage et al., 2011), identity matching (Quintana et al., 2003; Martin et al., 2005), and age discrimination (R. E. Gur et al., 2002) have been used primarily in imaging studies, while behavioral studies have focused on the priming paradigm (Höschel & Irle, 2001; Rauch et al., 2010) or the Stroop task (Roux, Christophe, & Passerieux, 2010).

Considering behavioral studies, large differences in emotion processing between patients and controls are observed in explicit and implicit tasks. For instance, in an emotion and also an identity matching task with patients, Martin et al. (2005) reported a performance decrease that was especially severe if both emotion and identity varied. Interestingly, this suggests that patients had difficulties processing only one stimulus dimension while ignoring the other. Furthermore, the performance deficit was larger in the emotion than in the identity task. A similar picture, namely a deficit in both tasks, albeit larger in an emotion identification task, was also found by several other authors (e.g. Sachs, Steger-Wuchse, Kryspin-Exner, Gur, & Katschnig, 2004). While these results suggest a combination of general and emotion specific impairments, other studies provide results pointing clearly in one or the other direction. Kerr and Neale (1993) reported no emotion-specific processing deficit, while Edwards, Pattison, Jackson, and Wales (2001) described a selective impairment in emotional processing. Evidence of both, a general as well as a specific impairment, has been reported for auditory and visual stimuli (Kerr & Neale, 1993; Edwards et al., 2001).

While the pattern of results remains unclear for studies employing explicit tasks, a clearer picture emerges when considering implicit processing tasks. In an affective priming paradigm, Höschel and Irle (2001) demonstrated that both healthy controls and patients with schizophrenia show increased priming by negative stimuli. However, this effect was even more strongly increased in the patient group, suggesting enhanced automatic processing of (negative) emotional information. In contrast, other studies did not find any difference in implicit emotion processing (Roux et al., 2010; Demily et al., 2010; Linden et al., 2010), or an impairment only for patients suffering from negative affective symptoms (Suslow, Droste, Roestel, & Arolt, 2005).

In sum, behavioral studies yield a rather heterogeneous picture regarding a possible emotion processing deficit in schizophrenia. The fact that implicit emotion processing seems less affected than explicit processing may suggest that pre-attentive processing is relatively spared. The reported deficits in explicit tasks may arise from rather late, evaluative processing. Furthermore, the ability to correctly discriminate emotional facial expressions under explicit conditions correlates with general cognitive performance in individuals with schizophrenia (Sachs et al., 2004; Kohler, Bilker, Hagendoorn, Gur, & Gur, 2000) suggesting a link between the performance in emotion discrimination and other tasks.

In order to shed light on the underlying mechanisms and potential deficits of emotion perception, numerous neuroimaging studies have been conducted in recent years. Overall, the results suggest deficits at multiple levels. Several studies have shown that patients with schizophrenia show a decrease in activation in the basal-limbic system, including the amygdala (F. Schneider et al., 1998; R. E. Gur et al., 2002; Hempel et al., 2003; Johnston et al., 2005; Williams et al., 2007; Leitman et al., 2007), the hippocampus (R. E. Gur et al., 2002; Hempel et al., 2003), the anterior cingulate cortex (Hempel et al., 2003; Williams et al., 2007), and the medial prefrontal cortex (MPFC, Williams et al., 2007; Das et al., 2007). Based on these findings, Williams and colleagues suggest a breakdown in the amygdala-MPFC connection to be related to the observed deficits in the processing of emotional, especially of fearful, stimuli. As abnormalities in functional integration of these regions were observed irrespective of whether fearful stimuli were processed consciously or subconsciously, this suggests deficits in early, fast processing as well as disturbed frontal control mechanisms impacting late evaluative processing.

An interesting differentiation between explicit and implicit tasks can be seen when considering a meta-analysis by Li, Chan, McAlonan, and Gong (2010), who report an overall decrease in activation particularly in the amygdala and in the fusiform gyrus in emotion perception. However, while changes in amygdala activation were observed irrespective of task, the fusiform gyrus was less activated only in explicit tasks. Similar results are reported in a recent study by Quintana et al. (2011), who report underactivation in the fusiform gyrus only when attention is directed to emotional features of a stimulus. These findings provide further support for the notion that at least two separate systems are impaired in emotion processing in patients with schizophrenia: a fast, pre-attentive system, involving the amygdala and its surrounding network, and, at least in visual emotion perception, an attention-modulated system, which also seems deficient but is only involved when participants have to consciously process facial features.

Some studies also report differences in sensory areas, such as the middle occipital gyrus (Johnston et al., 2005), in visual emotion perception, both in explicit and implicit tasks, and primary auditory areas for explicit emotion perception from voices (Leitman et al., 2007).

These findings suggest a third aspect that may also affect emotion processing at later stages, namely deficits in an early encoding phase.

A differentiation between an early, purely sensory, and later, more evaluative processing stages is also in line with electrophysiological evidence specifically investigating the time course of visual emotion perception. In addition to an fMRI study, Johnston et al. (2005) conducted an ERP study with the same patient group and paradigm. They reported a reduced vertex positive potential (VPP) amplitude in patients, associated with deficits in the encoding of emotion information. Furthermore, they observe differences in the P3a, which the authors interpret as being associated with the encoding deficits at earlier processing stages. Hence, these results support a more general view in which the encoding of facial features is disturbed, irrespective of specific emotional content. Similar results were observed by Turetsky et al. (2007), who reported N170 amplitude differences in an emotion recognition task, being indicative of deficient structural encoding, which also affects differences at the P3. Support for the hypothesis that early processing deficits are at least partly responsible for emotion processing deficits comes from a study by Kee, Kern, and Green (1998), who found a strong correlation with deficits in visual (and auditory) emotion recognition and measures of neurocognitive function assessing early deficits in perceptual processing. Other studies, however, did not report any N170 differences, but only differences at later emotion processing stages, irrespective of task specification (An et al., 2003; Wynn, Lee, Horan, & Green, 2008).

Overall, behavioral and neuroimaging results suggest deficits in at least two different subprocesses of emotion processing. Explicit and implicit tasks lead to differences in performance and point to separate brain structures underlying the behaviorally observed deficits. Hence, they provide a valuable tool in assessing the interplay between the observed networks.

7.2.2 Schizophrenia and Dynamic versus Static Stimuli

Another potentially important factor influencing visual emotion perception concerns the role of dynamic information; what happens if stimuli are not static pictures but dynamic videos, a situation much more comparable to emotion perception in an everyday life. It has been reported that healthy controls show an improvement in emotion recognition for dynamic over static point-light displays (Atkinson et al., 2004). Furthermore, imaging studies have shown that static stimuli activate different brain regions than dynamic stimuli (Kilts et al., 2003). In particular, static stimuli activate areas associated with motor imagery, for example motor and parietal areas (Decety & Grézes, 1999), suggesting that static stimuli require motion simulation to be processed completely. Dynamic stimuli therefore present an interesting case when investigating emotion perception in patients with schizophrenia. On the one hand, one could expect a decrease in performance, as the processing of dynamic information is an additional processing load and information needs to be integrated correctly in order to be beneficial rather than detrimental. Indeed, Archer, Hay, and Young (1994) showed that patients performed worse in an emotion identification task when stimuli were dynamic rather than static. On the other hand, one can argue that dynamic stimuli are more natural, utilize different brain networks, and lead to a processing benefit in healthy participants, and hence may also be beneficial to patients. Evidence in support of this view comes from a study by Tomlinson, Jones, Johnston, Meaden, and Wink (2006) who observed that patients with schizophrenia showed an overall worse performance in emotion identification, but nevertheless improved when dynamic stimuli were presented. In contrast to Archer et al. (1994), Tomlinson et al. (2006) used point-light stimuli and were thus able to investigate the perception of motion information selectively. Therefore, motion itself seems to provide a benefit to patients with schizophrenia, and the deficits observed by Archer and colleagues are likely to arise from a different stimulus aspect. Determining the interplay between different aspects, and establishing which aspects prove beneficial and which detrimental is particularly necessary when investigating early, sensory-driven emotion perception.

7.2.3 Schizophrenia and Multimodal Emotion Perception

One aspect that needs investigation is the perception of emotional information from multiple sensory modalities at the same time. On the one hand, multimodal presentation of congruent emotional information in healthy controls usually leads to facilitated perception (de Gelder & Vroomen, 2000), on the other hand, it increases the processing demands in terms of sorting out which piece of information from one processing stream is associated with which piece from the other processing stream. As patients with schizophrenia have trouble integrating different sensory streams (Magnée, Oranje, van Engeland, Kahn, & Kemner, 2009), multimodal perception may therefore do more harm than good. Previous studies have shown that patients with schizophrenia seem to be able to accomplish audiovisual integration for very simple sensory input but show strong deficits in more complex, social situations, such as audiovisual speech perception (de Gelder, Vroomen, Annen, Masthof, & Hodiamont, 2003).

A few studies have investigated the multimodal perception of emotional information in patients with schizophrenia (de Gelder et al., 2005; de Jong, Hodiamont, van den Stock, & de Gelder, 2009; de Jong, Hodiamont, & de Gelder, 2010; van den Stock, de Jong, Hodiamont, & de Gelder, in press). Using photos of emotional facial expressions accompanied by matching or mismatching prosodic information, all these studies describe an anomalous integration between auditory and visual information, but no emotion-specific effects. Regarding the exact nature of this anomalous integration, the evidence provided by the studies is inconclusive. When patients are instructed to determine the emotion conveyed by a facial expression and ignore simultaneously presented auditory information, reduced integration compared to healthy controls is observed (de Gelder et al., 2005). When participants are instructed to attend to the auditory modality while ignoring the visual information, the reverse pattern is seen; Schizophrenics then show stronger integration than healthy controls (de Gelder et al., 2005). In the study by de Jong et al. (2009), a slightly modified paradigm was used; again participants were instructed to attend to the auditory modality while ignoring the visual information. This time, decreased integration was found. If, however, an auditory distractor was included in the paradigm, integration increased in patients while it decreased in healthy controls (de Jong et al., 2010). In a fourth study by the same group, increased integration of emotional body language and emotional vocalizations was observed in patients instructed to pay attention to visual information (van den Stock et al., in press).

Overall, the studies described here seem to show that there is an abnormality in the multimodal integration of emotional information, but its exact nature is still unclear. As no comparison to neutral stimuli was made, it is still an open question whether the observed effects are specific to emotions, or are deficits in integration per se.

7.2.4 Emotion Perception in Schizophrenia: Appraisal and Outlook

In summary, task, stimulus dynamics, and multimodality are three factors that are likely to influence emotion perception in patients with schizophrenia. The heterogeneous pattern of results in explicit tasks and the relatively better performance in implicit tasks suggest several underlying mechanisms that may be differentially impaired. If this observation of intact implicit processing can be confirmed, it would necessitate a new discussion about what exactly emotion discrimination tasks are measuring, and whether it can still be claimed that patients with schizophrenia are impaired in emotion perception per se. This assumption is further supported by brain imaging results; at least two separate aspects, namely an early sensory and a late cognitive processing aspect, seem to be affected to different degrees. Further light on the interaction and impairment of these mechanisms can arise by investigating dynamic and multimodal stimuli, two features that in healthy participants lead to facilitated processing. Movement information seems to result in a processing benefit, providing evidence of an intact integration of motion information, probably at early processing steps. Furthermore, multimodal perception seems to be affected. However, the exact extent and emotion specificity remains unclear.

Taken together, it is of relevance for future studies to disentangle these different components in order to gain a clearer understanding of emotion processing deficits in patients with schizophrenia.

7.3 Depression

Major depressive disorder (MDD) is one of the most prevalent neuropsychiatric disorders in Western society. For example, around one sixth of the American population suffers from MDD at least once during life (Kessler et al., 2003). There is evidence that MDD leads to biased emotional processing, which is a maintaining factor of the disorder and may also be present in people at risk or in remission. Understanding how emotion is processed in MDD or MDD-prone individuals could thus be an important key for preventing and treating the disorder.

It should first be considered that depression severely alters cognitive functions. Problems thinking and concentrating are among the diagnostic criteria for MDD according to the Diagnostic and Statistic Manual of Mental Disorders (DSM-IV). MDD patients exhibit slowed information processing, as evidenced by reaction times (e.g. Leyman, Raedt, Schacht, & Koster, 2007) and enhanced latencies in the event-related potential, e.g. the P300 (Vandoolaeghe, Hunsel, Nuyten, & Maes, 1998). MDD patients are also more interference-prone than healthy controls, most clearly manifested in increased error rates (Elliott et al., 1997). These deficits need to be taken into account when running experiments with this patient population as they could considerably influence study results.

Generally, there are three views on emotion processing in depression:

1. negative potentiation posits that MDD patients are biased towards negatively valenced stimuli (e.g. Beck, 1967),
2. positive attenuation considers that processing of positively valenced information is impaired in MDD (e.g. Clark & Watson, 1991), and
3. general blunting of emotional processing irrespective of valence is also suggested (Emotion Context Insensitivity; Rottenberg & Gotlib, 2004).

Importantly, (1) and (2) do not exclude each other and can be present in combination, while (3) is in line with (2), but additionally posits blunted processing of negatively valenced stimuli. An overview of selected studies assessing emotion perception in MDD is provided in Table 7.2.

7.3.1 Depression and Explicit versus Implicit Tasks

A negativity bias in MDD is widely supported by studies applying explicit categorization of facial emotion as patients tend to classify neutral or ambiguous facial expressions as sad or negative (R. C. Gur et al., 1992; Hale, 1998; Hale, Jansen, Bouhuys, & van den Hoofdakker, 1998; Leppänen, Milders, Bell, Terriere, & Hietanen, 2004; P. Luck & Dowrick, 2004; Gollan, Pane, McCloskey, & Coccaro, 2008; Douglas & Porter, 2010; Naranjo et al., 2010), and happy faces as neutral (R. C. Gur et al., 1992; Surguladze et al., 2004). In the latter study by Surguladze et al. (2004), this was limited to happy expressions of 50% intensity, and only at longer stimulus durations (2000 ms instead of 100 ms). Thus, biases in MDD may benefit from long presentation times and non-prototypical emotion expressions providing more scope for an individual's own interpretations. This speaks in favor of biases at later, evaluative processing stages. This evaluative bias fits with the finding that morphed sad faces need less intensity to be categorized as sad by MDD patients while the intensity of happy facial expressions needs to be higher in order to be categorized as happy (Joormann & Gotlib, 2006). Even healthy subjects have been shown to perceive more sadness or rejection in ambiguous line drawings of faces after depressive mood induction (Bouhuys, Bloem, & Groothuis, 1995). Thus, evaluative biases can also occur in the absence of MDD; mood itself may play an essential role. A positivity bias in controls not present in patients has also been reported, reflected by an increased tendency of controls to classify neutral faces as happy (Surguladze et al., 2004; Gollan et al., 2008; Douglas & Porter, 2010).

Results from negative priming, a method in which a former distractor becomes task-relevant in a subsequent trial, revealed deficient inhibition in MDD patients when the distractor was a sad face (Goeleven, De Raedt, Baert, & Koster, 2006; Dai, Feng, & Koster, 2011). This suggests that patients allocated attention to the sad face even though it had to be ignored while the happy face had to be categorized.

The negative bias in MDD previously outlined for faces is also reported for the categorization of vocal stimuli (Kan, Mimura, Kamijima, & Kawamura, 2004; P. Luck & Dowrick, 2004; Naranjo et al., 2010). For example, MDD patients judge negatively-valenced stimuli as more negative than healthy controls (Naranjo et al., 2010). Importantly, stimulus characteristics may play a pivotal role in MDD. Uekermann, Abdel-Hamid, Lehmkämper, Vollmoeller, and Daum (2008) tested emotional prosody categorization and found broad impairments in MDD patients for almost all prosodic stimuli (with neutral semantics and with mismatching semantics). However, no group effects were found when semantics matched the emotional tone conveyed in a sentence. Additionally, the participants' executive functions were related to task performance. Thus, convergent semantic information may largely reduce between-group differences. Reduced impact on executive functions and richer information availability could account for this effect. Likewise, in a study by Emerson, Harrison, and Everhart (1999) examining schoolboys, it was shown that although all participants were better at emotional categorization when semantics and prosody of a sentence matched, the performance decline for mismatching sentences was much more pronounced in depressed than in healthy participants.

Preliminary ERP data from our laboratory on a group of elderly participants without clinically-relevant depression symptoms indicate that increased depression scores affect the integration of emotional-prosodic and semantic information, as reflected in smaller differences in the ERP response to pure prosodic (unintelligible) and normal speech. Based on these results, one may argue that depressive symptoms hamper the integration of information channels, and that the benefit MDD patients showed for semantically and prosodically congruent stimuli in Uekermann et al. (2008) rather reflects richer information availability and/or the absence of distracting information than integration per se.

Taken together, the results on emotional prosody perception in depression confirm the presence of a negative evaluation bias and thus extend the findings from facial expressions to speech, although stimulus characteristics may play an additional role in these findings.

Apart from explicit tasks, implicit methods have also been used at the behavioral level, most notably the dot-probe task (described in the introduction). At face presentation times of 1000 ms, an attentional bias towards sad faces has been reported for MDD patients (Gotlib, Krasnoperova, Yue, & Joormann, 2004; Gotlib, Kasch, et al., 2004; Joormann & Gotlib, 2007; Fritzsche et al., 2010), and this bias was preserved in remitted patients (Joormann & Gotlib, 2007; Fritzsche et al., 2010). Even in never-depressed daughters of depressed mothers, this effect was found after negative mood induction (Joormann, Talbot, & Gotlib, 2007; Kujawa et al., 2010). A related method, spatial cueing, revealed prolonged attending to angry versus neutral faces in MDD compared to controls, suggesting that the bias extends beyond expressions of sadness when participants are instructed to attend to the faces (Leyman et al., 2007). By contrast, in dot-probe studies including anger, the bias appears to be sadness-specific (Gotlib, Krasnoperova, et al., 2004; Gotlib, Kasch, et al., 2004); thus, instructions that direct attention may be significant. Some dot-probe studies also report a positivity bias in healthy controls not present in MDD or MDD-prone individuals (Joormann & Gotlib, 2007; Joormann et al., 2007; Fritzsche et al., 2010). These studies demonstrate that depression leads to preferred attention towards sad, or more generally, negatively-valenced faces and away from positive expressions at late processing stages, even when emotion is not task-relevant.

In short, both explicit and implicit behavioral studies revealed biased emotion processing in depression, and not only evaluative but also attentional biases are supported by the literature.

Electrophysiology has also yielded interesting results: Dai et al. (2011) reported abnormal ERP responses to sad faces in MDD already occurring around 100 milliseconds after stimulus onset when the face was preceded by a sad target that had to be categorized. Thus, after judging a face as sad, this kind of stimulus is processed abnormally in MDD, which can be observed during early sensory processing. Importantly, sad faces preceded by a happy target did not provoke such early effects, which speaks against an early sensory deficit or bias. In an electromyographic study (Sloan, Bradley, Dimoulas, & Lang, 2002), students who were dysphoric revealed reduced activity of the zygomaticus, a facial muscle which responds to positive stimuli, during the explicit categorization of happy faces. Instead, they showed more corrugator activity for happy faces, which is normally correlated with processing negative stimuli. In fact, positive and negative facial expressions were processed as if they were all negative in dysphorics, even though categorization performance was comparable to controls and thus there was no evaluative bias. This study, which temporally covers the seconds range, is complemented by an ERP study with millisecond resolution, in which students who were depressed exhibited reduced processing especially of mildly happy faces, starting around 350 milliseconds after face onset (Cavanagh & Geisler, 2006).

Taken together, results from electrophysiological studies using explicit tasks indicate that subclinical depression leads to reduced positive reactivity, and that sad stimuli are processed abnormally in MDD when attention has previously been drawn to them. The time course suggests that biases already emerge before evaluative decisions but no early, sensory processing deficits are confirmed.

ERP studies applying implicit tasks also support altered processing of facial emotion in MDD. Deldin, Keller, Gergen, and Miller (2000) observed a generally attenuated N200 component at right-posterior electrode sites in MDD, an effect which was strongest in response to happy facial expressions. In another study, MDD patients did not differ in their P300 response to previously seen happy versus sad faces while controls did (Deldin, Keller, Gergen, & Miller, 2001). The mismatch negativity (MMN), a measure of pre-attentive change detection, was not evident in MDD patients from around 220 milliseconds post-stimulus onset onwards, as their ERP response to sad, happy, and neutral schematic faces did not differ. However, the MMN was present before this time window, as neutral faces initially elicited brain responses that differed from the two emotional categories (Chang, Xu, Shi, Zhang, & Zhao, 2010). This means that early perceptual encoding including the face-related N170 component may be largely intact, and that blunting takes effect at a stage in which emotional significance is supposed to be extracted. Foti et al. (2010) reported a reduced differentiation of fearful and angry faces from neutral ones in the event-related potential in MDD at a late processing stage (late positive potential). This further supports the view of emotional blunting in MDD although it may also reflect the fact that MDD patients are more likely to perceive neutral faces as negative. Blunting is also supported in a study by J. Kayser, Bruder, Tenke, Stewart, and Quitkin (2000), who presented unpleasant (wounded) and neutral (intact) faces and reported a lack of a differentiation of these two stimulus types in MDD in a comparable time window, suggesting generally reduced emotional reactivity. Deveney and Deldin (2004) assessed memory retention of emotional faces during several seconds (slow-wave). While this component was increasingly reduced for sad relative to neutral and happy faces in healthy participants, no differences emerged in the depressed group, suggesting that MDD patients fail to disengage from sad faces over time. This is nicely complemented by an fMRI study on memory retention of positively and negatively valenced pictures: Successful retention of the latter was associated with enhanced amygdala activation during encoding in MDD patients (Hamilton & Gotlib, 2008); thus, amygdala hyperactivity might play a role in the enhanced retention of negative stimuli.

To sum up, electrophysiological studies indicate that after a few hundred milliseconds from the onset of a face, emotion perception is blunted in MDD, reflected in reduced neural differentiation between positive, negative, and neutral faces. Negative potentiation, in contrast, seems to occur in the seconds range and may reflect a lack of attentional deployment from negative stimuli. This can be observed with explicit and implicit tasks.

In fMRI studies, there is some indication that explicit compared to implicit tasks may reduce group effects at the neural level. Studies using implicit tasks or passive viewing provide ample evidence of negative potentiation, i.e. heightened neural responses in MDD compared to controls when individuals are presented with negative facial expressions (Sheline et al., 2001; Fu et al., 2004; Gotlib et al., 2005; Surguladze et al., 2005; Fu et al., 2008; Frodl et al., 2009; Victor, Furey, Fromm, Ohman, & Drevets, 2010, but see Lawrence et al., 2004; Lee et al., 2008). This extends to remitted patients (Neumeister et al., 2006) and individuals at genetic risk of MDD (Dannlowski et al., 2008; Hariri et al., 2005). Positive attenuation has also been corroborated by neuroimaging studies using implicit tasks (Lawrence et al., 2004; Surguladze et al., 2005; Fu et al., 2007; Victor et al., 2010).

In contrast, a study using emotional categorization reported no significant activation differences between MDD and control participants, for either positive or negative emotional faces (Almeida, Versace, Hassel, Kupfer, & Phillips, 2010). This fits well with Frodl et al. (2009), who observed that negative potentiation was greater when using a gender matching instead of an emotion matching task. It is also in line with a study by Monk et al. (2008) comparing passive viewing with different tasks in adolescents at risk of depression. While the data obtained from passive viewing support negative potentiation and positive attenuation, the data from the constrained attention conditions do not. The findings from Monk et al. also suggest that a challenging implicit task may diminish group differences. Thus, what has to be evaluated is probably not the explicit–implicit distinction but rather the potential cognitive load introduced by the task. Stronger frontal activations in the risk compared to the non-risk group when performing the tasks support this notion (Monk et al., 2008). Along these lines, an easy explicit task may also give rise to negative potentiation in MDD, compared to a more challenging one (van Wingen et al., 2011).

The amygdala, in particular, seems to play a central role in MDD, as most of the studies reporting negative potentiation find enhanced activations in this structure. Moreover, Dannlowski et al. (2007) observed that amygdala activation to sad and angry faces correlated positively with behaviorally-manifested negative biases in MDD. Connectivity studies with depressed individuals have shown abnormal interactions of prefrontal regions with brain structures attributed to emotional processing. More specifically, connections between the amygdala and prefrontal cortex appear to be impaired in MDD (Siegle, Thompson, Carter, Steinhauer, & Thase, 2007; Dannlowski et al., 2009), a connection which has been implicated in successful emotion regulation (Banks, Eddy, Angstadt, Nathan, & Phan, 2007). Low cognitive load, as in easy implicit tasks, may exacerbate the impact of this disturbed circuit, enabling excessive amygdala activations in response to negative stimuli. Findings from the cognitive domain indicate that tasks which strongly bind attention and suppress rumination, a core feature of the disorder, have been associated with reduced between-group differences (see Gotlib & Joormann, 2010, for a review).

7.3.2 Depression and Dynamic versus Static Stimuli

Categorization studies using dynamic facial expressions have yielded promising results. Kan et al. (2004) tested videos of six emotions (happiness, surprise, anger, disgust, fear, and sadness) and found no group effects in recognition performance, a result which is in line with a recent study (Schaefer, Baumann, Rich, Luckenbaugh, & Zarate Jr, 2010). One possible explanation for these results is that the information provided by dynamic stimuli is richer than from static displays and makes it easier for patients to recognize them, as discussed by the authors of both studies. Another explanation is that emotional information provided in moving stimuli that is continuously changing prevents MDD patients from focusing on a facial expression and starting to ruminate about it or project negativity onto it. This suggestion is in line with intact early processing of emotional cues and biases at later, more cognitive processing stages. In any case, these study results are promising and will have to be corroborated by further investigations.

However, as outlined in the previous section, impairments and negative biases have been reported in studies on emotional prosody categorization in MDD (Kan et al., 2004; P. Luck & Dowrick, 2004; Uekermann et al., 2008; Naranjo et al., 2010), and prosody is also dynamic in nature. In the realm of an apparent dissociation between dynamic speech and face stimuli, it must be considered that recognition performance is generally higher for emotions conveyed by faces than by voices (e.g. Kan et al., 2004). Likewise, cognitive impairment associated with MDD could potentiate challenging task demands in explicit tasks (i.e. prosody categorization), while congruent semantic information may reduce cognitive demands and decrease group differences (Emerson et al., 1999; Uekermann et al., 2008).

To sum up, dynamic facial expressions as well as prosodic stimuli with matching semantics have led to promising results in MDD. However, even in the absence of behavioral differences, there may still be alterations at the neural level (e.g. Sloan et al., 2002), as explicit categorization is not very informative about earlier processing steps. More research is needed to shed light on these open issues.

7.3.3 Depression and Multimodal Emotion Perception

To our knowledge, so far there are no studies on multimodal emotion perception in MDD patients. There is, however, one ERP experiment by Campanella et al. (2010) assessing a group of students displaying elevated but subclinical anxiety and depression scores. In an emotional target-detection oddball paradigm, they reported a reduced P300 amplitude when compared with students with low scores in these measures. Interestingly, this was only the case in the multimodal (prosodic speech cue and static facial expression) condition irrespective of emotion, but not in the unimodal conditions. Moreover, the P300 amplitude was negatively correlated with depression scores. This suggests that under multimodal input conditions, emotion processing deficits could be more likely to take effect than under unimodal input. However, this is but one experiment in a subclinical population, and one can only speculate about multimodal emotion processing in MDD. Furthermore, in the experiment, a static facial expression was combined with a prosodic stimulus. Even though these were matched in terms of the emotion they expressed, the static-dynamic combination represents a mismatch, and mismatching information may affect depressive individuals more strongly than controls (Emerson et al., 1999; Uekermann et al., 2008). As long as multimodal information is congruent, MDD patients should benefit from it, at least when it comes to explicit recognition tasks, while neural correlates of multimodal stimuli still need to be elucidated in MDD.

7.3.4 Emotion Perception in Depression: Appraisal and Outlook

While a considerable amount of work targeting facial emotion processing in MDD has accumulated in both implicit and explicit tasks, studies on emotional prosody or multimodal emotion displays are largely or completely missing to date. Furthermore, face stimuli have been presented statically rather than dynamically in the vast majority of studies.

As outlined in the introduction, emotion processing is not a unitary process but rather involves different processing steps in which different tasks are accomplished and different neural correlates may be engaged. Studies with MDD patients have helped to shape our understanding of how depression may influence emotion processing from faces. ERP studies indicate that processing alterations in MDD may start around a few hundred milliseconds after stimulus onset, where there is evidence of general blunting, as neural response patterns fail to distinguish stimuli according to their emotional content (Foti et al., 2010; Chang et al., 2010). At around one second post face onset, patients have directed their attention towards sad stimuli (Gotlib, Krasnoperova, et al., 2004; Fritzsche et al., 2010) and away from happy stimuli (Joormann & Gotlib, 2007; Joormann et al., 2007) even when faces are not task-relevant. This attentional focus then persists, as suggested by two ERP studies looking at late processing stages (Deveney & Deldin, 2004; Dai et al., 2011). When prompted to attend to the emotionality of a face, happy expressions may be processed as if they were negative (Sloan et al., 2002), a result which could explain biases evident in explicit categorization experiments. Thus, it seems that depression does not lead to preferential attention towards negative or attention deployment from positive stimuli, but once negative stimuli have captured a patient's attention, the patient cannot disengage from them (Gotlib & Joormann, 2010). This may be especially the case when an experimental task is not challenging (Frodl et al., 2009) and does not provide sufficient distraction to suppress rumination.

Even though there is no convincing evidence so far that MDD biases early perceptual encoding of facial emotion, support for early alterations has been provided by three recent fMRI studies using subliminally-presented masked face primes which revealed group-related activation differences in the amygdala (Sheline et al., 2001; Suslow et al., 2010; Victor et al., 2010). Another study reported similar results only when dividing the sample into genetic risk and non-risk groups rather than diagnostic groups (Dannlowski et al., 2008). The effects may correspond to the fast emotional face processing pathway directly feeding into the amygdala (Adolphs, 2002b; Vuilleumier & Pourtois, 2007). However, neutral faces were used to mask the emotional faces in these studies. Due to the low temporal resolution of fMRI, it cannot be ruled out that the activations reflect the processing of the neutral mask, influenced by emotional prime valence. Experiments using non-facial masks and the application of techniques providing higher temporal resolution, such as EEG, may be useful tools to further address this issue. ERP studies, despite their excellent temporal resolution, have so far failed to find alterations in early sensory processing.

In sum, at this point, we do not know whether depression affects early sensory processing. Evidence on later processing stages is much clearer, although the fact that the vast majority of studies are based on static face stimuli must be considered.

7.4 General Discussion

Accurately perceiving others' emotions is an essential component of successful social interaction, a skill, which is impaired in many if not all psychiatric disorders. However, emotion perception is not a single-step process but rather a complex mechanism involving several interacting subcomponents.

As elaborated in the introduction, at least three such sub-steps can be distinguished: (1) sensory processing, (2) integration of sensory cues, and (3) evaluation of the perceived cues (Adolphs, 2002b; Schirmer & Kotz, 2006). In principle, all of these steps are vulnerable to malfunction, raising the question at which processing stage the observed deficits in a given neuropsychiatric disorder may occur.

Here, schizophrenia and depression are of particular interest. While a common finding in patients with schizophrenia is the presence of early sensory processing deficits, depression is characterized by deficits at later, cognitive processing stages. This dissociation in general deficits suggests that the processing of emotional information expressed in faces and voices may also result in specific processing deficits. Thus, while at first glance, emotion perception appears disturbed in both disorders, the underlying mechanisms leading to these alterations may be different. It is therefore conceivable that deficits arise in several processing steps, or from a different mechanism altogether, such as cognitive deficits affecting response behavior. Furthermore, it is open to discussion to what extent observed deficits are modality-specific or supramodal, and whether multimodal emotion perception can alleviate symptoms or, in contrast, lead to a decrease in performance. All these questions cannot be fully answered by the current state of the art in psychiatric research on emotion expressions but the evidence reviewed here offers the possibility for some justified speculation, motivating further research.

Current studies suggest that emotion processing deficits occur at various processing levels. For schizophrenia, in particular, studies show very early processing differences (Johnston et al., 2005; Leitman et al., 2007), suggesting that deficits already occur at the level of sensory encoding. Considering the well-known problems in sensory gating that are often described as one characteristic finding in patients with schizophrenia, this raises the question to what extent these deficits are emotion-specific or rather a result of pathological early sensory encoding per se. In depression, evidence on deficits in early perceptual encoding is, to date, not convincing (e.g. Gotlib & Joormann, 2010).

Regarding later processing stages, emotional processing deficits have been reported for both patient populations (An et al., 2003; Deveney & Deldin, 2004; Wynn et al., 2008; Dai et al., 2011). While emotional valence appears to play a role in depression, as positively valenced material is attenuated and negatively valenced material is processed in an enhanced manner, no such distinction can be described for schizophrenia patients.

These patterns suggest that the observed deficits in emotion processing arise from different underlying mechanisms. The observations regarding MDD speak in favor of an emotional bias, as processing differences are found at later stages and appear to be valence-driven. In schizophrenic patients, however, the picture is not that clear. While there is evidence for deficits at later evaluative steps, disturbances in early sensory processes are also commonly observed. On the one hand, this complex pattern may indicate that two separate processing mechanisms are affected. On the other, the late processing deficits could also be interpreted as a consequence of deficits at earlier stages. Thus, alterations in emotion processing per se seem more likely in depression than in schizophrenia.

To shed more light on emotional processing in these two disorders, it seems important to more closely examine several factors influencing emotion perception. Here, we focused on the three aspects – task, stimulus dynamics, and multimodality – as they seem to be especially relevant to emotion perception (e.g. Atkinson et al., 2004; Giard & Peronnet, 1999). Distinguishing between explicit and implicit tasks is important when investigating patient populations, as explicit task instructions may introduce differences between patient and healthy populations which are unrelated to emotional processing, but rather reflect cognitive deficits. In fact, schizophrenia patients generally perform worse in emotional as well as control tasks, suggesting a strong effect of cognitive factors (Kerr & Neale, 1993; Pomarol-Clotet et al., 2010). This is supported by the finding that difficulties are more pronounced in explicit than in implicit emotional processing tasks (Suslow et al., 2005; Linden et al., 2010). Patients with depression seem to be less affected by task settings and may even benefit from challenging tasks at the neural level (Frodl et al., 2009).

A second aspect to be taken into consideration is multisensory emotion perception. In healthy participants, providing congruent emotional information via several modalities usually leads to facilitated processing (Giard & Peronnet, 1999; Paulmann et al., 2009). Can similar benefits be observed in patient populations? Or do deficits rather increase, as multisensory stimuli are more complex than unisensory, thus requiring more elaborate processing? While investigating multisensory emotion perception may thus provide crucial insights into the relation between sensory and affective deficits, this dimension has hardly been investigated in patient populations. While no studies are known that have investigated multimodal emotion perception in MDD patients, a few studies have addressed this issue at the behavioral level in patients with schizophrenia (de Gelder et al., 2005; de Jong et al., 2009, 2010; van den Stock et al., in press). However, no clear picture results from these studies with respect to whether these patients can benefit from the multimodal perception of emotional information. For both disorders, it remains an open question whether multimodal emotional input leads to improved processing or introduces new problems.

A last aspect, also improving emotion perception in healthy controls, is the use of dynamic visual stimuli. Dynamic visual stimuli have higher ecological validity and provide more information than static pictures. In relation to the face stimuli used, it has been shown that, just as healthy controls, patient groups benefit from dynamic information (Atkinson et al., 2004; Kan et al., 2004; Tomlinson et al., 2006; Schaefer et al., 2010). Nevertheless, they are not widely used but should be considered for further research aiming at more naturalistic stimulus displays.

To conclude, emotion perception in patient populations is influenced by numerous factors, such as task specificity and stimulus dynamics. However, it seems that patients with schizophrenia and depression are affected differentially; while in schizophrenic patients, emotion processing deficits occur at early as well as later stages in processing in patients with schizophrenia, MDD patients show impairments primarily at later evaluative stages. Acknowledging this fact could not only affect our understanding of a given disorder but may also play a crucial role in the correct assessment and intervention.

Acknowledgements

The authors would like to thank three anonymous reviewers for helpful comments on an earlier version of this manuscript. We also thank Rosie Wallis for carefully checking our English.

Table 7.1: *Comparison of Selected Studies Investigating Emotion Perception in Schizophrenia.*

	Participants		*Medication*				*Emotions*								
Authors	*Patients*	*H. Controls*	*Medicated*	*Unmedicated*	*Method*	*Stimuli*	*Anger*	*Fear*	*Sadness*	*Happiness*	*Disgust*	*Neutral*	*Other*	*Task*	*Results*
de Gelder et al. (2005)	13	13	13	0	Behav	AV			+	+				e	Weaker influence of voice but stronger influence of face
de Jong et al. (2009)	55^1	50	52	0	Behav	AV		+	+	+				e	Weaker influence of face
de Jong et al. (2010)	55^1	50	52	0	Behav	AV		+	+	+				e,i	Stronger integration when auditory distractor present, no impairment by visual distractor
Linden et al. (2010)	34	34	30	4	Behav	V	+			+		+		e,i	No difference for i task but impairment for e task
Suslow et al. (2005)	88^2	30	88	0	Behav	V			+	+		+		e	Differences only for patients with affective symptoms
Tomlinson et al. (2006)	16	24	?	?	Behav	V^3	+	+	+	+	+		+	e	Motion benefit for both groups, but smaller for patients
van den Stock et al. (in press)	31	21	28	3	Behav	V^4	+	+	+					e	Deficits for perception of body expressions
	16	16	16	0	Behav	AV^4		+		+				e	Increased audiovisual integration
Das et al. (2007)	14	14	10	4	fMRI	V		+				+		p	Reduced activation/ differences in connectivity of AMG
Johnston et al. (2005)	11	15	11	0	fMRI/ EEG	V	+	+	+	+	+	+	+	e, i	ERP: reduction of VPP and P3; fMRI: reduced activity of FFG
Salgado-Pineda et al. (2010)	14	14	14	0	fMRI	V	+	+						e	Reduced activation of AMG for sustained stimulation
An et al. (2003)	20^5	20	15	11	EEG	V						+	+	i	Reduced P3
Turetsky et al. (2007)	16	16	10	6	EEG	V			+	+		+		e	Larger N170
Wynn et al. (2008)	26	27	26	0	EEG	V	+	+	+	+			+	e, i	Reduced N250

Notes: H. Controls: healthy controls; Behav: behavioral; V: visual; A: auditory (prosody); e: explicit; i: implicit; p: passive viewing; FFG: fusiform gyrus; AMG: amygdala. 1: 53 paranoid, 2 residual; 2: 30 schizophrenic patients with flat affect, 30 schizophrenic patients with anhedonia, 28 schizophrenic patients without affective symptoms; 3: point-light faces; 4: body language; 5: 16 paranoid, 4 undifferentiated.

Table 7.2: *Comparison of Selected Studies Testing Emotion Perception in Major Depression.* Note that due to a very high number of facial emotion categorization studies in the literature, they are treated exclusively in the text. In the case of treatment studies, only results at baseline are reported

	Participants		*Medication*				*Emotions*								
Authors	*Patients*	*Controls*	*Medicated*	*Unmedicated*	*Method*	*Stimuli*	*anger*	*fear*	*sadness*	*happiness*	*disgust*	*neutral*	*other*	*Task*	*Results*
Fritzsche et al. (2010)	20	20	2	18	Behav	V			+	+		+		i	Attentional bias toward sad and away from happy faces
Goeleven et al. (2006)	20	20	?	?	Behav	V			+	+				e	Reduced negative priming of sad faces, indicating reduced inhibition
Gotlib, Kasch et al. (2004)	88	55	?	?	Behav	V	+		+	+		+		i	Attentional bias toward sad faces
Gotlib, Kransnoperova et al. (2004)	19	16	8	11	Behav	V	+		+	+		+		i	Attentional bias toward sad faces
Joormann et al. (2007)	26	19	16	10	Behav	V			+	+		+		i	Attentional bias toward sad and away from happy faces
Kan et al. (2004)	16	20	15	1	Behav	A	+	+	+	+	+		+	e	Tendency to categorize surprise as negative
Leyman et al. (2007)	20	20	16	4	Behav	V	+					+		i	Reduced attentional disengagement from angry faces
Luck et al. (2004)	49	30	0	49	Behav	A	+		+	+		+		e	More negative classifications
Naranjo et al. (2010)	23	23	23	0	Behav	A	+	+	+	+	+	+	+	e	Tendency to categorize neutral and surprise as negative
Uekermann et al. (2008)	29	29	12	17	Behav	A	+	+	+	+		+		e	Impaired for all subtests except for stimuli with congruent semantics and prosody; all categories except sadness affected
Almeida et al. (2010)	15	15	13	2	fMRI	V		+	+	+		+		e	No group differences
Dannlowski et al. (2008)	28	28	28	0	fMRI	V[1]	+		+	+		+		p	No group differences; increased AMG response in risk allele carriers

Frodl et al. (2009)	12	12	8	4	fMRI	V	+		+			e, i	Higher neural response, reduced with explicit task
Fu et al. (2004)	19	19	0	19	fMRI	V^2			+			i	Stronger neural response, slower RT
Fu et al. (2007)	19	19	0	19	fMRI	V^2				+		i	Range of neural response to different intensities reduced in limbic-subcortical and visual areas
Fu et al. (2008)	16	16	0	16	fMRI	V^2			+			i	Higher overall AMG-hippocampus activity, stronger reaction to intensity increase
Gotlib et al. (2005)	18	18	9	9	fMRI	V			+	+	+	i	Stronger neural response to sad and happy
Lawrence et al. (2004)	9	11	9	0	fMRI	V^2		+	+	+	+	i	Weaker neural response to all faces, subcortically and frontally
Lee et al. (2008)	21	15	10	11	fMRI	V	+		+		+	p	Weaker neural response, subcortically and frontally
Sheline et al. (2001)	11	11	0	11	fMRI	V^1		+		+	+	p	Greater AMG activation to emotional faces
Surguladze et al. (2005)	16	14	16	0	fMRI	V^2			+	+	+	i	Happy: reduced activation in FFG and putamen; sad: increased activation in FFG, putamen, parahippocampal gyrus/ AMG
Suslow et al. (2010)	30	26	30	0	fMRI	V^1			+	+	+	e	Stronger AMG reaction to sad primes; reduced to happy
van Wingen et al. (2010)	18	30	0	18	fMRI	V	+	+				e^3	Increased activation in AMG, IFG, and ACC to labeling task
Victor et al. (2010)	22	25	0	22	fMRI	V^1			+	+	+	i	Stronger AMG reaction to sad primes; reduced to happy; faster RT to sad primes
Chang et al. (2010)	15	15	0	15	EEG	V			+	+	+	i	No difference between neutral and emotional stimuli from around 220 ms after onset
Dai et al. (2011)	19	20	?	?	EEG/ Behav	V			+	+		e	Enhanced P1 to sad face preceded by sad target; no negative priming for sad faces
Deldin et al. (2001)	19	15	14	5	EEG	V			+	+	+	i	Abolished P300 difference between previously seen sad and happy faces
Deveney & Deldin (2004)	17	17	6	11	EEG	V			+	+	+	i	No slow wave amplitude reduction for sad faces
Foti et al. (2010)	19	25	0	19	EEG	V	+	+			+	p	No difference of emotional versus neutral faces in late positive potential

Notes: H. Controls: healthy controls; Behav: behavioral; V: visual; A: auditory (prosody); e: explicit; i: implicit; p: passive viewing; ACC: anterior cingulate cortex; IFG: inferior frontal gyrus; FFG: fusiform gyrus; AMG: amygdala; RT: reaction time. 1: subliminally presented, masked faces; 2: morphed to different emotional intensities; 3: two explicit tasks: matching of facial expressions, matching of facial expression to verbal label.

Chapter 8

General Discussion and Outlook

8.1 Summary of Empirical Studies

The aim of the current thesis was to investigate the neural bases that underlie the multisensory perception of other's emotions. In particular, the perception of anger and fear from vocal and body expressions was investigated by using dynamic displays of emotional information. To this end, two EEG studies and one fMRI study were conducted, flanked by behavioral pilot studies. The experimental paradigm was comparable for all three studies; participants were presented with emotional information in three different modality conditions. In an auditory condition, the participants heard emotional or neutral interjections; in a visual condition, they saw short video clips showing emotional or non-emotional body expressions; and in an audiovisual condition, they saw video clips and heard matching interjections.

In the first EEG study, the auditory N100 was strongly reduced in amplitude for audiovisual compared to auditory stimuli. Based on comparable results by Stekelenburg and Vroomen (2007), this reduction can be interpreted as a very early facilitation in processing due to an interaction between predictive visual and auditory information. Likewise, a reduced N100 amplitude was observed for emotional compared to neutral stimuli, again indicative of facilitated processing (Besle et al., 2004; van Wassenhove et al., 2005). An interaction between both factors revealed that in the auditory but not in the audiovisual condition, an additional differentiation between anger and fear stimuli occurred. This provides first evidence that unimodal stimuli may necessitate a more detailed processing in comparison to audiovisual stimuli. In addition to these ERP results, a time–frequency-analysis was computed, which revealed that visual and audiovisual stimuli result in highly similar processing patterns in the alpha- and beta-band.

While a very early influences of modality as well as affective content was observed in the first EEG study, one can draw only limited conclusions about involved brain areas due to the poor spatial resolution of EEG. Therefore, a very similar paradigm was implemented in an fMRI study in order to gain closer insights into the neural networks involved. For both unimodal conditions, activations were observed in areas commonly reported in emotion processing from body and vocal expressions, respectively (Grèzes et al., 2007; Pichon et al., 2008, 2009). For emotional compared to nonemotional body expressions, an increased BOLD response was found in occipital areas, the fusiform gyrus, the extrastriate body area, and the hippocampus. In the auditory condition, emotional stimuli elicited an increased response in the amygdala, the inferior frontal gyrus and the anterior cingulate cortex.

In the audiovisual condition, a strong decrease in activation was found for emotional stimuli, particularly in the bilateral STG/STS. At the same time, activity in the fusiform gyrus was increased for emotional stimuli. Hence, the preceding and ongoing visual information seems to result in a reduced auditory activation for emotional stimuli.

Differential effects for emotional and neutral stimuli also are visible in the conjunction analysis computed to investigate the multisensory perception of auditory and visual information. While neutral stimuli elicited a clear activation in a network consisting of bilateral STG, bilateral thalamus, and bilateral cerebellum, hardly any activation was observed for emotional stimuli. Hence, the results of the fMRI study extend previous studies on audiovisual emotion perception by showing that a combination of emotional and multimodal processing may indeed result predominantly in a decrease, rather than an increase in BOLD response in the STG/STS.

Both, the first EEG study and the fMRI study clearly show an interaction between auditory and visual processing of body and voice information. However, it has been a longstanding debate in the field of multisensory perception research how such an interaction between two modalities can be qualified as integration. Integration is defined as a truly combined process rather than two parallel processing streams. It has been suggested that a true integration should be enhanced under suboptimal sensory conditions, a criterion that has been termed “inverse effectiveness” (B. Stein & Meredith, 1993). While this criterion has been amply employed in fMRI studies (Holle et al., 2010; Werner & Noppeney, 2010b), it had only been used in one

EEG-study, investigating the integration of perceptually less complex material (Senkowski et al., 2011). In the second EEG study, the signal quality of the auditory stimuli was therefore manipulated to investigate audiovisual interactions at high and low noise levels. Only in the high-noise condition, smaller N100 peak latencies were observed for audiovisual compared to auditory stimuli. This result confirms the predictions of the inverse effectiveness principle and suggests a true integration of body and voice information. Further support for this assumption arises from the pattern observed in the beta-band suppression. Larger suppression differences between audiovisual and auditory stimuli were found in high-noise emotional conditions compared to low-noise emotional conditions. No such difference occurred for nonemotional/neutral stimuli. Not only is this observation predicted by the inverse effectiveness principle, it additionally suggests a modulation of the integration process by emotional content.

Overall, the data in the thesis demonstrate an interaction between body and voice information, which occurs early in processing and is characterized by an integration between both modalities. Furthermore, a clear effect of emotional content was found on the processing of body and voice information. On a neural level, this emotion processing invokes a network consisting of early modality-specific as well as later, modality-independent areas involved in the evaluation of emotional content. Furthermore, the data suggest different processes underlying the multisensory perception of emotional and nonemotional information.

8.2 Discussion

8.2.1 Visual Information Influences Auditory Processing

Figure 8.1 schematically summarizes the findings of the thesis. It contrasts differences in emotional processing for unisensory auditory information versus audiovisual information. Broadly speaking, processing of emotional information can be categorized into an *early, sensory*, and a *later, evaluative* phase, as has been previously suggested by various models on emotion processing (Adolphs, 2002b; Schirmer & Kotz, 2006). In the current studies, early processing was predominantly indicated by changes in the N100 amplitude. Differences were found between auditory and audiovisual processing; while for example in EEG-study 1, no difference between anger and fear stimuli was observed in the audiovisual condition, fear stimuli elicited significantly smaller N100 amplitudes in the audio condition.

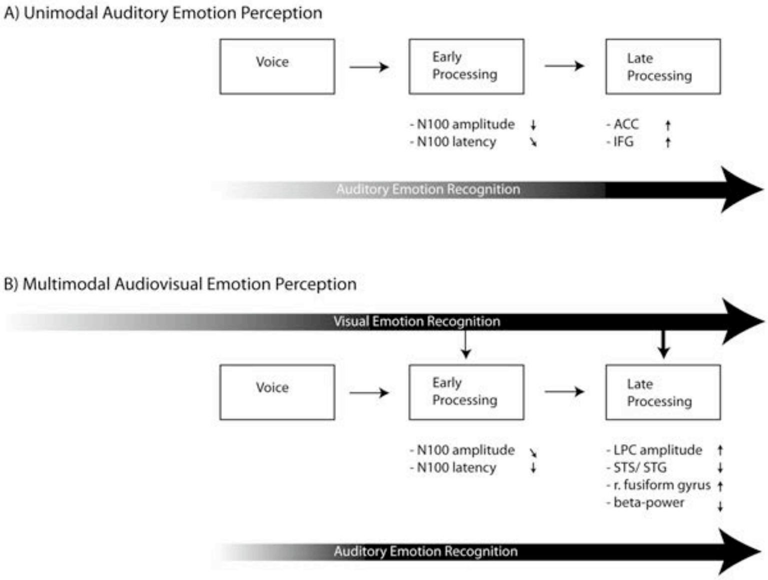

Figure 8.1: *Auditory vs. Audiovisual Emotion Perception.* The figure summarizes the results from all three studies presented, contrasting auditory emotion processing (A) with audiovisual emotion processing (B). Up-going arrows denote an increase in activation/ amplitude/latency, while arrows pointing downwards show a decrease. Tilted arrows denote an ambiguous effect; further details can be found in the text. Processing of voice information can be divided into an early and a late stage, temporally but also with respect to lower compared to higher processing hierarchy. While processing clearly differs at later stages, emotion effects early in processing are similar between auditory and audiovisual processing. The gradient of the arrows depicting visual and auditory emotion recognition signals the recognition process of the respective emotion. Note that auditory emotions are recognized faster in the multimodal condition. Furthermore, visual emotion recognition is already advanced when the auditory onset occurs.

However, in general, the response pattern was very similar across audio and audiovisual conditions, in both, EEG study 1 and 2:

- Emotional processing is characterized by a reduction in N100 amplitude.
- Emotional processing is characterized by a reduction in N100 latency.
- Audiovisual and auditory processing irrespective of emotional content lead to a robust reduction in N100 amplitude.

Hence, emotion processing as well as audiovisual integration are clearly reflected at this very early processing stage, yet the interplay between both factors seems to be rather small, if present at all. Thus, early auditory processing clearly is affected by accompanying visual information though not specifically by its emotional content.

Later processing steps in contrast show a strong interaction between emotional and multisensory processing.

- Larger LPC amplitudes for fearful stimuli are observed only in the audiovisual condition.
- Beta-power is reduced for emotional information in the audiovisual condition, but not in the audio condition (note, however, that beta-oscillations primarily indicate processing of biological motion rather than auditory processing, see below).
- While activation in the ACC and IFG increases for auditory emotional stimuli, activation for audiovisual emotion processing decreases in the STS/STG and increases in the right fusiform gyrus.

Emotional content therefore seems to influence multisensory processing rather at later, evaluative processing steps than at an early sensory level.

This interaction seems to crucially rely on the relative timing of visual and auditory information. In a natural setting, visual information typically precedes the onset of auditory information (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Vroomen & Stekelenburg, 2010), and previous studies suggest that this delay is crucial for efficient multisensory integration (Stekelenburg & Vroomen, 2007; Vroomen & Stekelenburg, 2010). Furthermore, in a congruent audiovisual setting, preceding visual information predicts not only the temporal onset of auditory information, but – at least in the present studies – also to a certain degree the content of the auditory signal: Depending on the body posture before the sound onset, the viewer builds up certain expectations for instance regarding the loudness of the sound. In the stimuli used in the current studies, the participants may even be able to predict the emotional content of the vocal expression to follow. Regarding the underlying mechanisms, it can therefore be assumed that the participant has already a certain representation regarding the emotional state of the observed person when the vocal expression starts. While this previous information does not affect early auditory processing, later stages seem to be strongly facilitated by the preceding visual information.

This assumption generates several hypotheses that will have to be tested in future experiments. First of all, the role of visual predictivity has to be confirmed. If the visual information before the sound onset differentially affects auditory processing depending on the emotional content, differences in visual processing depending on the emotion should also be visible. This should hold for the entire video, but, even more importantly, also for the initial segment before the sound onset. Furthermore, participants may even be able to determine above chance level the depicted emotion purely based on this initial subpart of the video.

Interesting insights could also arise by manipulating the visual predictability at various levels. Comparable to the studies by Stekelenburg and Vroomen (2007), shifting the sound onset in relation to the video should reduce the overall integration effect at the N100. However, if early auditory emotion processing is indeed not influenced by preceding visual information, the emotion effect at the N100 should remain largely unaffected. Later stages of multisensory processing should, however, be affected as a consequence of the earlier failure to properly integrate auditory and visual information.

Another interesting manipulation of visual predictability would be a change in congruency, that is, pairing visual stimuli expressing one emotion with auditory stimuli expressing a different one while retaining the temporal predictability. Multisensory interaction would probably be hampered irrespective of emotional content, albeit not as strongly as when temporal predictability is changed. In fact, previous studies manipulating content congruency yield inconsistent results, with some studies reporting differences in early processing (Molholm, Ritter, Javitt, & Foxe, 2004), while other only find differences at later stages, starting around 260 ms after sound onset (Yuval-Greenberg & Deouell, 2007); see also Doehrmann and Naumer (2008) for a review on that topic. Again, emotion processing should be unaffected at early stages, and only show differences at later stages.

The assumption that early emotion processing should not be influenced by incongruent visual information receives support from various previous studies. Primarily, these studies have investigated audiovisual interaction using facial expressions combined with matching or mismatching emotional words. Pourtois et al. (2002) report differences between congruent and incongruent stimuli at the P200b, but failed to find any difference at the N100. In contrast in an earlier study, Pourtois et al. (2000) in a previous study report larger auditory N100 amplitudes for congruent compared to incongruent face–voice pairings. However, all these studies used static photographs of facial expressions and included data from only seven respectively ten subjects. Thus, further studies are necessary to investigate the robustness of this effect, and whether dynamic, and hence more predictive, visual stimuli result in a similar pattern.

A second aspect that has to be investigated in future studies is the relation between temporal unfolding of processing, as shown in the EEG studies, and brain networks involved in such processing, as suggested by fMRI. Do brain areas posited for later emotion processing, such as ACC, IFG, but also STS/STG indeed activate *after* early sensory processing? What is the time course of the interaction between visual and auditory stimuli, as revealed by the conjunction? To address these questions, a combination of techniques offering both high temporal and spatial resolution would be ideal. On the one hand, this might be done by recording magnetoencephalography (MEG) (or also EEG, along with individual electrode locations) and computing a source analysis. On the other hand, a combined recording of EEG and fMRI could also shed light on these questions.

Therefore, the results obtained provide a solid starting point for the future investigation of multisensory emotion perception from dynamic stimuli.

8.2.2 What about Visual Processing?

The main part of the current thesis, as can be seen also in Figure 8.1, focuses on the influence of visual information on auditory processing. However, an influence may of course also be observed and analyzed in the opposite direction. That is, as an influence of auditory information on visual processing. While most principles described for the auditory perspective should also hold in the reverse case, visual processing differs from auditory processing in several respects.

First, as outlined above, visual information in a natural environment precedes auditory information. Hence, visual information can have a predictive influence on auditory processing, which auditory information cannot have on visual processing. This is also reflected in empirical studies manipulating the time-window for audiovisual integration. While an earlier onset of visual information compared to auditory information is tolerated for a rather large time-window, the time-window for a lag of visual onset is much smaller (van Wassenhove, Grant, & Poeppel, 2007; Maier, Luca, & Noppeney, 2011). Furthermore, while predictive visual information can provide a benefit in auditory speech processing, no benefit of auditory information on visual speech processing can be observed (Sánchez-García, Alsius, Enns, & Soto-Faraco, 2011). Hence, the predictive function that visual information fulfills in the scheme outlined in Figure 8.1 cannot be taken over by auditory information.

Second, at least in the stimulus material used in the present studies, vocal expressions are shorter and have a steeper and more clearly defined onset. In contrast, dynamic visual information unfolds over time, and while a clear onset of the visual signal or also of the motion in the video can clearly be determined, the emotional content can only be recognized over time. In principle, this is also the case in the auditory domain. However, it can be assumed that the temporal delay between auditory onset and auditory emotion recognition is much smaller. This view is supported by the finding that brain processes are affected by emotional content within 100 ms after sound onset, also in the unisensory auditory condition. Hence, it can be assumed that slowly accumulating visual information is processed at a different timescale than fast auditory information; this may have an impact on the integration of both modalities.

Additionally, from an empirical point of view, slow visual development renders the investigation of visual processing challenging. If no single point of emotion recognition can be determined, variations in brain responses signaling such as recognition are hard to detect. One way to partly circumvent this problem is the investigation of induced oscillatory brain responses, which was employed in both EEG studies. In order to observe a possible interplay between auditory and visual processing, the analyses were focused on the time range following the sound onset. In both studies only comparably small differences between visual and audiovisual processing were found, suggesting that additional sound information has only a limited influence on the ongoing processing.

This further supports the view that visual processing may operate on a different time scale, and a possible impact of auditory processing at such a rather late stage is rather small. Nevertheless, a larger suppression of oscillatory power for emotional stimuli was observed in the unisensory visual compared to the audiovisual condition in EEG-study 1. Hence, adding auditory information may indeed elicit facilitation in processing of emotional stimuli, indicated by a smaller suppression and thus a reduction in processing effort. This finding is compatible with the results of the fMRI study, showing stronger emotion effects in the unimodal than in the multimodal conditions.

However, to fully understand the impact of auditory information on visual processing, it will be necessary to monitor the unfolding of visual processing from the onset of the video on. Only then it will be possible to determine at what point in visual processing auditory information is added, and how this influences the unisensory processing. Another interesting approach would be the manipulation of visual information similar to the auditory manipulation in EEG-study 2. Several previous studies have used degraded visual information to investigate multisensory integration (Werner & Noppeney, 2010a, 2010b), suggesting that a comparable increase in interaction with decreasing signal quality can be observed. Thus, it remains subject to future studies how the specific role of emotional content would be affected by a degradation of visual information.

8.2.3 Interpretation of Brain Responses

Emotional content and multisensory perception both have the potential to facilitate processing, and a combination of both is assumed to result in even stronger facilitation. However, the question arises how such facilitation is characterized, and how it can be reflected in brain responses.

At a behavioral level, the concept of facilitation seems straightforward. A process is facilitated, if participants respond faster and/or with a higher accuracy (e.g. Burton et al., 2005; Briesemeister, Kuchinke, & Jacobs, 2011). A brain process causing such a speed–up or performance increase would be a brain signature of facilitation. However, causal relations between brain activations and behavior are hard to determine, especially on a higher cognitive level, without using lesion studies or temporary lesions induced by transcranial magnetic stimulation. The next-best evidence for the involvement of a brain response in a facilitation process would be correlation with behavioral performance. If for instance N100 amplitude decreases with emotion recognition accuracy, this indicates that the process underlying the N100 is indeed related to recognition accuracy. Such a correlation however requires a certain variability in the behavioral data, and as the behavioral tasks in the present studies were not designed to generate such a variability (i.e. all participants performed very well), no stable correlation with any brain responses could be observed.

Hence, the interpretation of a facilitated brain process mainly rests on previous studies investigating the respective brain response in the context of task performance.

N100 amplitude and latency reductions in response to multisensory stimuli have been reported in numerous previous studies (Besle et al., 2004; van Wassenhove et al., 2005; Stekelenburg & Vroomen, 2007; Vroomen & Stekelenburg, 2010), and both have been linked to benefits in behavioral performance (Besle et al., 2004). Stekelenburg and Vroomen (2007) state that "[t]he observed cortical deactivation [in the N100] to bimodal speech reflects facilitation of auditory processing as it is associated with behavioral facilitation, [...]" (p. 1964). Therefore, based on the literature, an interpretation of the N100 reduction as indicating facilitation for multisensory perception is justified.

In the context of emotion perception, the interpretation of a reduction in N100 amplitude and latency is more challenging, as emotion effects for congruent stimuli are typically observed later in auditory processing (see Discussion of EEG-study 1, p. 56, for a discussion of the N100 amplitude reduction for emotional stimuli). Previous studies, however, suggest a correlation between N100 amplitude and processing effort, showing that N100 amplitude increases with decreases in performance (Obleser & Kotz, 2011). Hence, more difficult, or less facilitated, processing seems to result in larger N100 amplitudes. Facilitated emotional stimuli may therefore elicit smaller N100 amplitudes compared to non-facilitated neutral ones. Clearly future studies are necessary to confirm such an early emotional facilitation, ideally by correlating neural responses and performance. Increased task difficulty would for example result in a larger behavioral variance, which should in turn lead to an inverse correlation between behavioral performance and N100 amplitude.

Comparable to the N100 amplitude, a robust decrease in beta-power was observed for emotional stimuli. Again, a direct link between task-performance and emotion processing in the beta-band is missing. An increase in beta-power has been correlated with a decrease in reaction time and an increase in multisensory integration (Senkowski et al., 2006), while a decreased beta-power corresponds to increasing task difficulty (Pfurtscheller & Silva, 1999). Hence, a beta-desynchronization, as observed in the present studies, most likely does not reflect facilitation. It may rather reflect a different aspect in emotion processing. Interestingly, beta-suppression was hardly modulated by adding auditory information, and is commonly found in biological motion processing (Muthukumaraswamy et al., 2006; Muthukumaraswamy & Singh, 2008). Therefore, the observed suppression may indicate rather an adaptive response to emotional information, possibly an action preparation (Pineda, 2005; Grabska-Barwinska & Zygierewicz, 2006).

Finally, an increase in BOLD signal was observed for unimodal emotional stimuli. In the audiovisual condition, a more complex pattern consisting of a deactivation for emotional stimuli in the STS/STG and an activation in the fusiform gyrus occurred. Increased BOLD responses have been linked to an improved performance (Kreifelts et al., 2007; Pichon et al., 2009) suggesting that indeed a facilitation occurs.

The link between decreased BOLD response and facilitated processing is less straightforward. Studies investigating processing at varying signal-to-noise ratios have reported an increase in BOLD response if information was presented suboptimally (Callan et al., 2003; Sekiyama et al., 2003; Werner & Noppeney, 2010b). Hence, a higher processing effort seems to correspond to an increased BOLD response, or, conversely, easier processing is associated with a reduced BOLD response. However, future investigation of this observation is necessary, in particular in combination with behavioral correlations. Do the observed deactivations correlate with an improvement in performance?

Interestingly it has been argued, that beta-power correlates inversely with BOLD signal (Ritter, Moosmann, & Villringer, 2009). Hence, a decrease in beta-power, as observed for emotional stimuli, may indicate the same processes as revealed by an increase in BOLD response for emotional body expressions. Nevertheless, this assumption as well needs further investigation, in particular a more sophisticated source localization of the beta-band suppression (i.e. using individual MR images as source models) to investigate a possible correspondence to activation increases observed in the fMRI.

8.3 Future Directions

8.3.1 Implications for Clinical Settings

Deficits in emotion processing are a key symptom of numerous psychiatric disorders. The two diseases used as examples in our review (Garrido-Vásquez, Jessen, & Kotz, 2011, p.105ff), major depression and schizophrenia, are of particular interest for the current discussion, as the observed deficits may indicate an impairment at different processing steps. Patients with schizophrenia are known to have deficits also in early sensory processing (Johnston et al., 2005). Hence, if initial sensory processing goes astray, this will likely impact on later processing steps. In contrast, patients suffering from major depression are assumed to show deficits in later, evaluative steps of emotion processing (Gotlib & Joormann, 2010), while early perception appears relatively spared.

As outlined in our review article (p.105ff), emotion processing in psychiatric patients has mainly been investigated using words or static photographs and explicit tasks in a single modality. The use of dynamic multisensory stimuli would potentially help to disentangle actual deficits in emotion processing from artifacts generated by the artificial test situation.

Beyond the aspects addressed in the review article, the results of the present studies also provide a link to other clinical settings, which will be outlined in the following.

Deficits in emotion processing are not only observed in psychiatric disorders, but also in developmental disorders such as autism spectrum disorder (ASD, Riby, Doherty-Sneddon, & Bruce, 2008; Bal et al., 2010; Kuchinke, Schneider, Kotz, & Jacobs, 2011). While a comprehensive discussion of potential multisensory emotion deficits in ASD patients goes beyond the scope of this dissertation, this patient group might form a particularly interesting case for several reasons.

It has long been known that ASD patients show a radically different gaze pattern from healthy controls when observing faces. While healthy controls for instance preferentially scan eyes when observing a person's face, no such pattern can be observed in people suffering from ASD (Spezio, Adolphs, Hurley, & Piven, 2007). The eyes, however, are a crucial source of emotional information (Buchan, Paré, & Munhall, 2007). If participants ignore these cues, it is not surprising that they cannot determine a person's emotional state. However, this also raises the question whether people suffering from ASD are equally impaired when emotional information is not only present in the facial expression, but also in the body movements or the voice, and if emotional information is presented via multiple modalities simultaneously. Alternatively, patients suffering from autism might even perform worse if emotional information is presented in too many modalities at the same time, as ASD patients tend to focus on local rather than global features (Happé, 1999; Wang, Mottron, Peng, Berthiaume, & Dawson, 2007).

With respect to multisensory integration ASD patients, evidence is inconsistent: while some authors report deficits in multisensory integration (Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2011), others report no impairment (David, Schneider, Vogeley, & Engel, 2011). Furthermore, also for the single modalities, results from previous studies are heterogeneous. Some studies report a deficit in biological motion perception (Blake, Turner, Smoski, Pozdol, &

Stone, 2003), whereas others did not find any impairment (Murphy, Brady, Fitzgerald, & Troje, 2009). Also emotion perception from voices seems to be impaired, though not consistently (Chevallier, Noveck, Happé, & Wilson, 2011).

Hence, the investigation of emotion perception from complex dynamic stimuli is of high clinical interest for various neuropsychiatric as well as developmental disorders. Applying the stimuli used in the present study in such a setting may thus help to differentiate which aspects of emotion processing are affected in which disorder.

8.3.2 Multisensory Integration and Neural Oscillations

Brain oscillations were investigated in the present study mainly with the aim of assessing changes in ongoing visual processing. However, they also offer a new approach to the mechanisms underlying multisensory integration, which may open a new perspective for future studies.

Similar to the unisensory problem of feature binding in object perception (W. Singer & Gray, 1995), multisensory integration can be considered as a problem of information binding between modalities. Different aspects of incoming sensory information are processed in different specialized brain regions: visual information in occipital and inferior temporal areas, auditory information in temporal areas, to stay on a very broad level. Yet, under normal conditions we have no problem determining that visual and auditory information originate from the same source, and typically perceive that information as a single event despite the different time scales at which the unisensory information arrives. How is such a binding realized in the brain?

Regarding unisensory feature binding, Engel, Fries, and Singer (2001) posit that one possible mechanisms may be the synchronization of oscillations originating from different neuronal populations. Cells responding to the same object should fire in synchrony, which creates a temporal coupling between cells and sets them apart from other neurons not involved in processing that particular object. In particular, gamma oscillations have turned out to be a prime candidate realizing this principle (Tallon-Baudry & Bertrand, 1999; Fries, Reynolds, Rorie, & Desimone, 2001).

A similar principle has recently been suggested to underlie the binding of information from different modalities (Senkowski et al., 2008). If neurons from different early sensory areas respond to the same object, they synchronize, which in turn leads to an increased activation of multisensory areas receiving inputs from these neurons. Again, gamma oscillations seem to play a key role in such a mechanism (Senkowski, Schneider, Tandler, & Engel, 2009; T. Schneider, Lorenz, Senkowski, & Engel, 2011).

Furthermore, a complementary approach emphasizes the role of phase-coupling between high and low frequencies in multisensory facilitation (Lakatos et al., 2007; Schroeder et al., 2008). The perception of information from one modality can apparently reset the phase of low frequency oscillations, so that information from the second modality arrives at an optimal time-point within an oscillation cycle. Schroeder et al. (2008) for instance propose that visual facial information can influence the phase of the theta-oscillations in the primary auditory cortex so that newly incoming auditory information can be processed optimally. Interestingly, such a mechanisms can also explain the biological basis of several key features of multisensory integration outlined above.

First, it is essential that one modality, in the case of audiovisual integration the visual one, precedes the other. A phase reset requires a certain, albeit small, amount of time. If information from both modalities starts at the identical point in time, no prior phase reset is possible, and auditory information may or may not arrive at an optimal point within the oscillation cycle. Second, and related to that, the auditory onset has to be predictable by the visual information. Probably due to lifelong experience, our brain knows about the natural time-lag between visual and auditory onset, and this information is taken as the basis for a phase reset. If it is violated, auditory information does not arrive at an optimal time-point and can not be integrated optimally. In fact, this may even create an audiovisual suppression below the unimodal case, as has been shown in electrophysiological studies in nonhuman animals (Ghazanfar et al., 2005; Lakatos et al., 2007). If no visual prediction can elicit a phase-reset, oscillations in the auditory cortex will sometimes be in an optimal phase and sometime not, purely by chance. If, however, visual information induces an incorrect phase reset, auditory information will definitely arrive at a suboptimal point in time, causing a suppression in activation below the normal average.

Therefore, the closer analysis of brain oscillations offers an extremely promising approach to understanding the neural basis of multisensory integration. However, the research in this field has only emerged in the past 15 years, and many open questions remain. In particular, most knowledge originates from electrophysiological recordings in nonhuman animals (e.g.

Ghazanfar et al., 2005; Lakatos et al., 2007). While this allows for the investigation of response properties of very small groups of neurons, findings are limited to the cognitive and behavioral realm of a monkey. Therefore, problems arise particularly when it comes to more difficult task settings, or higher cognitive abilities such as language processing. Studies in humans, on the other hand, have often employed very simple stimulus material (Senkowski, Talsma, Grigutsch, Herrmann, & Woldorff, 2007; Besle et al., 2011), which allows for a large degree of control over physical stimulus properties, but also limits the ecological validity (see below for a more extensive discussion of this issue). However, several recent studies using more complex stimulus material in humans have also demonstrated similar effects (Senkowski et al., 2009; J. Keil, Müller, Ihssen, & Weisz, 2011), such as an influence of oscillatory brain activity prior to sound onset on the perception of ambiguous auditory information (J. Keil et al., 2011).

An interesting future direction to the studies presented here would therefore be the more extensive analysis of oscillatory brain responses. Does the phase of low frequency oscillations change during the course of a video, thereby facilitating auditory processing? How large is the tolerable time-lag between auditory and visual information, and how does oscillatory activity change depending on the time-lag? Can a coupling between different frequency bands (i.e. between their phase and amplitude) be observed? How does emotional content influence these measures? A closer investigation of these questions could yield important insights into the mechanisms underlying complex multisensory perception and communication.

8.3.3 Simple and Complex Stimuli

The main emphasis of the studies presented here lies on the investigation of multisensory information from ecologically valid or complex stimuli. It is assumed that such a stimulus set allows for the investigation of brain processes that approximate “real life” brain mechanisms as close as possible. This approach can be extended in two directions. On the one hand, it may be interesting to simplify the stimulus material used, thereby decreasing ecological validity but increasing control over physical stimulus properties (Sabatinelli et al., 2011). On the other hand, the ecological validity may be increased, thereby avoiding abstractions still present in the stimulus material used.

Regarding the first point, the limited control of physical stimulus properties is an inevitable caveat of increasing ecological validity. The stimulus material used in the present studies obeyed a certain number of control criteria: all videos were recorded in the same, neutral setting; all actors wore similar clothing; every video started with a still frame of the actor; vocal expressions were limited to “ah”,“oh”, and “mh” and presented at the same loudness to all participants; all stimuli were rated and yielded comparable recognition rates. Nevertheless, a number of physical aspects remained uncontrolled. While the amount of motion, the pitch, and the spectral center of gravity were computed and did not consistently set one condition apart from the others, some differences were observed. In principle, it would be possible to control these features, and thereby reduce a possible confound. However, these features are characteristic for and a defining feature of different emotions. Pitch variations are for instance larger for emotional compared to neutral voice (Breitenstein, Lancker, Daum, & Waters, 2001), and levelling all differences between different emotional and neutral stimuli would render them indistinguishable. Yet, a controlled manipulation of physical features in future studies may help to shed light on the individual contributions of these features. Increasing the loudness of auditory stimuli may for instance increase effects for emotional but not for neutral stimuli.

Conversely, while the stimulus material was designed to closely approximate an ecologically valid situation, it still differed from a real-world setting in numerous aspects. Emotional expressions were presented by actors; yet, studies have shown that posed emotions are processed differently from “real” emotions (Allen & Atkinson, 1981; McGettigan, Agnew, Walsh, Jessop, & Scott, 2011). Likewise, emotional expressions were presented out of context, which also does not occur in everyday life, but allowed for the disentanglement of context and body expressions. A different approach for future studies would therefore be to increase the complexity of the stimuli used. For example by comparing responses to stimuli with or without context, the role of context in the perception of emotional stimuli may further be explored.

8.3.4 Different Emotions

Anger and fear were used as two examples of emotional states in the present studies. They are particularly suited, as they typically elicit strong responses, are easily recognized, and can naturally be expressed by the interjections used in the auditory condition. Yet, an interesting question for future studies would be the comparison to various other emotional states. One approach may be to compare anger and fear to other basic emotions, such as happiness (to contrast positive and negative states) or sadness (to vary the arousal level). Another interesting take however is the comparison to variations of anger and fear. In the stimulus material used in the present studies, the actors depicted a high level of emotionality. Both emotions can however also be expressed to varying degrees, and in everyday situations people rarely express their emotions to the full degree. Hence, a stepwise increase in emotionality in a parametric approach may yield interesting results regarding more subtle influences of emotionality. Furthermore, the terms “anger” and “fear” subsume a host of different emotional states. Someone afraid of a dog will act very different from someone afraid of giving a public talk. Likewise, anger can be expressed in overt physical threat but also in a rather subtle tension (commonly referred to as “hot” and “cold” anger). In the present studies such differences were excluded as much as possible by providing the actors with specific situations they should imagine to express the emotions. While this reduced the variability in the stimulus material, it also focuses the experiments on a very small and specific segment of emotional states.

Especially in the context of social communication, it may be promising to investigate affective states beyond the classical basic emotions. Pride, shame, guilt, and other “social emotions” are particularly relevant in the maintenance and establishment of social relations and group hierarchies. They are not as easily associated with distinct facial expressions as basic emotions (Adolphs, 2002b), and hence the communication via multiple modalities as well as dynamic changes may be of increased importance (App, McIntosh, Reed, & Hertenstein, 2011).

8.3.5 Action Tendencies

Besides the features listed in the previous section, anger and fear are particularly interesting in the present study for another reason. Both emotions have a strong link to behavioral responses, and for both, the directedness of the emotion can be highly important. Recognizing that someone is angry can have a totally different impact on the observer dependent on whether that someone is angry at the observer or at someone else. If someone is angry at me, I might have to defend myself, while if someone is angry at a different person, that does not necessarily affect me. Likewise in the case of fear: If someone is afraid of me, this might be good to know but does not have to be of direct behavioral relevance. If, however, someone is afraid of something next to me, that “something” may also be a threat to me and require a fast response. Hence, anger directed at the observer and fear directed away from the observer may be of increased behavioral relevance.

This is also reflected in results of previous studies on the interplay between eye gaze direction and emotion expression. While facial expressions of anger are recognized faster if the eyes are directed at the observer, fear is recognized faster if the gaze is averted (Adams et al., 2003; Adams & Kleck, 2003, 2005).

As body expressions provide a yet more direct link to someone's intended behavior or "action tendencies" (de Gelder, 2009), it seems likely that the just cited effects may be even enhanced for directed body expressions compared to directed gaze. An interesting way to investigate the connection between emotion and behavioral relevance would therefore be to investigate the perception of body expressions displaying anger or fear either directed at the observer or elsewhere.

Furthermore, manipulating spatial information provides a different and interesting approach to investigate multisensory interactions. Unlike visual information, which can contain a clear spatial direction even if presented at the same spot on a computer screen, auditory information presented binaurally over headphones provides only very limited spatial information. A first interesting question therefore is whether the effect described above, namely an improved processing of directed anger and averted fear information, is modulated by accompanying auditory information. On the one hand, the auditory information does not contain any new spatial information, and may therefore improve overall recognition performance, but not impact on the directionality effect. On the other hand, the improved recognition performance of emotional content may also allow for a faster processing of other types of information, for instance spatial direction, and thereby enhance the directionality effect. However, a third option may also be a decreased directionality effect, as a mismatch between the spatially directed visual information and the spatially undirected auditory information may hamper processing.

Finally, it is of course also possible to manipulate the spatial information contained in the auditory signal. This can either be done by simulating spatial differences via a head-related transfer function (e.g. J. Chan, Freed, Vermiglio, & Soli, 2008; Zahorik, 2009) or by presenting auditory information at different locations via separate loudspeakers (freefield, e.g. Tian, Reser, Durham, Kustov, & Rauschecker, 2001; Lewald & Getzmann, 2011). In any case, the interplay between emotional content, spatial directedness, and multimodality offers an interesting direction for the further investigation of emotional communication in complex environments.

8.4 Conclusion

The aim of the current thesis was to investigate the interplay between face, body, and voice expressions in the communication of emotional information. In two EEG studies and one fMRI study flanked by behavioral pilot studies, an early integration between auditory and visual information was observed, as well as early differences in the processing of nonemotional compared to emotional information. Multimodal perception and emotional content furthermore interact in processing, which leads to an increased facilitation.

The studies presented here are the first to investigate the interaction between all three channels – face, body, and voice – in a systematic way. By combining dynamic multisensory information and emotional content, it was possible to compare the interaction between these two factors. Furthermore, the combination of more traditional (ERP) as well as recent (time–frequency analysis) EEG analysis techniques and fMRI allowed for an assessment of interactions on temporal as well as on a spatial level. The results thus provide a valuable starting point for the investigation of emotional communication under complex multisensory conditions.

References

Adams, R., Gordon, H., Baird, A., Ambady, N., & Kleck, R. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science*, *300*, 1536.

Adams, R., & Kleck, R. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychol Sci*, *14*(6), 644–647.

Adams, R., & Kleck, R. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion*, *5(1)*, 3–11.

Adolphs, R. (2002a). Neural systems for recognizing emotion. *Curr Opin Neurobiol*, *12*(2), 169–177.

Adolphs, R. (2002b). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behav Cogn Neurosci Rev*, *1*(1), 21–62.

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J Neurosci*, *20*(7), 2683–2690.

Adolphs, R., Schul, R., & Tranel, D. (1998). Intact recognition of facial emotion in Parkinson's disease. *Neuropsychology*, *12*(2), 253–258.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., et al. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, *37*(10), 1111–1117.

Alegre, M., Gurtubay, I. G., Labarga, A., Iriarte, J., Valencia, M., & Artieda, J. (2004). Frontal and central oscillatory changes related to different aspects of the motor process: A study in go/no-go paradigms. *Exp Brain Res*, *159*(1), 14–22.

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nat Neurosci*, *14*(10), 1338–1344.

Alfimova, M. V., & Uvarova, L. G. (2008). Changes in EEG spectral power on perception of neutral and emotional words in patients with schizophrenia, their relatives, and healthy subjects from the general population. *Neurosci Behav Physiol*, *38*(5), 533–540.

Allen, V., & Atkinson, M. (1981). Identification of spontaneous and deliberate behavior. *Journal of Nonverbal Behavior*, *5(4)*, 224–237.

Allison, Puce, & McCarthy. (2000). Social perception from visual cues: Role of the STS region. *Trends Cogn Sci*, *4*(7), 267–278.

Almeida, J. R. C., Versace, A., Hassel, S., Kupfer, D., & Phillips, M. (2010). Elevated amygdala activity to sad facial expressions: A state marker of bipolar but not unipolar depression. *Biological Psychiatry*, *67*(5), 414–421.

Alter, K., Rank, E., Kotz, S., Toepel, U., Besson, M., Schirmer, A., et al. (2003). Affective encoding in the speech signal and in event-related brain potentials. *Speech Communication*, *40*, 61–70.

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face. *Psychological Science*, *16*(5), 403–410.

Amedi, A., von Kriegstein, K., van Atteveldt, N., Beauchamp, M., & Naumer, M. (2005). Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res*, *166*, 559–571.

An, S. K., Lee, S. J., Lee, C. H., Cho, H. S., Lee, P. G., Lee, C., et al. (2003). Reduced P3 amplitudes by negative facial emotional photographs in schizophrenia. *Schizophrenia Research*, *64*(2-3), 125–135.

App, B., McIntosh, D. N., Reed, C. L., & Hertenstein, M. J. (2011). Nonverbal channel use in communication of emotion: How may depend on why. *Emotion*, *11*(3), 603–617.

Archer, J., Hay, D. C., & Young, A. W. (1994). Movement, face processing and schizophrenia: Evidence of a differential deficit in expression analysis. *Br J Clin Psychol*, *33 (Pt 4)*, 517–528.

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A.-L. (2009). Dual neural routing of visual facilitation in speech processing. *J Neurosci*, *29*(43), 13445–13453.

Ashley, V., Vuilleumier, P., & Swick, D. (2004). Time course and specificity of event-related potentials to emotional expressions. *Neuroreport*, *15*(1), 211–216.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*(6), 717–746.

Atkinson, A. P., Vuong, Q. C., & Smithson, H. E. (2011). Modulation of the face- and body-selective visual regions by the motion and emotion of point-light face and body stimuli. *Neuroimage*.

Babiloni, C., Babiloni, F., Carducci, F., Cincotti, F., Cocozza, G., Del Percio, C., et al. (2002). Human cortical electroencephalography (EEG) rhythms during the observation of simple aimless movements: a high-resolution EEG study. *Neuroimage*, *17*(2), 559–572.

Bach, D. R., Herdener, M., Grandjean, D., Sander, D., Seifritz, E., & Strik, W. K. (2009). Altered lateralisation of emotional prosody processing in schizophrenia. *Schizophrenia Research*, *110*(1-3), 180–187.

Bal, E., Harden, E., Lamb, D., Hecke, A. V. V., Denver, J. W., & Porges, S. W. (2010). Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *J Autism Dev Disord*, *40*(3), 358–370.

Banks, S. J., Eddy, K. T., Angstadt, M., Nathan, P. J., & Phan, K. L. (2007). Amygdala–frontal connectivity during emotion regulation. *Social Cognitive and Affective Neuroscience*, *2*(4), 303–312.

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Brain Res Cogn Brain Res*, *17*(3), 613–620.

Baumann, O., & Mattingley, J. B. (2010). Scaling of neural responses to visual and auditory motion in the human cerebellum. *J Neurosci*, *30*(12), 4489–4495.

Beauchamp, M. S. (2005a). See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex. *Curr Opin Neurobiol*, *15*(2), 145–153.

Beauchamp, M. S. (2005b). Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics*, *3*(2), 93–113.

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), 809–823.

Beaucousin, V., Lacheret, A., Turbelin, M.-R., Morel, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). FMRI study of emotional speech comprehension. *Cereb Cortex*, *17*(2), 339–352.

Beck, A. (1967). *Depression: Clinical, experimental, and theoretical aspects* (A. Beck, Ed.). New York, NY: Heber Medical Division, Harper & Row.

Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proc Biol Sci*, *275*(1634), 473–481.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behav Res Methods*, *40*(2), 531–539.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *J Cogn Neurosci*, *8*(6), 551–565.

Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (p. 141-177). Oxford University Press.

Besle, J., Fort, A., Delpuech, C., & Giard, M. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *Eur J Neurosci*, *20*(8), 2225–2234.

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *J Neurosci*, *31*(9), 3176–3185.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annu Rev Psychol*, *58*, 47–73.

Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., & Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychol Sci*, *14*(2), 151–157.

Bobes, M. A., Martin, M., Olivares, E., & Valdés-Sosa, M. (2000). Different scalp topography of brain potentials related to expression and identity matching of faces. *Cognitive Brain Research*, *9*(3), 249–260.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5 (9/10)*, 341–345.

Bolognini, N., Leo, F., Passamonti, C., Stein, B., & Làdavas, E. (2007). Multisensory-mediated auditory localization. *Perception*, *36*, 1477–1485.

Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event–related brain potentials to emotional exclamations. *Psychophysiology*, *41*(2), 259–268.

Bouhuys, A. L., Bloem, G. M., & Groothuis, T. G. G. (1995). Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects. *Journal of Affective Disorders*, *33*(4), 215–226.

Bould, E., & Morris, N. (2008). Role of motion signals in recognizing subtle facial expressions of emotion. *Br J Psychol*, *99*(Pt 2), 167–189.

Bradley, M., & Lang, P. (2007). Emotion and motivation. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (p. 581-607). Cambridge: Cambridge University Press.

Bramon, E., Rabe-Hesketh, S., Sham, P., Murray, R. M., & Frangou, S. (2004). Meta-analysis of the P300 and P50 waveforms in schizophrenia. *Schizophrenia Research*, *70*(2-3), 315–329.

Breitenstein, C., Lancker, D. V., Daum, I., & Waters, C. H. (2001). Impaired Perception of Vocal Emotions in Parkinson's Disease: Influence of Speech Time Processing and Executive Functioning. *Brain and Cognition*, *45*(2), 277–314.

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2011). Discrete emotion effects on lexical decision response times. *PLoS One*, *6*(8), e23743.

Brosch, T., Grandjean, D., Sander, D., & Scherer, K. (2009). Cross-modal emotional attention: Emotional voices modulate early stages of visual processing. *Journal of Cognitive Neuroscience*, *21(9)*, 1670–1679.

Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Soc Neurosci*, *2*(1), 1–13.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., et al. (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Brain Res Cogn Brain Res*, *9*(3), 227–238.

Burton, L. A., Rabin, L., Wyatt, G., Frohlich, J., Bernstein Vardy, S., & Dimitri, D. (2005). Priming effects for affective vs. neutral faces. *Brain Cogn*, *59*(3), 322–329.

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

Buzsáki, G. (2006). *Rhythms of the brain.* Oxford University Press.

Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures.

Neuroreport, *14*(17), 2213–2218.

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport*, *10*(12), 2619–2623.

Calvert, G. A., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal identification. *Trends Cogn Sci*, *2*(7), 247–253.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593–596.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol*, *10*(11), 649–657.

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

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends Cogn Sci*, *11*(12), 535–543.

Campanella, S., Bruyer, R., Froidbise, S., Rossignol, M., Joassin, F., Kornreich, C., et al. (2010). Is two better than one? A cross-modal oddball paradigm reveals greater sensitivity of the P300 to emotional face-voice associations. *Clinical Neurophysiology*, *121*(11), 1855–1862.

Cappe, C., Morel, A., Barone, P., & Rouiller, E. M. (2009). The thalamocortical projection systems in primate: An anatomical support for multisensory and sensorimotor interplay. *Cereb Cortex*, *19*(9), 2025–2037.

Cappe, C., Rouiller, E. M., & Barone, P. (2009). Multisensory anatomical pathways. *Hear Res*, *258*(1-2), 28–36.

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia*, *47*(4), 1045–1052.

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: Timing, topography, directionality, and sources. *J Neurosci*, *30*(38), 12572–12580.

Carretié, L., Hinojosa, J. A., Lépez-Martín, S., Albert, J., Tapia, M., & Pozo, M. A. (2009). Danger is worse when it moves: Neural and behavioral indices of enhanced attentional capture by dynamic threatening stimuli. *Neuropsychologia*, *47*(2), 364–369.

Cavanagh, J., & Geisler, M. W. (2006). Mood effects on the ERP processing of emotional intensity in faces: A P3 investigation with depressed students. *International Journal of Psychophysiology*, *60*(1), 27–33.

Chan, J., Freed, D. J., Vermiglio, A. J., & Soli, S. D. (2008). Evaluation of binaural functions in bilateral cochlear implant users. *Int J Audiol*, *47*(6), 296–310.

Chan, R., Li, H., Cheung, E. F. C., & Gong, Q.-Y. (2010). Impaired facial emotion perception in schizophrenia: A meta-analysis. *Psychiatry Research*, *178*(2), 381–390.

Chang, Y., Xu, J., Shi, N., Zhang, B., & Zhao, L. (2010). Dysfunction of processing task-irrelevant emotional faces in major depressive disorder patients revealed by expression-related visual MMN. *Neuroscience Letters*, *472*(1), 33–37.

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2011). What's in a voice? Prosody as a test case for the theory of mind account of autism. *Neuropsychologia*, *49*(3), 507–517.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *J Abnorm Psychol*, *100*(3), 316–336.

Cochin, S., Barthelemy, C., Lejeune, B., Roux, S., & Martineau, J. (1998). Perception of motion and qEEG activity in human adults. *Electroencephalogr Clin Neurophysiol*, *107*(4), 287–295.

Costafreda, S. G., Brammer, M. J., David, A. S., & Fu, C. H. Y. (2008). Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies. *Brain Res Rev*, *58*(1), 57–70.

Critchley, H., Daly, E., Phillips, M., Brammer, M., Bullmore, E., Williams, S., et al. (2000). Explicit and implicit neural mechanisms for processing of social information from facial expressions: A functional magnetic resonance imaging study. *Human Brain Mapping*, *9*(2), 93–105.

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clin Neurophysiol*, *115*(4), 732–744.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective pictures processing: Covariation with autonomic arousal and affective report. *Biological Psychology*, *52*, 95-111.

Dai, Q., Feng, Z., & Koster, E. H. W. (2011). Deficient distracter inhibition and enhanced facilitation for emotional stimuli in depression: An ERP study. *Int J Psychophysiol*, *79*(2), 249–258.

Dannlowski, U., Ohrmann, P., Bauer, J., Deckert, J., Hohoff, C., Kugel, H., et al. (2008). 5-HTTLPR biases amygdala activity in response to masked facial expressions in major depression. *Neuropsychopharmacology*, *33*(2), 418–424.

Dannlowski, U., Ohrmann, P., Bauer, J., Kugel, H., Arolt, V., Heindel, W., et al. (2007). Amygdala reactivity to masked negative faces is associated with automatic judgmental bias in major depression: A 3 T fMRI study. *Journal of Psychiatry & Neuroscience*, *32*(6), 423–429.

Dannlowski, U., Ohrmann, P., Konrad, C., Domschke, K., Bauer, J., Kugel, H., et al. (2009). Reduced amygdala–prefrontal coupling in major depression: Association with MAOA genotype and illness severity. *International Journal of Neuropsychopharmacology*, *12*(1), 11–22.

Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray, UK.

Das, P., Kemp, A. H., Flynn, G., Harris, A. W. F., Liddell, B. J., Whitford, T. J., et al. (2007). Functional disconnections in the direct and indirect amygdala pathways for fear processing in schizophrenia. *Schizophrenia Research*, *90*(1-3), 284–294.

David, N., Schneider, T. R., Vogeley, K., & Engel, A. K. (2011). Impairments in multisensory processing are not universal to the autism spectrum: No evidence for crossmodal priming deficits in Asperger syndrome. *Autism Res*, *4*(5), 383–388.

Davis, K. D., Taylor, K. S., Hutchison, W. D., Dostrovsky, J. O., McAndrews, M. P., Richter, E. O., et al. (2005). Human anterior cingulate cortex neurons encode cognitive and emotional demands. *J Neurosci*, *25*(37), 8402–8406.

de Gelder, B. (2006a). Toward a biological theory of emotional body language. *Biological Theory*, *1(2)*, 130–133.

de Gelder, B. (2006b). Towards the neurobiology of emotional body language. *Nat Rev Neurosci*, *7*(3), 242–249.

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philos Trans R Soc Lond B Biol Sci*, *364*(1535), 3475–3484.

de Gelder, B., Böcker, K., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: Early interaction revealed by human electric brain responses. *Neurosci Lett*, *260*, 133–136.

de Gelder, B., Snyder, J., Greve, D., Gerard, G., & Hadjikhani, N. (2004). Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. *PNAS*, *101(47)*, 16701–16706.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, *14(3)*, 289-311.

de Gelder, B., Vroomen, J., Annen, L., Masthof, E., & Hodiamont, P. (2003). Audio-visual integration in schizophrenia. *Schizophrenia Research*, *59*(2-3), 211–218.

de Gelder, B., Vroomen, J., de Jong, S. J., Masthoff, E. D., Trompenaars, F. J., & Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics. *Schizophrenia Research*, *72*(2-3), 195–203.

de Jong, J. J., Hodiamont, P. P. G., & de Gelder, B. (2010). Modality-specific attention and multisensory integration of emotions in schizophrenia: Reduced regulatory effects. *Schizophrenia Research*, *122*(1-3), 136–143.

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

Decety, J., & Grézes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends Cogn Sci*, *3*(5), 172–178.

Dekio-Hotta, S., Kojima, T., Karino, S., Yamasoba, T., Dekio, I., Ito, Y. M., et al. (2009). N1 component reflects difference of terminal chords in three-chord sequences. *Neuroreport*, *20*(3), 251–256.

Deldin, P. J., Keller, J., Gergen, J. A., & Miller, G. A. (2000). Right-posterior face processing anomaly in depression. *Journal of Abnormal Psychology*, *109*(1), 116–121.

Deldin, P. J., Keller, J., Gergen, J. A., & Miller, G. A. (2001). Cognitive bias and emotion in neuropsychological models of depression. *Cognition & Emotion*, *15*(6), 787–802.

Demily, C., Attala, N., Fouldrin, G., Czernecki, V., Ménard, J.-F., Lamy, S., et al. (2010). The Emotional Stroop task: A comparison between schizophrenic subjects and controls. *Eur Psychiatry*, *25*(2), 75–79.

Deveney, C. M., & Deldin, P. J. (2004). Memory of faces: A slow wave ERP study of major depression. *Emotion*, *4*(3), 295–304.

Dietrich, S., Hertrich, I., Alter, K., Ischebeck, A., & Ackermann, H. (2008). Understanding the emotional expression of verbal interjections: A functional MRI study. *Neuroreport*, *19*(18), 1751–1755.

Dimitrijevic, A., Michalewski, H. J., Zeng, F.-G., Pratt, H., & Starr, A. (2008). Frequency changes in a continuous tone: Auditory cortical potentials. *Clin Neurophysiol*, *119*(9), 2111–2124.

Doallo, S., Cadaveira, F., & Holguín, S. R. (2007). Time course of attentional modulations on automatic emotional processing. *Neurosci Lett*, *418*(1), 111–116.

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

Dolan, R. J., Morris, J., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proc Natl Acad Sci USA*, *98*(17), 10006–10010.

Douglas, K. M., & Porter, R. J. (2010). Recognition of disgusted facial expressions in severe depression. *The British Journal of Psychiatry*, *197*(2), 156–157.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473.

Downing, P. E., Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastriate body area in action perception. *Soc Neurosci*, *1*(1), 52–62.

Edwards, J., Pattison, P. E., Jackson, H. J., & Wales, R. J. (2001). Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophrenia Research*, *48*(2-3), 235–253.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, *45*(1), 15–31.

Ekman, P., Sorenson, E., & Friesen, W. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*, 86–88.

Elliott, R., Baker, S. C., Rogers, R. D., O'Leary, D. A., Paykel, E. S., Frith, C. D., et al. (1997). Prefrontal dysfunction in depressed patients performing a complex planning task: A study using positron emission tomography. *Psychological Medicine*, *27*(4), 931–942.

Emerson, C. S., Harrison, D. W., & Everhart, D. E. (1999). Investigation of receptive affective prosodic ability in school-aged boys with and without depression. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *12*(2), 102–109.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations – signalling the status quo? *Curr Opin Neurobiol*, *20*(2), 156–165.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nat Rev Neurosci*, *2*(10), 704–716.

Ethofer, T., Anders, S., Erb, M., Droll, C., Royen, L., Saur, R., et al. (2006). Impact of voice on emotional judgment of faces: An event-related fMRI study. *Hum Brain Mapp*, *27*(9), 707–714.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., et al. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *Neuroimage*, *30*(2), 580–587.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., et al. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *J Cogn Neurosci*, *21*(7), 1255–1268.

Ethofer, T., Pourtois, G., & Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Prog Brain Res*, *156*, 345–361.

Ethofer, T., Ville, D. V. D., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Curr Biol*, *19*(12), 1028–1033.

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *TICS*, *15(2)*, 85–93.

Evans, A. C., Marrett, S., Neelin, P., Collins, L., Worsley, K., Dai, W., et al. (1992). Anatomical mapping of functional activation in stereotactic coordinate space. *Neuroimage*, *1*(1), 43–53.

Fawcett, I. P., Hillebrand, A., & Singh, K. D. (2007). The temporal sequence of evoked and induced cortical responses to implied-motion processing in human motion area V5/MT+. *Eur J Neurosci*, *26*(3), 775–783.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2005). Sensitivity to voice in human prefrontal cortex. *J Neurophysiol*, *94*(3), 2251–2254.

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage*, *36*(2), 480–487.

Foti, D., Olvet, D. M., Klein, D. N., & Hajcak, G. (2010). Recduced electrocortical response to threatening faces in major depressive disorder. *Depression and Anxiety*, *27*, 813-820.

Fox, E., Lester, V., Russo, R., Bowles, R., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cogn Emot.*, *14(1)*, 61–92.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res*, *147*(3), 332–343.

Freedman, R., Adler, L. E., Gerhardt, G. A., Walkdo, M., Baker, N., Rose, G. M., et al. (1987). Neurobiological studies of sensory gating in schizophrenia. *Schizophrenia Bulletin*, *13*(4), 669–679.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, *291*(5508), 1560–1563.

Friston, K., Ashburner, J., Frith, C., Poline, J., Heather, J., & Frackowiak, R. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, *2(4)*, 189–210.

Fritzsche, A., Dahme, B., Gotlib, I. H., Joormann, J., Magnussen, H., Watz, H., et al. (2010). Specificity of cognitive biases in patients with current depression and remitted depression and in patients with asthma. *Psychological Medicine*, *40*(5), 815–826.

Frodl, T., Scheuerecker, J., Albrecht, J., Kleemann, A., Müller-Schunk, S., Koutsouleris, N., et al. (2009). Neuronal correlates of emotional processing in patients with major depression. *World Journal of Biological Psychiatry*, *10*(3), 202–208.

Fu, C. H. Y., Williams, S. C. R., Brammer, M. J., Suckling, J., Kim, J., Cleare, A. J., et al. (2007). Neural responses to happy facial expressions in major depression following antidepressant treatment. *American Journal of Psychiatry*, *164*(4), 599–607.

Fu, C. H. Y., Williams, S. C. R., Cleare, A. J., Brammer, M. J., Walsh, N. D., Kim, J., et al. (2004). Attenuation of the neural response to sad faces in major depression by antidepressant treatment: A prospective, event-related functional magnetic resonance imaging study. *Archives of General Psychiatry*, *61*(9), 877–889.

Fu, C. H. Y., Williams, S. C. R., Cleare, A. J., Scott, J., Mitterschiffthaler, M. T., Walsh, N. D., et al. (2008). Neural responses to sad facial expressions in major depression following cognitive behavioral therapy. *Biological Psychiatry*, *64*(6), 505–512.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., et al. (2009). Functional atlas of emotional faces processing: A voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *J Psychiatry Neurosci*, *34*(6), 418–432.

Garrido-Vásquez, P., Jessen, S., & Kotz, S. A. (2011). Perception of emotion in psychiatric disorders: On the possible role of task, dynamics, and multimodality. *Soc Neurosci*, *6*(5-6), 515–536.

Gastaut, H. J., & Bert, J. (1954). EEG changes during cinematographic presentation; moving picture activation of the EEG. *Electroencephalogr Clin Neurophysiol*, *6*(3), 433–444.

Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci*, *25*(20), 5004–5012.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn Sci*, *10*(6), 278–285.

Giard, M., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11(5)*, 473–490.

Giard, M., Perrin, F., Echallier, J., Thévenet, M., Froment, J., & Pernier, J. (1994). Dissociation of temporal and frontal components in the human auditory N1 wave: A scalp current density and dipole model analysis. *Electroencephalogr Clin Neurophysiol*, *92*(3), 238–252.

Gläscher, J., & Adolphs, R. (2003). Processing of the arousal of subliminal and supraliminal emotional stimuli by the human amygdala. *J Neurosci*, *23*(32), 10274–10282.

Glauser, E. S. D., & Scherer, K. R. (2008). Neuronal processes involved in subjective feeling emergence: Oscillatory activity during an emotional monitoring task. *Brain Topogr*, *20*(4), 224–231.

Gliga, T., & Dehaene-Lambertz, G. (2005). Structural encoding of body and face in human infants and adults. *J Cogn Neurosci*, *17*(8), 1328–1340.

Goeleven, E., De Raedt, R., Baert, S., & Koster, E. H. W. (2006). Deficient inhibition of emotional information in depression. *J Affect Disord*, *93*(1-3), 149–157.

Gollan, J. K., Pane, H. T., McCloskey, M. S., & Coccaro, E. F. (2008). Identifying differences in biased affective information processing in major depression. *Psychiatry Res*, *159*(1-2), 18–24.

Gorno-Tempini, M. L., Pradelli, S., Serafini, M., Pagnoni, G., Baraldi, P., Porro, C., et al. (2001). Explicit and incidental facial expression processing: An fMRI study. *Neuroimage*, *14*(2), 465–473.

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Current status and future directions. *Annu Rev Clin Psychol*, *6*, 285–312.

Gotlib, I. H., Kasch, K. L., Traill, S., Arnow, B. A., Johnson, S. L., & Joormann, J. (2004). Coherence and specificity of information-processing biases in depression and social phobia. *Journal of Abnormal Psychology*, *113*(3), 386–398.

Gotlib, I. H., Krasnoperova, E., Yue, D. N., & Joormann, J. (2004). Attentional biases for negative interpersonal stimuli in clinical depression. *Journal of Abnormal Psychology*, *113*(1), 127–135.

Gotlib, I. H., Sivers, H., Gabrieli, J. D. E., Whitfield-Gabrieli, S., Goldin, P., Minor, K. L., et al. (2005). Subgenual anterior cingulate activation to valenced emotional stimuli in major depression. *Neuroreport*, *16*(16), 1731–1734.

Grabska-Barwinska, A., & Zygierewicz, J. (2006). A model of event-related EEG synchronization changes in beta and gamma frequency bands. *Journal of Theoretical Biology*, *238*, 901–913.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*(2), 145–146.

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95-112.

Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Hum Brain Mapp*, *12*, 1-19.

Grèzes, J., Pichon, S., & de Gelder, B. (2007). Perceiving fear in dynamic body expressions. *Neuroimage*, *35*, 959–967.

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proc Natl Acad Sci U S A*, *98*(2), 694–699.

Gur, R. C., Erwin, R. J., Gur, R. E., Zwil, A. S., Heimberg, C., & Kraemer, H. C. (1992). Facial emotion discrimination: II. Behavioral findings in depression. *Psychiatry Research*, *42*(3), 241–251.

Gur, R. C., Schroeder, L., Turner, T., McGrath, C., Chan, R., Turetsky, B. I., et al. (2002). Brain activation during facial emotion processing. *Neuroimage*, *16*(3), 651–662.

Gur, R. E., McGrath, C., Chan, R., Schroeder, L., Turner, T., Turetsky, B. I., et al. (2002). An fMRI study of facial emotion processing in patients with schizophrenia. *American Journal of Psychiatry*, *159*(12), 1992–1999.

Hadjikhani, N., & de Gelder, B. (2003). Seeing fearful body expressions activates the fusiform cortex and amygdala. *Curr Biol*, *13*(24), 2201–2205.

Hadjikhani, N., Hoge, R., Snyder, J., & de Gelder, B. (2008). Pointing with the eyes: The role of gaze in communicating danger. *Brain Cogn*, *68(1)*, 1–8.

Haegens, S., Osipova, D., Oostenveld, R., & Jensen, O. (2010). Somatosensory working memory performance in humans depends on both engagement and disengagement of regions in a distributed network. *Hum Brain Mapp*, *31*(1), 26–35.

Hagan, C. C., Woods, W., Johnson, S., Calder, A. J., Green, G. G. R., & Young, A. W. (2009). MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus. *Proc Natl Acad Sci U S A*, *106*(47), 20010–20015.

Hajcak, G., & Nieuwenhuis, S. (2006). Reappraisal modulates the electrocortical response to unpleasant pictures. *Cognitive, Affective, & Behavioral Neuroscience*, *6(4)*, 291-297.

Hale, W. W. (1998). Judgment of facial expressions and depression persistence. *Psychiatry Research*, *80*(3), 265–274.

Hale, W. W., Jansen, J. H. C., Bouhuys, A. L., & van den Hoofdakker, R. H. (1998). The judgment of facial expressions by depressed patients, their partners and controls. *Journal of Affective Disorders*, *47*(1-3), 63–70.

Hamilton, J. P., & Gotlib, I. H. (2008). Neural substrates of increased memory sensitivity for negative stimuli in major depression. *Biological Psychiatry*, *63*(12), 1155–1162.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *J Pers Soc Psychol*, *54*(6), 917–924.

Happé, F. (1999). Autism: Cognitive deficit or cognitive style? *Trends Cogn Sci*, *3*(6), 216–222.

Hariri, A., Drabant, E., Munoz, K., Kolachana, B., Mattay, V., Egan, M., et al. (2005). A susceptibility gene for affective disorders and the response of the human amygdala. *Archives of General Psychiatry*, *62*(2), 146–152.

Harris, K. C., Mills, J. H., & Dubno, J. R. (2007). Electrophysiologic correlates of intensity discrimination in cortical evoked potentials of younger and older adults. *Hear Res*, *228*(1-2), 58–68.

Hauser, M. (2004). *The evolution of communication*. Cambridge, MA: MIT Press.

Hegerl, U., & Juckel, G. (1993). Intensity dependence of auditory evoked potentials as an indicator of central serotonergic neurotransmission: A new hypothesis. *Biol Psychiatry*, *33*(3), 173–187.

Hempel, A., Hempel, E., Schönknecht, P., Stippich, C., & Schröder, J. (2003). Impairment in basal limbic function in schizophrenia during affect recognition. *Psychiatry Research*, *122*(2), 115–124.

Herbert, C., Kissler, J., Junghöfer, M., Peyk, P., & Rockstroh, B. (2006). Processing of emotional adjectives: Evidence from startle EMG and ERPs. *Psychophysiology*, *43*(2), 197–206.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(108), 177–180.

Hinojosa, J. A., Carretié, L., Valcárcel, M. A., Méndez-Bértolo, C., & Pozo, M. A. (2009). Electrophysiological differences in the processing of affective information in words and pictures. *Cogn Affect Behav Neurosci*, *9*(2), 173–189.

Holle, H., Obleser, J., Rueschemeyer, S.-A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage*, *49*(1), 875–884.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6(2)*, 65-70.

Horstmann, G., & Ansorge, U. (2009). Visual search for facial expressions of emotions: A comparison of dynamic and static faces. *Emotion*, *9*(1), 29–38.

Höschel, K., & Irle, E. (2001). Emotional priming of facial affect identification in schizophrenia. *Schizophr Bull*, *27*(2), 317–327.

Hoshi, E., & Tanji, J. (2004). Functional specialization in dorsal and ventral premotor areas. *Prog Brain Res*, *143*, 507–511.

Huang, Y.-X., & Luo, Y.-J. (2006). Temporal course of emotional negativity bias: An ERP study. *Neurosci Lett*, *398*(1-2), 91–96.

Huettel, S., Song, A., & McCarthy, G. (2004). *Functional magnetic resonance imaging* (S. Huettel, Ed.). Sinauer: Sunderland, Mass.

Hurtado, E., Haye, A., González, R., Manes, F., & Ibanez, A. (2009). Contextual blending of ingroup/outgroup face stimuli and word valence: LPP modulation and convergence of measures. *BMC Neurosci*, *10*, 69.

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., et al. (1997). Vocal identification of speaker and emotion activates differerent brain regions. *Neuroreport*, *8*(12), 2809–2812.

Jensen, O., & Colgin, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends Cogn Sci*, *11*(7), 267–269.

Jensen, O., Gelfand, J., Kounios, J., & Lisman, J. E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. *Cereb Cortex*, *12*(8), 877–882.

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

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *Neuroimage*, *58*(2), 665–674.

Joassin, F., Pesenti, M., Maurage, P., Verreckt, E., Bruyer, R., & Campanella, S. (2011). Cross-modal interactions between human faces and voices involved in person recognition. *Cortex*, *47*(3), 367–376.

Johnston, P. J., Stojanov, W., Devir, H., & Schall, U. (2005). Functional MRI of facial emotion recognition deficits in schizophrenia and their electrophysiological correlates. *European Journal of Neuroscience*, *22*(5), 1221–1232.

Joormann, J., & Gotlib, I. H. (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology*, *115*(4), 705–714.

Joormann, J., & Gotlib, I. H. (2007). Selective attention to emotional faces following recovery from depression. *Journal of Abnormal Psychology*, *116*(1), 80–85.

Joormann, J., Talbot, L., & Gotlib, I. H. (2007). Biased processing of emotional information in girls at risk for depression. *Journal of Abnormal Psychology*, *116*(1), 135–143.

Kan, Y., Mimura, M., Kamijima, K., & Kawamura, M. (2004). Recognition of emotion from moving facial and prosodic stimuli in depressed patients. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(12), 1667–1671.

Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Res*, *1148*, 138–148.

Kanske, P., & Kotz, S. A. (2011). Emotion speeds up conflict resolution: A new role for the ventral anterior cingulate cortex? *Cereb Cortex*, *21*(4), 911–919.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci*, *17*(11), 4302–4311.

Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Struct Funct*, *212*(2), 121–132.

Kayser, J., Bruder, G. E., Tenke, C. E., Stewart, J. E., & Quitkin, F. M. (2000). Event-related potentials (ERPs) to hemifield presentations of emotional stimuli: Differences between depressed patients and healthy adults in P3 amplitude and asymmetry. *International Journal of Psychophysiology*, *36*(3), 211–236.

Kee, K. S., Kern, R. S., & Green, M. F. (1998). Perception of emotion and neurocognitive functioning in schizophrenia: What's the link? *Psychiatry Research*, *81*(1), 57–65.

Keil, A., Bradley, M. M., Hauk, O., Rockstroh, B., Elbert, T., & Lang, P. J. (2002). Large-scale neural correlates of affective picture processing. *Psychophysiology*, *39*(5), 641–649.

Keil, J., Müller, N., Ihssen, N., & Weisz, N. (2011). On the variability of the McGurk effect: Audiovisual integration depends on prestimulus brain states. *Cereb Cortex*.

Kerr, S. L., & Neale, J. M. (1993). Emotion perception in schizophrenia: Specific deficit or further evidence of generalized poor performance? *Journal of Abnormal Psychology*, *102*(2), 312–318.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., et al. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity

Survey Replication (NCS-R). *Journal of the American Medical Association*, *289*(23), 3095-3105.

Kilner, J. M., Baker, S. N., Salenius, S., Jousmäki, V., Hari, R., & Lemon, R. N. (1999). Task-dependent modulation of 15–30 Hz coherence between rectified EMGs from human hand and forearm muscles. *J Physiol*, *516 (Pt 2)*, 559–570.

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *Neuroimage*, *18*(1), 156–168.

Kissler, J., Herbert, C., Winkler, I., & Junghofer, M. (2009). Emotion and attention in visual word processing: An ERP study. *Biol Psychol*, *80*(1), 75–83.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Res Rev*, *53*(1), 63–88.

Kohler, C. G., Bilker, W., Hagendoorn, M., Gur, R. E., & Gur, R. C. (2000). Emotion recognition deficit in schizophrenia: Association with symptomatology and cognition. *Biological Psychiatry*, *48*(2), 127–136.

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2010). Facial emotion perception in schizophrenia: A meta-analytic review. *Schizophr Bull*, *36*(5), 1009–1019.

Komura, Y., Tamura, R., Uwano, T., Nishijo, H., & Ono, T. (2005). Auditory thalamus integrates visual inputs into behavioral gains. *Nat Neurosci*, *8*(9), 1203–1209.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., Cramon, D. Y. von, & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, *86*, 366—-376.

Kotz, S. A., Meyer, M., & Paulmann, S. (2006). Lateralization of emotional prosody in the brain: An overview and synopsis on the impact of study design. *Progress in Brain Research*, *156*, 285–294.

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Res*, *1151*, 107–118.

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, *5(3)*, 108-125.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *Neuroimage*, *37*(4), 1445–1456.

Kreifelts, B., Ethofer, T., Shiozawa, T., Grodd, W., & Wildgruber, D. (2009). Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia*, *47*(14), 3059–3066.

Kret, M. E., Pichon, S., Grèzes, J., & de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage*, *54*(2), 1755–1762.

Kucharska-Pietura, K., David, A. S., Masiak, M., & Phillips, M. L. (2005). Perception of facial and vocal affect by people with schizophrenia in early and late stages of illness. *British Journal of Psychiatry*, *187*, 523–528.

Kucharska-Pietura, K., Phillips, M. L., Gernand, W., & David, A. S. (2003). Perception of emotions from faces and voices following unilateral brain damage. *Neuropsychologia*, *41*(8), 1082–1090.

Kuchinke, L., Schneider, D., Kotz, S. A., & Jacobs, A. M. (2011). Spontaneous but not explicit processing of positive sentences impaired in Asperger's syndrome: Pupillometric evidence. *Neuropsychologia*, *49*(3), 331–338.

Kujawa, A. J., Torpey, D., Kim, J., Hajcak, G., Rose, S., Gotlib, I. H., et al. (2010). Attentional biases for emotional faces in young children of mothers with chronic or recurrent depression. *Journal of abnormal child psychology*, 125–135.

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2011). Altered auditory and multisensory temporal processing in autism spectrum disorders. *Front Integr Neurosci*, *4*, 129.

LaBar, K. S., Crupain, M. J., Voyvodic, J. T., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cereb Cortex*, *13*(10), 1023–1033.

Lachaux, J.-P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, *8*, 194-208.

Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*(2), 279–

292.

Lang, J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In *Technology in mental health care delivery systems* (pp. 119–137). Ablex, Norwood, NJ.

Latinus, M., & Belin, P. (2011). Human voice perception. *Curr Biol*, *21*(4), R143–R145.

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

Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B., & Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Hum Brain Mapp*, *19*(4), 213–223.

Lawrence, N. S., Williams, A. M., Surguladze, S., Giampietro, V., Brammer, M. J., Andrew, C., et al. (2004). Subcortical and ventral prefrontal cortical neural responses to facial expressions distinguish patients with bipolar disorder and major depression. *Biological Psychiatry*, *55*(6), 578–587.

LeDoux, J. E. (2000). Emotion Circuits in the Brain. *Annual Review in Neuroscience*, *23*, 155–184.

Lee, B. T., Seok, J. H., Lee, B. C., Cho, S. W., Yoon, B. J., Lee, K. U., et al. (2008). Neural correlates of affective processing in response to sad and angry facial stimuli in patients with major depressive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *32*(3), 778–785.

Leitman, D. I., Hoptman, M. J., Foxe, J. J., Saccente, E., Wylie, G. R., Nierenberg, J., et al. (2007). The neural substrates of impaired prosodic detection in schizophrenia and its sensorial antecedents. *American Journal of Psychiatry*, *164*(3), 474–482.

Lepage, M., Sergerie, K., Benoit, A., Czechowska, Y., Dickie, E., & Armony, J. L. (2011). Emotional face processing and flat affect in schizophrenia: Functional and structural neural correlates. *Psychol Med*, *41*(9), 1833–1844.

Leppänen, J. M., Milders, M., Bell, J. S., Terriere, E., & Hietanen, J. K. (2004). Depression biases the recognition of emotionally neutral faces. *Psychiatry Research*, *128*(2), 123–133.

Lewald, J., & Getzmann, S. (2011). When and where of auditory spatial processing in cortex: A novel approach using electrotomography. *PLoS One*, *6*(9), e25146.

Leyman, L., Raedt, R. D., Schacht, R., & Koster, E. H. W. (2007). Attentional biases for angry faces in unipolar depression. *Psychological Medicine*, *37*(3), 393–402.

Li, H., Chan, R., McAlonan, G. M., & Gong, Q.-y. (2010). Facial emotion processing in schizophrenia: A meta-analysis of functional neuroimaging data. *Schizophr Bull*, *36*(5), 1029–1039.

Linden, S. C., Jackson, M. C., Subramanian, L., Wolf, C., Green, P., Healy, D., et al. (2010). Emotion-cognition interactions in schizophrenia: Implicit and explicit effects of facial expression. *Neuropsychologia*, *48*(4), 997–1002.

Liu, B., Wang, Z., & Jin, Z. (2009). The integration processing of the visual and auditory information in videos of real-world events: An ERP study. *Neurosci Lett*, *461*(1), 7–11.

Liu, H., Hu, Z., Peng, D., Yang, Y., & Li, K. (2010). Common and segregated neural substrates for automatic conceptual and affective priming as revealed by event-related functional magnetic resonance imaging. *Brain Lang*, *112*(2), 121–128.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *J Exp Psychol Hum Percept Perform*, *31*(1), 210–220.

Love, S. A., Pollick, F. E., & Latinus, M. (2011). Cerebral correlates and statistical criteria of cross-modal face and voice integration. *Seeing Perceiving*, *24*(4), 351–367.

Luck, P., & Dowrick, C. F. (2004). Don't look at me in that tone of voice! Disturbances in the perception of emotion in facial expression and vocal intonation by depressed patients. *Primary Care and Mental Health*, *2*(2), 99–106.

Luck, S. (2005). An introduction to event-related potentials and their neural origins. In *An introduction to the event-related potential technique* (p. 1-49). Cambridge, Mass., MIT Press.

MacMillan, N., & Creelman, C. (2005). *Detection theory: A user's guide (2nd ed.).* Mahwah, N.J.: Lawrence Erlbaum Associates.

Magnée, M. J. C. M., Oranje, B., van Engeland, H., Kahn, R. S., & Kemner, C. (2009). Cross-sensory gating in schizophrenia and autism spectrum disorder: EEG evidence for impaired brain connectivity? *Neuropsychologia*, *47*(7), 1728–1732.

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

Maier, J. X., Luca, M. D., & Noppeney, U. (2011). Audiovisual asynchrony detection in human speech. *J Exp Psychol Hum Percept Perform*, *37*(1), 245–256.

Maier, J. X., Neuhoff, J. G., Logothetis, N. K., & Ghazanfar, A. A. (2004). Multisensory integration of looming signals by rhesus monkeys. *Neuron*, *43*(2), 177–181.

Marinkovic, K., & Halgren, E. (1998). Human brain potentials related to the emotional expression, repetition, and gender of faces. *Psychobiology*, *26(4)*, 348–356.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods*, *164*(1), 177–190.

Marrufo, M. V., Vaquero, E., Cardoso, M. J., & Gómez, C. M. (2001). Temporal evolution of alpha and beta bands during visual spatial attention. *Brain Res Cogn Brain Res*, *12*(2), 315–320.

Martin, F., Baudouin, J.-Y., Tiberghien, G., & Franck, N. (2005). Processing emotional expression and facial identity in schizophrenia. *Psychiatry Res*, *134*(1), 43–53.

Mayer, A. R., Teshiba, T. M., Franco, A. R., Ling, J., Shane, M. S., Stephen, J. M., et al. (2011). Modeling conflict and error in the medial frontal cortex. *Hum Brain Mapp*.

Mayes, A. K., Pipingas, A., Silberstein, R. B., & Johnston, P. (2009). Steady state visually evoked potential correlates of static and dynamic emotional face processing. *Brain Topography*, *22*(3), 145–157.

McGettigan, C., Agnew, Z. K., Walsh, E., Jessop, R., & Scott, S. (2011). Activation of medial brodmann area 10 predicts behavioural accuracy in the classification of posed and genuine laughter samples. In *Program no. 606.11. 2011 neuroscience meeting planner. washington, dc: Society for neuroscience, 2011. online.*

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

Medendorp, W. P., Kramer, G. F. I., Jensen, O., Oostenveld, R., Schoffelen, J.-M., & Fries, P. (2007). Oscillatory activity in human parietal and occipital cortex shows hemispheric lateralization and memory effects in a delayed double-step saccade task. *Cereb Cortex*, *17*(10), 2364–2374.

Meeren, H. K. M., Hadjikhani, N., Ahlfors, S. P., Hämäläinen, M. S., & de Gelder, B. (2008). Early category-specific cortical activation revealed by visual stimulus inversion. *PLoS One*, *3*(10), e3503.

Meeren, H. K. M., van Heijnsbergen, C. C. R. J., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proc Natl Acad Sci U S A*, *102*(45), 16518–16523.

Mishkin, M., Ungerleider, L., & Macko, K. (1983). Object vision and spatial vision: Two cortial pathways. *Trends Neurosci*, *6*, 414–417.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cereb Cortex*, *14*(4), 452–465.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Brain Res Cogn Brain Res*, *14*(1), 115–128.

Monk, C. S., Klein, R. G., Telzer, E. H., Schroth, E. A., Mannuzza, S., Moulton III, J. L., et al. (2008). Amygdala and nucleus accumbens activation to emotional facial expressions in children and adolescents at risk for major depression. *American Journal of Psychiatry*, *165*(1), 90–98.

Morris, J., Ohman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc Natl Acad Sci U S A*, *96*(4), 1680–1685.

Morris, J., Scott, S., & Dolan, R. (1999). Saying it with feeling: Neural responses to emotional vocalizations. *Neuropsychologia*, *37*, 1155-1163.

Moser, J., Hajcak, G., Bukay, E., & Simons, R. F. (2006). Intentional modulation of emotional responding to unpleasant pictures: An ERP study. *Psychophysiology*, *43*, 92-296.

Müller, V. I., Habel, U., Derntl, B., Schneider, F., Zilles, K., Turetsky, B. I., et al. (2011). Incongruence effects in crossmodal emotional integration. *Neuroimage*, *54*(3), 2257–2266.

Münte, T. F., Brack, M., Grootheer, O., Wieringa, B. M., Matzke, M., & Johannes, S. (1998). Brain potentials reveal the timing of face identity and expression judgments. *Neuroscience Research*, *30*(1), 25–34.

Murphy, P., Brady, N., Fitzgerald, M., & Troje, N. F. (2009). No evidence for impaired perception of biological motion in adults with autistic spectrum disorders. *Neuropsychologia*, *47*(14), 3225–3235.

Muthukumaraswamy, S. D., Johnson, B. W., Gaetz, W. C., & Cheyne, D. O. (2006). Neural processing of observed oro-facial movements reflects multiple action encoding strategies in the human brain. *Brain Res*, *1071*(1), 105–112.

Muthukumaraswamy, S. D., Johnson, B. W., & McNair, N. A. (2004). Mu rhythm modulation during observation of an object-directed grasp. *Brain Res Cogn Brain Res*, *19*(2), 195–201.

Muthukumaraswamy, S. D., & Singh, K. D. (2008). Modulation of the human mirror neuron system during cognitive activity. *Psychophysiology*, *45*(6), 896–905.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol*, *118*(12), 2544–2590.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425.

Nakagawa, K., Aokage, Y., Fukuri, T., Kawahara, Y., Hashizume, A., Kurisu, K., et al. (2011). Neuromagnetic beta oscillation changes during motor imagery and motor execution of skilled movements. *Neuroreport*, *22*(5), 217–222.

Naranjo, C., Kornreich, C., Campanella, S., Noël, X., Vandriette, Y., Gillain, B., et al. (2010). Major depression is associated with impaired processing of emotion in music as well as in facial and vocal stimuli. *Journal of Affective Disorders*.

Neumeister, A., Drevets, W. C., Belfer, I., Luckenbaugh, D. A., Henry, S., Bonne, O., et al. (2006). Effects of a α2C-adrenoreceptor gene polymorphism on neural responses to facial expressions in depression. *Neuropsychopharmacology*, *31*(8), 1750–1756.

Neuper, C., Scherer, R., Wriessnegger, S., & Pfurtscheller, G. (2009). Motor imagery and action observation: Modulation of sensorimotor brain rhythms during mental control of a brain-computer interface. *Clin Neurophysiol*, *120*(2), 239–247.

Nichols, S. (2005). Innateness and moral psychology. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind, Vol. II: Structure and content.* New York: Oxford University

Press.

Nishijo, H., Ono, T., & Nishino, H. (1988). Topographic distribution of modality-specific amygdalar neurons in alert monkey. *J Neurosci*, *8*(10), 3556–3569.

Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cereb Cortex*, *18*(3), 598–609.

Oatley, K., & Johnson-Laird, P. (1987). Towards a cognitive theory of emotions. *Cognition & Emotion*, *1(1)*, 29–50.

Oberman, L., McCleery, J., Ramachandran, V., & Pineda, J. A. (2007). EEG evidence for mirror neuron activity during the observation of human and robot actions: Towards an analysis of the human qualities of interactive robots. *Neurocomputing*, *70*, 2194-2203.

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage*, *55*(2), 713–723.

O'Doherty, J., Rolls, E., & Kringelbach, M. (2004). Neuroimaging studies of cross-modal integration for emotion. In G. Calvert, C. Spence, & B. Stein (Eds.), *The handbook of multisensory processes* (pp. 563–579). Cambridge, M.A., MIT Press.

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology*, *77*(3), 247–265.

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

Otten, L., & Rugg, M. (2005). Interpreting event-related brain potentials. In Handy (Ed.), *Event-related potential. A methods handbook.* (p. 3-16). Cambridge, Mass., MIT Press.

Palva, S., & Palva, J. M. (2007). New vistas for alpha-frequency band oscillations. *Trends Neurosci*, *30*(4), 150–158.

Park, J.-Y., Gu, B.-M., Kang, D.-H., Shin, Y.-W., Choi, C.-H., Lee, J.-M., et al. (2011). Integration of cross-modal emotional information in the human brain: An fMRI study. *Cortex*, *46*, 161-169.

Paulmann, S., Jessen, S., & Kotz, S. (2009). Investigating the multimodal nature of human communication. Insights from ERPs. *Journal of Psychophysiology*, *23(2)*, 63-76.

Paulmann, S., & Kotz, S. A. (2008a). Early emotional prosody perception based on different speaker voices. *Neuroreport*, *19*(2), 209–213.

Paulmann, S., & Kotz, S. A. (2008b). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain Lang*, *105*(1), 59–69.

Peelen, M. V., Atkinson, A. P., Andersson, F., & Vuilleumier, P. (2007). Emotional modulation of body-selective visual areas. *Soc Cogn Affect Neurosci*, *2*(4), 274–283.

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *J Neurosci*, *30*(30), 10127–10134.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *J Neurophysiol*, *93*(1), 603–608.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nat Rev Neurosci*, *8*(8), 636–648.

Perry, A., Bentin, S., Shalev, I., Israel, S., Uzefovsky, F., Bar-On, D., et al. (2010). Intranasal oxytocin modulates EEG mu/alpha and beta rhythms during perception of biological motion. *Psychoneuroendocrinology*, *35*(10), 1446–1453.

Perry, A., Troje, N. F., & Bentin, S. (2010). Exploring motor system contributions to the perception of social information: Evidence from EEG activity in the mu/alpha frequency range. *Social Neuroscience*, *5(3)*, 272-284.

Petrini, K., Pollick, F. E., Dahl, S., McAleer, P., McKay, L. S., McKay, L., et al. (2011). Action expertise reduces brain activity for audiovisual matching actions: An fMRI study with expert drummers. *Neuroimage*, *56*(3), 1480–1492.

Pfurtscheller, G., & Silva, F. H. L. da. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin Neurophysiol*, *110*(11), 1842–1857.

Pichon, S., de Gelder, B., & Grèzes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Soc Neurosci*, *3*(3-4), 199–212.

Pichon, S., de Gelder, B., & Grèzes, J. (2009). Two different faces of threat. Comparing the neural systems for recognizing fear and anger in dynamic body expressions. *Neuroimage*, *47*, 1873–1883.

Pineda, J. A. (2005). The functional significance of mu rhythms: Translating "seeing" and "hearing" into "doing". *Brain Res Brain Res Rev*, *50*(1), 57–68.

Pizzagalli, D., Regard, M., & Lehmann, D. (1999). Rapid emotional face processing in the human right and left brain hemispheres: An ERP study. *Neuroreport*, *10*(13), 2691–

2698.

Pollick, F. E., Kay, J. W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. *J Exp Psychol Hum Percept Perform*, *31*(6), 1247–1265.

Pomarol-Clotet, E., Hynes, F., Ashwin, C., Bullmore, E. T., McKenna, P. J., & Laws, K. R. (2010). Facial emotion processing in schizophrenia: A non-specific neuropsychological deficit? *Psychological Medicine*, *40*(6), 911–919.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, *11*(6), 1329–1333.

Pourtois, G., Debatisse, D., Despland, P.-A., & de Gelder, B. (2002). Facial expressions modulate the time course of long latency auditory brain potentials. *Brain Res Cogn Brain Res*, *14*(1), 99–105.

Pourtois, G., Gelder, B. de, Bol, A., & Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, *41*(1), 49–59.

Puce, A., Epling, J. A., Thompson, J. C., & Carrick, O. K. (2007). Neural responses elicited to face motion and vocalization pairings. *Neuropsychologia*, *45*(1), 93–106.

Quintana, J., Lee, J., Marcus, M., Kee, K., Wong, T., & Yerevanian, A. (2011). Brain dysfunctions during facial discrimination in schizophrenia: Selective association to affect decoding. *Psychiatry Res*, *191*(1), 44–50.

Quintana, J., Wong, T., Ortiz-Portillo, E., Marder, S. R., & Mazziotta, J. C. (2003). Right lateral fusiform gyrus dysfunction during facial information processing in schizophrenia. *Biological Psychiatry*, *53*(12), 1099–1112.

Rauch, A. V., Reker, M., Ohrmann, P., Pedersen, A., Bauer, J., Dannlowski, U., et al. (2010). Increased amygdala activation during automatic processing of facial emotion in schizophrenia. *Psychiatry Research*, *182*(3), 200–206.

Riby, D. M., Doherty-Sneddon, G., & Bruce, V. (2008). Exploring face perception in disorders of development: Evidence from Williams syndrome and autism. *J Neuropsychol*, *2*(Pt 1), 47–64.

Ritter, P., Moosmann, M., & Villringer, A. (2009). Rolandic alpha and beta EEG rhythms' strengths are inversely related to fMRI-BOLD signal in primary somatosensory and motor cortex. *Hum Brain Mapp*, *30*(4), 1168–1187.

Robins, D. L., Hunyadi, E., & Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain Cogn*, *69*(2), 269–278.

Romei, V., Murray, M. M., Cappe, C., & Thut, G. (2009). Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Curr Biol*, *19*(21), 1799–1805.

Rosburg, T., Boutros, N. N., & Ford, J. M. (2008). Reduced auditory evoked potential component N100 in schizophrenia – a critical review. *Psychiatry Res*, *161*(3), 259–274.

Rosenthal, R. (1991). *Meta-analytic procedures for social research (revised).* Newbury Park, CA: Sage.

Rottenberg, J., & Gotlib, I. (2004). Mood disorders: A handbook of science and practice. In M. Power (Ed.), (pp. 61–77). Chichester, UK: Wiley.

Roux, P., Christophe, A., & Passerieux, C. (2010). The emotional paradox: Dissociation between explicit and implicit processing of emotional prosody in schizophrenia. *Neuropsychologia*.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol Bull*, *115*(1), 102–141.

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., et al. (2011). Emotional perception: Meta-analyses of face and natural scene processing. *Neuroimage*, *54*(3), 2524–2533.

Sachs, G., Steger-Wuchse, D., Kryspin-Exner, I., Gur, R. C., & Katschnig, H. (2004). Facial recognition deficits and cognition in schizophrenia. *Schizophrenia Research*, *68*(1), 27–35.

Salgado-Pineda, P., Fakra, E., Delaveau, P., Hariri, A. R., & Blin, O. (2010). Differential patterns of initial and sustained responses in amygdala and cortical regions to emotional stimuli in schizophrenia patients and healthy participants. *Journal of Psychiatry & Neuroscience*, *35*(1), 41–48.

Sánchez-García, C., Alsius, A., Enns, J. T., & Soto-Faraco, S. (2011). Cross-modal prediction in speech perception. *PLoS One*, *6*(10), e25198.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Rev Neurosci*, *14*(4), 303–316.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *Neuroimage*, *28*(4), 848–858.

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: An fMRI study. *Brain Res Cogn Brain Res*, *20*(1), 81–91.

Sato, W., & Yoshikawa, S. (2007). Enhanced experience of emotional arousal in response to dynamic facial expressions. *J Nonverbal Behav*, *31(2)*, 119–135.

Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *J Cogn Neurosci*, *22*(3), 474–481.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc Natl Acad Sci U S A*, *107*(6), 2408–2412.

Saygin, A. P., Wilson, S. M., Hagle Jr., D. J., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *The Journal of Neuroscience*, *24(27)*, 6181-6188.

Schaefer, K. L., Baumann, J., Rich, B. A., Luckenbaugh, D. A., & Zarate Jr, C. A. (2010). Perception of facial emotion in adults with bipolar or unipolar depression and controls. *Journal of Psychiatric Research*.

Scherer, K. R. (1995). Expression of emotion in voice and music. *J Voice*, *9*(3), 235–248.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends Cogn Sci*, *10*(1), 24–30.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*(2), 228–233.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research*, *24*(3), 442–452.

Schneider, F., Weiss, U., Kessler, C., Salloum, J. B., Posse, S., Grodd, W., et al. (1998). Differential amygdala activation in schizophrenia during sadness. *Schizophrenia Research*, *34*(3), 133–142.

Schneider, T., 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. *J Neurosci*, *31*(7), 2502–2510.

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

Schupp, H., Cuthbert, B., Bradley, M., Cacioppo, J., Ito, T., & Lang, P. (2000). Affective picture processing: The late positive potential is modulated by motivational relevance. *Psychophysiology*, *37*, 257-261.

Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biol Psychol*, *80*(1), 95–104.

Scott, S., Sauter, D., & McGettigan, C. (2009). Brain mechanisms for processing perceived emotional vocalizations in humans. In S. Brudzynksi (Ed.), *Handbook of mammalian vocalizations* (pp. 187–198). Oxford: Academic Press.

Scott, S., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature*, *385*(6613), 254–257.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res*, *47*(3), 277–287.

Senkowski, D., Gomez-Ramirez, M., Lakatos, P., Wylie, G. R., Molholm, S., Schroeder, C. E., et al. (2007). Multisensory processing and oscillatory activity: Analyzing non-linear electrophysiological measures in humans and simians. *Exp Brain Res*, *177*, 184-195.

Senkowski, D., Molholm, S., Gomez-Ramirez, M., & Foxe, J. J. (2006). Oscillatory beta activity predicts response speed during a multisensory audiovisual reaction time task: A high-density electrical mapping study. *Cereb Cortex*, *16*(11), 1556–1565.

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage*, *56*, 2200-2208.

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

Senkowski, D., Schneider, T. R., Tandler, F., & Engel, A. K. (2009). Gamma-band activity reflects multisensory matching in working memory. *Exp Brain Res*, *198*(2-3), 363–372.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia*, *45*(3), 561–571.

Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neurosci Biobehav Rev*, *32*(4), 811–830.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain, and cognitive control in the cingulate cortex. *Nat Rev Neurosci*, *12(3)*, 154–167.

Shahin, A. J., Picton, T. W., & Miller, L. M. (2009). Brain oscillations during semantic evaluation of speech. *Brain Cogn*, *70*(3), 259–266.

Sharbrough, F., Chatrian, G., Lesser, R., Lüders, H., Nuwer, M., & Picton, T. (1991). Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *8*, 200-202.

Sheline, Y. I., Barch, D. M., Donnelly, J. M., Ollinger, J. M., Snyder, A. Z., & Mintun, M. A. (2001). Increased amygdala response to masked emotional faces in depressed subjects resolves with antidepressant treatment: An fMRI study. *Biological Psychiatry*, *50*(9), 651–658.

Shepherd, G. M. (2004). The human sense of smell: Are we better than we think? *PLoS Biol*, *2*(5), E146.

Sherman, S. M., & Guillery, R. W. (2002). The role of the thalamus in the flow of information to the cortex. *Philos Trans R Soc Lond B Biol Sci*, *357*(1428), 1695–1708.

Siegle, G. J., Thompson, W., Carter, C. S., Steinhauer, S. R., & Thase, M. E. (2007). Increased amygdala and decreased dorsolateral prefrontal BOLD responses in unipolar depression: Related and independent features. *Biological Psychiatry*, *61*(2), 198–209.

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends Cogn Sci*, *13*(8), 334–340.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci*, *18*, 555–586.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, *25*(1), 76–89.

Sloan, D. M., Bradley, M. M., Dimoulas, E., & Lang, P. J. (2002). Looking at facial expressions: Dysphoria and facial EMG. *Biological Psychology*, *60*(2-3), 79–90.

Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res Cogn Brain Res*, *17*(1), 75–82.

Sokolov, A. A., Erb, M., Gharabaghi, A., Grodd, W., Tatagiba, M. S., & Pavlova, M. A. (2011). Biological motion processing: The left cerebellum communicates with the right superior temporal sulcus. *Neuroimage*.

Sokolov, A. A., Gharabaghi, A., Tatagiba, M. S., & Pavlova, M. (2010). Cerebellar engagement in an action observation network. *Cereb Cortex*, *20*(2), 486–491.

Sperdin, H. F., Cappe, C., Foxe, J. J., & Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Front Integr Neurosci*, *3*, 2.

Spezio, M. L., Adolphs, R., Hurley, R. S. E., & Piven, J. (2007). Abnormal use of facial information in high-functioning autism. *J Autism Dev Disord*, *37*(5), 929–939.

Stanford, T., & Stein, B. (2007). Superadditivity in multisensory integration: Putting the computation in context. *Neuroreport*, *18*(8), 787–792.

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

Stein, B., & Meredith, M. (1993). *The merging of the senses.* MIT Press, Cambridge, Massachusetts.

Stein, B., Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: Alternative criteria, models, and inverse effectiveness. *Exp Brain Res*, *198*(2-3), 113–126.

Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: An ERP study on the body-inversion effect. *Neuroreport*, *15*(5), 777–780.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci*, *19*(12), 1964–1973.

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.

Sugimoto, S., Nittono, H., & Hori, T. (2007). Visual emotional context modulates brain potentials elicited by unattended tones. *Int J Psychophysiol*, *66*(1), 1–9.

Suied, C., Bonneel, N., & Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Exp Brain Res*, *194*(1), 91–102.

Sumby, W., & Pollack, I. (1954). Visual contributions to speech intelligibility in noise. *J Acoust Soc Am*, *26*, 212-215.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends Cogn Sci*, *13*(9), 403–409.

Surguladze, S., Brammer, M. J., Keedwell, P., Giampietro, V., Young, A. W., Travis, M. J., et al. (2005). A differential pattern of neural response toward sad versus happy facial expressions in major depressive disorder. *Biological Psychiatry*, *57*(3), 201–209.

Surguladze, S., Young, A. W., Senior, C., Brébion, G., Travis, M. J., & Phillips, M. L. (2004). Recognition accuracy and response bias to happy and sad facial expressions in patients with major depression. *Neuropsychology*, *18*(2), 212–218.

Suslow, T., Droste, T., Roestel, C., & Arolt, V. (2005). Automatic processing of facial emotion in schizophrenia with and without affective negative symptoms. *Cognitive Neuropsychiatry*, *10*(1), 35–56.

Suslow, T., Konrad, C., Kugel, H., Rumstadt, D., Zwitserlood, P., Schöning, S., et al. (2010). Automatic mood-congruent amygdala responses to masked facial expressions in major depression. *Biological Psychiatry*, *67*(2), 155–160.

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

Thomas, J. P., & Shiffrar, M. (2010). I can see you better if I can hear you coming: Action-consistent sounds facilitate the visual detection of human gait. *Journal of Vision*, *10(12):14*, 1-11.

Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, *292*(5515), 290–293.

Tomlinson, E. K., Jones, C. A., Johnston, R. A., Meaden, A., & Wink, B. (2006). Facial emotion recognition from moving and static point-light images in schizophrenia. *Schizophrenia Research*, *85*(1-3), 96–105.

Trautmann, S. A., Fehr, T., & Herrmann, M. (2009). Emotions in motion: Dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Research*, *1284*, 100–115.

Turetsky, B. I., Kohler, C. G., Indersmitten, T., Bhati, M. T., Charbonnier, D., & Gur, R. C. (2007). Facial emotion recognition in schizophrenia: When and why does it go awry? *Schizophrenia Research*, *94*(1-3), 253–263.

Tzagarakis, C., Ince, N. F., Leuthold, A. C., & Pellizzer, G. (2010). Beta-band activity during motor planning reflects response uncertainty. *J Neurosci*, *30*(34), 11270–11277.

Uekermann, J., Abdel-Hamid, M., Lehmkämper, C., Vollmoeller, W., & Daum, I. (2008). Perception of affective prosody in major depression: A link to executive functions? *Journal of the International Neuropsychological Society*, *14*(04), 552–561.

Ulloa, E. A., & Pineda, J. A. (2007). Recognition of point-light biological motion: Mu rhythm and mirror neuron activity. *Behavioral Brain Research*, *183*, 188-194.

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*(2), 271–282.

van de Riet, W. A. C., Grèzes, J., & de Gelder, B. (2009). Specific and common brain regions involved in the perception of faces and bodies and the representation of their emotional expressions. *Soc Neurosci*, *4*(2), 101–120.

van den Stock, J., de Jong, J., Hodiamont, P., & de Gelder, B. (in press). Perceiving emotions from bodily expressions and multisensory integration of emotion cues in schizophrenia. *Social Neuroscience*.

van den Stock, J., Grèzes, J., & de Gelder, B. (2008). Human and animal sounds influence recognition of body language. *Brain Res*, *1242*, 185–190.

van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion*, *7*(3), 487–494.

van Heijnsbergen, C. C. R. J., Meeren, H. K. M., Grèzes, J., & de Gelder, B. (2007). Rapid detection of fear in body expressions: An ERP study. *Brain Res*, *1186*, 233–241.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A*, *102*(4), 1181–1186.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607.

van Wingen, G. A., van Eijndhoven, P., Tendolkar, I., Buitelaar, J., Verkes, R. J., & Fernández, G. (2011). Neural basis of emotion recognition deficits in first-episode major depression. *Psychol Med*, *41*(7), 1397–1405.

Vandoolaeghe, E., Hunsel, F. van, Nuyten, D., & Maes, M. (1998). Auditory event related potentials in major depression: Prolonged P300 latency and increased P200 amplitude. *Journal of Affective Disorders*, *48*(2–3), 105–113.

Victor, T. A., Furey, M. L., Fromm, S. J., Ohman, A., & Drevets, W. C. (2010). Relationship between amygdala responses to masked faces and mood state and treatment in major depressive disorder. *Arch Gen Psychiatry*, *67*(11), 1128–1138.

von Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche. *Annalen der Physik und Chemie, Leipzig*, *89*, 211–233.

Vroomen, J., Driver, J., & Gelder, B. de. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cogn Affect Behav Neurosci*, *1*(4), 382–387.

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J Cogn Neurosci*, *22*(7), 1583–1596.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron*, *30*(3), 829–841.

Vuilleumier, P., & Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia*, *45*(1), 174–194.

Wambacq, I. J., Shea-Miller, K. J., & Abubakr, A. (2004). Non-voluntary and voluntary processing of emotional prosody: An event–related potentials study. *Neuroreport*, *15*(3), 555–559.

Wang, L., Mottron, L., Peng, D., Berthiaume, C., & Dawson, M. (2007). Local bias and local-to-global interference without global deficit: A robust finding in autism under various conditions of attention, exposure time, and visual angle. *Cogn Neuropsychol*, *24*(5), 550–574.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn Sci*, *7*(12), 553–559.

Werner, S., & Noppeney, U. (2010a). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J Neurosci*, *30*(7), 2662–2675.

Werner, S., & Noppeney, U. (2010b). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex*, *20*(8), 1829–1842.

Westbrook, G. (2000). Seizures and epilepsy. In E. Kandel, J. Schwartz, & T. Jessel (Eds.), *Principles of neural science. 4th edition.* (pp. 910–936). New York, NY, McGraw-Hill Pub.

Wheaton, K. J., Thompson, J. C., Syngeniotis, A., Abbott, D. F., & Puce, A. (2004). Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. *Neuroimage*, *22*(1), 277–288.

Wiethoff, S., Wildgruber, D., Grodd, W., & Ethofer, T. (2009). Response and habituation of the amygdala during processing of emotional prosody. *Neuroreport*, *20*(15), 1356–1360.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, *156*, 249–268.

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex*, *14*(12), 1384–1389.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., et al. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage*, *24*(4), 1233–1241.

Williams, L. M., Das, P., Liddell, B. J., Olivieri, G., Peduto, A. S., David, A. S., et al. (2007). Fronto-limbic and autonomic disjunctions to negative emotion distinguish schizophrenia subtypes. *Psychiatry Research*, *155*(1), 29–44.

Winston, J. S., O'Doherty, J., & Dolan, R. J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *Neuroimage*, *20*(1), 84–97.

Wunderlich, J. L., & Cone-Wesson, B. K. (2001). Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *J Acoust Soc Am*, *109*(4), 1526–1537.

Wynn, J. K., Lee, J., Horan, W. P., & Green, M. F. (2008). Using event related potentials to explore stages of facial affect recognition deficits in schizophrenia. *Schizophrenia Bulletin*, *34*(4), 679–687.

Yoshikawa, S., & Sato, W. (2006). Enhanced perceptual, emotional, and motor processing in response to dynamic facial expressions of emotions. *Japanese Psychological Research*, *48(3)*, 213–222.

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What you see is not (always) what you hear: Induced gamma band responses reflect cross-modal interactions in familiar object recognition. *J Neurosci*, *27*(5), 1090–1096.

Zahorik, P. (2009). Perceptually relevant parameters for virtual listening simulation of small room acoustics. *J Acoust Soc Am*, *126*(2), 776–791.

Curriculum Vitae

Name	Sarah Jessen
Date of Birth	26.11.1983
Place of Birth	Bielefeld, Germany

Academic Education

since 2008	PhD Student Max Planck Institute for Human Cognitive and Brain Sciences and Graduate School “Languages of Emotion” at the Free University of Berlin
2008	Master of Science in Neural and Behavioral Sciences International Max Planck Research School University of Tübingen
2006	Bachelor of Science in Cognitive Science University of Osnabrück
2002	Abitur (Diploma from German secondary school) Berlin, Germany

Career

2006-2007	Certificate Program “Biological and Cultural Principles of Human Thought” Forum Scientiarum, University Tübingen
10/07–12/07	Intern at the Interdisciplinary Institute for Biochemistry, University of Tübingen
08/07–10/07	Intern at the Max Planck Institute for Human Cognitive and Brain Sciences
02/07–04/07	Intern at the MEG-Center, University of Tübingen
2005–2006	Tutor for Computational Linguistics, University of Osnabrück

Awards

2011	DAAD travel grant (Japan)
2011	DRS travel grant (USA)
2008	3-year PhD-stipend of the Graduate School “Languages of Emotion”

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorgelegte Arbeit selbständig verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht verwendet. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Berlin, d. 15.12.2011

Sarah Jessen

Aus dieser Dissertation hervorgegangene Publikationen

Kapitel 4 (EEG-Study 1) ist erschienen in *NeuroImage*:

Jessen, S., & Kotz, S. A. (2011). The Temporal Dynamics of Processing Emotions from Vocal, Facial, and Bodily Expressions. Neuroimage, 58(2), 665-674.

Kapitel 5 (fMRI-Study) ist in modifizierter Fassung eingereicht bei *Cerebral Cortex*:

Jessen, S., & Kotz, S. A. (submitted to Cerebral Cortex). Differential Audiovisual Interactions for Emotional and Non-emotional Information.

Kapitel 6 (EEG-Study 2) ist erschienen in *PLoS ONE*:

Jessen, S., Obleser, J., & Kotz, S. A. (2012). How Bodies and Voices Interact in Early Emotion Perception. PLoS ONE, 7(4), e36070.

Kapitel 7 (Review-Artikel über Emotionswahrnehmung in Psychiatrischen Erkrankungen) ist erschienen in *Social Neuroscience*:

Garrido-Vásquez, P.*, Jessen, S.*, & Kotz, S. A. (2011). Perception of Emotion in Psychiatric Disorders: On the Possible Role of Task, Dynamics, and Multimodality. Social Neuroscience, 6(5-6), 515-536. * = both authors contributed equally.

MPI Series in Human Cognitive and Brain Sciences:

1 Anja Hahne
Charakteristika syntaktischer und semantischer Prozesse bei der auditiv Sprachverarbeitung: Evidenz aus ereigniskorrelierten Potentialstudien

2 Ricarda Schubotz
Erinnern kurzer Zeitdauern: Behaviorale und neurophysiologische Korrelate einer Arbeitsgedächtnisfunktion

3 Volker Bosch
Das Halten von Information im Arbeitsgedächtnis: Dissoziationen langsamer corticaler Potentiale

4 Jorge Jovicich
An investigation of the use of Gradient- and Spin-Echo (GRASE) imaging for functional MRI of the human brain

5 Rosemary C. Dymond
Spatial Specificity and Temporal Accuracy in Functional Magnetic Resonance Investigations

6 Stefan Zysset
Eine experimentalpsychologische Studie zu Gedächtnisabrufprozessen unter Verwendung der funktionellen Magnetresonanztomographie

7 Ulrich Hartmann
Ein mechanisches Finite-Elemente-Modell des menschlichen Kopfes

8 Bertram Opitz
Funktionelle Neuroanatomie der Verarbeitung einfacher und komplexer akustischer Reize: Integration haemodynamischer und elektrophysiologischer Maße

9 Gisela Müller-Plath
Formale Modellierung visueller Suchstrategien mit Anwendungen bei der Lokalisation von Hirnfunktionen und in der Diagnostik von Aufmerksamkeitsstörungen

10 Thomas Jacobsen
Characteristics of processing morphological structural and inherent case in language comprehension

11 Stefan Kölsch
Brain and Music
A contribution to the investigation of central auditory processing with a new electrophysiological approach

12 Stefan Frisch
Verb-Argument-Struktur, Kasus und thematische Interpretation beim Sprachverstehen

13 Markus Ullsperger
The role of retrieval inhibition in directed forgetting – an event-related brain potential analysis

14 Martin Koch
Measurement of the Self-Diffusion Tensor of Water in the Human Brain

15 Axel Hutt
Methoden zur Untersuchung der Dynamik raumzeitlicher Signale

16 Frithjof Kruggel
Detektion und Quantifizierung von Hirnaktivität mit der funktionellen Magnetresonanztomographie

17 Anja Dove
Lokalisierung an internen Kontrollprozessen beteiligter Hirngebiete mithilfe des Aufgabenwechselparadigmas und der ereigniskorrelierten funktionellen Magnetresonanztomographie

18 Karsten Steinhauer
Hirnphysiologische Korrelate prosodischer Satzverarbeitung bei gesprochener und geschriebener Sprache

19 Silke Urban
Verbinformationen im Satzverstehen

20 Katja Werheid
Implizites Sequenzlernen bei Morbus Parkinson

21 Doreen Nessler
Is it Memory or Illusion? Electrophysiological Characteristics of True and False Recognition

22 Christoph Herrmann
Die Bedeutung von 40-Hz-Oszillationen für kognitive Prozesse

23 Christian Fiebach
Working Memory and Syntax during Sentence Processing.
A neurocognitive investigation with event-related brain potentials and functional magnetic resonance imaging

24 Grit Hein
Lokalisation von Doppelaufgabendefiziten bei gesunden älteren Personen und neurologischen Patienten

25 Monica de Filippis
Die visuelle Verarbeitung unbeachteter Wörter. Ein elektrophysiologischer Ansatz

26 Ulrich Müller
Die katecholaminerge Modulation präfrontaler kognitiver Funktionen beim Menschen

27 Kristina Uhl
Kontrollfunktion des Arbeitsgedächtnisses über interferierende Information

28 Ina Bornkessel
The Argument Dependency Model: A Neurocognitive Approach to Incremental Interpretation

29 Sonja Lattner
Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformationen

30 Christin Grünewald
Die Rolle motorischer Schemata bei der Objektrepräsentation: Untersuchungen mit funktioneller Magnetresonanztomographie

31 Annett Schirmer
Emotional Speech Perception: Electrophysiological Insights into the Processing of Emotional Prosody and Word Valence in Men and Women

32 André J. Szameitat
Die Funktionalität des lateral-präfrontalen Cortex für die Verarbeitung von Doppelaufgaben

33 Susanne Wagner
Verbales Arbeitsgedächtnis und die Verarbeitung ambiger Wörter in Wort- und Satzkontexten

34 Sophie Manthey
Hirn und Handlung: Untersuchung der Handlungsrepräsentation im ventralen prämotorischen Cortex mit Hilfe der funktionellen Magnet-Resonanz-Tomographie

35 Stefan Heim
Towards a Common Neural Network Model of Language Production and Comprehension: fMRI Evidence for the Processing of Phonological and Syntactic Information in Single Words

36 Claudia Friedrich
Prosody and spoken word recognition: Behavioral and ERP correlates

37 Ulrike Lex
Sprachlateralisierung bei Rechts- und Linkshändern mit funktioneller Magnetresonanztomographie

38 Thomas Arnold
Computergestützte Befundung klinischer Elektroenzephalogramme

39 Carsten H. Wolters
Influence of Tissue Conductivity Inhomogeneity and Anisotropy on EEG/MEG based Source Localization in the Human Brain

40 Ansgar Hantsch
Fisch oder Karpfen? Lexikale Aktivierung von Benennungsalternative bei der Objektbenennung

41 Peggy Bungert
Zentralnervöse Verarbeitung akustischer Informationen
Signalidentifikation, Signallateralisation und zeitgebundene Informationsverarbeitung bei Patienten mit erworbenen Hirnschädigungen

42 Daniel Senkowski
Neuronal correlates of selective attention: An investigation of electro-physiological brain responses in the EEG and MEG

43 Gert Wollny
Analysis of Changes in Temporal Series of Medical Images

44 Angelika Wolf
Sprachverstehen mit Cochlea-Implantat: EKP-Studien mit postlingual ertaubten erwachsenen CI-Trägern

45 Kirsten G. Volz
Brain correlates of uncertain decisions: Types and degrees of uncertainty

46 Hagen Huttner
Magnetresonanztomographische Untersuchungen über die anatomische Variabilität des Frontallappens des menschlichen Großhirns

47 Dirk Köster
Morphology and Spoken Word Comprehension: Electrophysiological Investigations of Internal Compound Structure

48 Claudia A. Hruska
Einflüsse kontextueller und prosodischer Informationen in der auditorischen Satzverarbeitung: Untersuchungen mit ereigniskorrelierten Hirnpotentialen

49 Hannes Ruge
Eine Analyse des raum-zeitlichen Musters neuronaler Aktivierung im Aufgabenwechselparadigma zur Untersuchung handlungssteuernder Prozesse

50 Ricarda I. Schubotz
Human premotor cortex: Beyond motor performance

51 Clemens von Zerssen
Bewusstes Erinnern und falsches Wiedererkennen: Eine funktionelle MRT Studie neuroanatomischer Gedächtniskorrelate

52 Christiane Weber
Rhythm is gonna get you.
Electrophysiological markers of rhythmic processing in infants with and without risk for Specific Language Impairment (SLI)

53 Marc Schönwiesner
Functional Mapping of Basic Acoustic Parameters in the Human Central Auditory System

54 Katja Fiehler
Temporospatial characteristics of error correction

55 Britta Stolterfoht
Processing Word Order Variations and Ellipses: The Interplay of Syntax and Information Structure during Sentence Comprehension

56 Claudia Danielmeier
Neuronale Grundlagen der Interferenz zwischen Handlung und visueller Wahrnehmung

57 Margret Hund-Georgiadis
Die Organisation von Sprache und ihre Reorganisation bei ausgewählten, neurologischen Erkrankungen gemessen mit funktioneller Magnetresonanztomographie – Einflüsse von Händigkeit, Läsion, Performanz und Perfusion

58 Jutta L. Mueller
Mechanisms of auditory sentence comprehension in first and second language: An electrophysiological miniature grammar study

59 Franziska Biedermann
Auditorische Diskriminationsleistungen nach unilateralen Läsionen im Di- und Telenzephalon

60 Shirley-Ann Rüschemeyer
The Processing of Lexical Semantic and Syntactic Information in Spoken Sentences: Neuroimaging and Behavioral Studies of Native and Non-Native Speakers

61 Kerstin Leuckefeld
The Development of Argument Processing Mechanisms in German.
An Electrophysiological Investigation with School-Aged Children and Adults

62 Axel Christian Kühn
Bestimmung der Lateralisierung von Sprachprozessen unter besondere Berücksichtigung des temporalen Cortex, gemessen mit fMRT

63 Ann Pannekamp
Prosodische Informationsverarbeitung bei normalsprachlichem und deviantem Satzmaterial: Untersuchungen mit ereigniskorrelierten Hirnpotentialen

64 Jan Derrfuß
Functional specialization in the lateral frontal cortex: The role of the inferior frontal junction in cognitive control

65 Andrea Mona Philipp
The cognitive representation of tasks – Exploring the role of response modalities using the task-switching paradigm

66 Ulrike Toepel
Contrastive Topic and Focus Information in Discourse – Prosodic Realisation and Electrophysiological Brain Correlates

67 Karsten Müller
Die Anwendung von Spektral- und Waveletanalyse zur Untersuchung der Dynamik von BOLD-Zeitreihen verschiedener Hirnareale

68 Sonja A.Kotz
The role of the basal ganglia in auditory language processing: Evidence from ERP lesion studies and functional neuroimaging

69 Sonja Rossi
The role of proficiency in syntactic second language processing: Evidence from event-related brain potentials in German and Italian

70 Birte U. Forstmann
Behavioral and neural correlates of endogenous control processes in task switching

71 Silke Paulmann
Electrophysiological Evidence on the Processing of Emotional Prosody: Insights from Healthy and Patient Populations

72 Matthias L. Schroeter
Enlightening the Brain – Optical Imaging in Cognitive Neuroscience

73 Julia Reinholz
Interhemispheric interaction in object- and word-related visual areas

74 Evelyn C. Ferstl
The Functional Neuroanatomy of Text Comprehension

75 Miriam Gade
Aufgabeninhibition als Mechanismus der Konfliktreduktion zwischen Aufgabenrepräsentationen

76 Juliane Hofmann
Phonological, Morphological, and Semantic Aspects of Grammatical Gender Processing in German

77 Petra Augurzky
Attaching Relative Clauses in German – The Role of Implicit and Explicit Prosody in Sentence Processing

78 Uta Wolfensteller
Habituelle und arbiträre sensomotorische Verknüpfungen im lateralen prämotorischen Kortex des Menschen

79 Päivi Sivonen
Event-related brain activation in speech perception: From sensory to cognitive processes

80 Yun Nan
Music phrase structure perception: the neural basis, the effects of acculturation and musical training

81 Katrin Schulze
Neural Correlates of Working Memory for Verbal and Tonal Stimuli in Nonmusicians and Musicians With and Without Absolute Pitch

82 Korinna Eckstein
Interaktion von Syntax und Prosodie beim Sprachverstehen: Untersuchungen anhand ereigniskorrelierter Hirnpotentiale

83 Florian Th. Siebörger
Funktionelle Neuroanatomie des Textverstehens: Kohärenzbildung bei Witzen und anderen ungewöhnlichen Texten

84 Diana Böttger
Aktivität im Gamma-Frequenzbereich des EEG: Einfluss demographischer Faktoren und kognitiver Korrelate

85 Jörg Bahlmann
Neural correlates of the processing of linear and hierarchical artificial grammar rules: Electrophysiological and neuroimaging studies

86 Jan Zwickel
Specific Interference Effects Between Temporally Overlapping Action and Perception

87 Markus Ullsperger
Functional Neuroanatomy of Performance Monitoring: fMRI, ERP, and Patient Studies

88 Susanne Dietrich
Vom Brüllen zum Wort – MRT-Studien zur kognitiven Verarbeitung emotionaler Vokalisationen

89 Maren Schmidt-Kassow
What's Beat got to do with ist? The Influence of Meter on Syntactic Processing: ERP Evidence from Healthy and Patient populations

90 Monika Lück
Die Verarbeitung morphologisch komplexer Wörter bei Kindern im Schulalter: Neurophysiologische Korrelate der Entwicklung

91 Diana P. Szameitat
Perzeption und akustische Eigenschaften von Emotionen in menschlichem Lachen

92 Beate Sabisch
Mechanisms of auditory sentence comprehension in children with specific language impairment and children with developmental dyslexia: A neurophysiological investigation

93 Regine Oberecker
Grammatikverarbeitung im Kindesalter: EKP-Studien zum auditorischen Satzverstehen

94 Şükrü Barış Demiral
Incremental Argument Interpretation in Turkish Sentence Comprehension

95 Henning Holle
The Comprehension of Co-Speech Iconic Gestures: Behavioral, Electrophysiological and Neuroimaging Studies

96 Marcel Braß
Das inferior frontale Kreuzungsareal und seine Rolle bei der kognitiven Kontrolle unseres Verhaltens

97 Anna S. Hasting
Syntax in a blink: Early and automatic processing of syntactic rules as revealed by event-related brain potentials

98 Sebastian Jentschke
Neural Correlates of Processing Syntax in Music and Language – Influences of Development, Musical Training and Language Impairment

99 Amelie Mahlstedt
The Acquisition of Case marking Information as a Cue to Argument Interpretation in German An Electrophysiological Investigation with Pre-school Children

100 Nikolaus Steinbeis
Investigating the meaning of music using EEG and fMRI

101 Tilmann A. Klein
Learning from errors: Genetic evidence for a central role of dopamine in human performance monitoring

102 Franziska Maria Korb
Die funktionelle Spezialisierung des lateralen präfrontalen Cortex: Untersuchungen mittels funktioneller Magnetresonanztomographie

103 Sonja Fleischhauer
Neuronale Verarbeitung emotionaler Prosodie und Syntax: die Rolle des verbalen Arbeitsgedächtnisses

104 Friederike Sophie Haupt
The component mapping problem: An investigation of grammatical function reanalysis in differing experimental contexts using event-related brain potentials

105 Jens Brauer
Functional development and structural maturation in the brain's neural network underlying language comprehension

106 Philipp Kanske
Exploring executive attention in emotion: ERP and fMRI evidence

107 Julia Grieser Painter
Music, meaning, and a semantic space for musical sounds

108 Daniela Sammler
The Neuroanatomical Overlap of Syntax Processing in Music and Language - Evidence from Lesion and Intracranial ERP Studies

109 Norbert Zmyj
Selective Imitation in One-Year-Olds: How a Model's Characteristics Influence Imitation

110 Thomas Fritz
Emotion investigated with music of variable valence – neurophysiology and cultural influence

111 Stefanie Regel
The comprehension of figurative language: Electrophysiological evidence on the processing of irony

112 Miriam Beisert
Transformation Rules in Tool Use

113 Veronika Krieghoff
Neural correlates of Intentional Actions

114 Andreja Bubić
Violation of expectations in sequence processing

115 Claudia Männel
Prosodic processing during language acquisition: Electrophysiological studies on intonational phrase processing

116 Konstanze Albrecht
Brain correlates of cognitive processes underlying intertemporal choice for self and other

117 Katrin Sakreida
Nicht-motorische Funktionen des prämotorischen Kortex: Patientenstudien und funktionelle Bildgebung

118 Susann Wolff
The interplay of free word order and pro-drop in incremental sentence processing: Neurophysiological evidence from Japanese

119 Tim Raettig
The Cortical Infrastructure of Language Processing: Evidence from Functional and Anatomical Neuroimaging

120 Maria Golde
Premotor cortex contributions to abstract and action-related relational processing

121 Daniel S. Margulies
Resting-State Functional Connectivity fMRI: A new approach for assessing functional neuroanatomy in humans with applications to neuroanatomical, developmental and clinical questions

122 Franziska Süß
The interplay between attention and syntactic processes in the adult and developing brain: ERP evidences

123 Stefan Bode
From stimuli to motor responses: Decoding rules and decision mechanismsin the human brain

124 Christiane Diefenbach
Interactions between sentence comprehension and concurrent action: The role of movement effects and timing

125 Moritz M. Daum
Mechanismen der frühkindlichen Entwicklung des Handlungsverständnisses

126 Jürgen Dukart
Contribution of FDG-PET and MRI to improve Understanding, Detection and Differentiation of Dementia

127 Kamal Kumar Choudhary
Incremental Argument Interpretation in a Split Ergative Language: Neurophysiological Evidence from Hindi

128 Peggy Sparenberg
Filling the Gap: Temporal and Motor Aspects of the Mental Simulation of Occluded Actions

129 Luming Wang
The Influence of Animacy and Context on Word Order Processing: Neurophysiological Evidence from Mandarin Chinese

130 Barbara Ettrich
Beeinträchtigung frontomedianer Funktionen bei Schädel-Hirn-Trauma

131 Sandra Dietrich
Coordination of Unimanual Continuous Movements with External Events

132 R. Muralikrishnan
An Electrophysiological Investigation Of Tamil Dative-Subject Constructions

133 Christian Obermeier
Exploring the significance of task, timing and background noise on gesture-speech integration

134 Björn Herrmann
Grammar and perception: Dissociation of early auditory processes in the brain

135 Eugenia Solano-Castiella
In vivo anatomical segmentation of the human amygdala and parcellation of emotional processing

136 Marco Taubert
Plastizität im sensomotorischen System – Lerninduzierte Veränderungen in der Struktur und Funktion des menschlichen Gehirns

137 Patricia Garrido Vásquez
Emotion Processing in Parkinson's Disease: The Role of Motor Symptom Asymmetry

138 Michael Schwartze
Adaptation to temporal structure

139 Christine S. Schipke
Processing Mechanisms of Argument Structure and Case-marking in Child Development: Neural Correlates and Behavioral Evidence