Analyzing the Perception of Natural Music with EEG and ECoG

vorgelegt von
Dipl.-Inf. Irene Sturm
geb. in München

von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr. Klaus Obermayer
1. Gutachter: Prof. Dr. Benjamin Blankertz
2. Gutachter: Prof. Dr. Klaus-Robert Müller
3. Gutachter: Prof. Dr. Gabriel Curio

Tag der wissenschaftlichen Aussprache: 9. Dezember 2015

Berlin 2016

D83

Abstract

Brain states during real-word experiences have attracted growing research interest in the past two decades. Listening to music is one example of an on-going real-world experience that, on the one hand, relies on structured auditory input, on the other hand, often involves strong emotional responses. Since, obviously, the brain is the mediator between sound wave and subjective experience, it is an interesting question whether the comparative analysis of brain signals and music signals is a way to understand this fascinating process. Electrophysiological recordings of the brain are particularly suited for this, as they offer a temporal resolution in the millisecond range, a time scale that potentially gives access to the processing of the fine details of the rapidly changing musical surface. Deriving, however, interpretable information from electrophysiological signals recorded during music listening is a challenging task. Extracting stimulus-related brain activity from the electroencephalogram typically requires averaging of a high number of stimulus repetitions. If full-length pieces of music are presented, and, moreover, the unique listening experience is of interest, more sensitive methods for extracting neural activity from the continuous brain signal are required.

This thesis makes several contributions toward the development of such methods, by addressing relevant issues that arise whenever brain signals are analyzed in combination with music signals.

Taking advantage of the compelling properties of invasive ECoG recordings, the first part of this thesis presents a simple, but efficient method to derive a detailed reflection of an original rock song in the brain signal.

A core contribution of this thesis aims at further promoting the more widely applicable recording modality of (scalp) EEG for investigating the relationship between music signal and brain signal. Utilizing an evoked brain response to low-level constituents of music, i.e., to note onsets, we propose a multi-variate regression-based method for mapping the continuous EEG signal back onto the music signal. This so-called stimulus reconstruction approach is highly suitable for EEG recordings of the presentation of full-length, acoustically very complex music pieces. The resulting stimulus reconstructions can be used to determine the level of Cortico-Acoustic Correlation (CACor): the degree of synchronization of the brain signal with the stimulus. Through CACor, this thesis explores the potential of the stimulus reconstruction approach in several music-related research scenarios. A simple, repetitive sound stimulus is used to demonstrate the connection from the extracted brain signatures to canonical ERP components. Subsequently, the method is applied in a more complex setting that relates to auditory stream segregation. We first demonstrate that three monophonic semi-musical stimuli can be reconstructed from the listener's EEG. Next, we show to what extent such learned mappings can be utilized to trace a neural correlate of the separate streams that the listener perceives when the three voices play together forming a polyphonic semi-musical sound pattern.

Finally, we progress to the most 'natural' experimental setting in this context: the analysis of EEG recordings during listening to 'real' music without a specific task. We examine CACor for a range of naturalistic music and non-music sounds. We show that differences between stimuli with respect to observed significant CACor can be related to their acoustic/higher-level musical properties. Finally, with a complementary analysis of behavioral reports of perceived tension we provide first evidence on the experiential relevance of the observed relationship between brain signal and music signal.

Zusammenfassung

Das Interesse an der Frage, wie unser Gehirn die komplexen Reize, die in unserer Umwelt auf uns einströmen, verarbeitet, ist in den letzten zwei Dekaden stetig gewachsen.
Musikhören, ein Teil unseres Alltags, beginnt mit solch einem komplex strukturierten auditorischen Reiz und führt zu einem subjektiven, oft emotionalen Empfinden. Die vermittelnde Rolle des Gehirns in diesem Prozess ist offensichtlich. Unklar jedoch ist, inwiefern der Vergleich zwischen Gehirnsignalen und Musiksignalen Aufschluss geben kann über den Weg von der Schallwelle zum persönlichen Hörerlebnis. Elektrophysiologische Messverfahren wie Elektroenzephalogramm (EEG) und Elektrokortikogramm (ECoG) sind ideal für eine detaillierte Untersuchung von Musikverarbeitung im Gehirn, da ihre Zeitauflösung im Millisekundenbereich liegt. Aus den Gehirnströmen eines musikhörenden Menschen interpretierbare Informationen zu ziehen, ist jedoch eine datenanalytische Herausforderung. Typischerweise werden reiz-spezifische Muster im EEG sichtbar gemacht, indem gemittelte Zeitkurven vieler Wiederholungen eines Stimulus betrachtet werden. Wenn lange, komplexe Musikstücke präsentiert werden sollen, und, darüber hinaus, das individuelle Hörerlebnis von Interesse ist, sind alternative Methoden nötig, um Informationen aus den kontinuierlichen Gehirnsignalen zu extrahieren.

Die vorliegende Dissertation liefert eine Reihe von Beiträgen, die typische Probleme der kombinierten Analyse von Musik- und Gehirnsignalen thematisieren und Lösungsansätze vorschlagen. Die erste in dieser Arbeit beschriebene Studie nutzt die besondere Datenqualität des invasiven ECoG und stellt eine einfache, aber effektive Methode vor, um ein hochdifferenziertes Abbild der mehrdimensionaler Struktur eines Rockmusikstücks in den Gehirndaten sichtbar zu machen.

Ein weiterer Beitrag dieser Arbeit zielt darauf ab, durch Fortschritte in der Datenanalyse das wesentlich universaler anwendbare EEG für die vergleichende Analyse von Gehirnsignal und Musiksignalen auszunutzen. Ausgehend von sogenannten 'obligatorischen' evozierten Potentialen auf einzelne Töne, wird eine multi-variate Analysemethode vorgestellt, die die neurale Representation der Abfolge von Tönen eines Musikstücks extrahiert. Diese Methode zur 'Stimulusrekonstruktion' eignet sich für die Anwendung auf kontinuierlichen EEG-Daten und stellt somit eine Alternative zu konventionellen Methoden dar. Mittels dieser neuralen Repräsentation kann der Cortico-Acoustic Correlation Coefficient (CACor) bestimmt werden, der als Maß für die Synchronisation von EEG-Signal und Musiksignal dient. Konkrete Anwendungsbeispiele, die vorgestellt und diskutiert werden, sind (i) ein Vergleich der mit der vorgeschlagenen Methode extrahierten kortikalen Signaturen mit konventionell ermittelten ereigniskorrelierten Potentialen, der in Falle eines einfachen repetitiven Stimulus möglich ist. Des weiteren wird (ii) die Methode auf Daten aus einer komplexeren Hörsituation evaluiert, nämlich anhand einer Musik-Variante des 'Cocktail-Party-Problems' mit dem das Phänomen der auditory stream separation (ASS) erforscht werden kann. Die letzte Studie (iii) evaluiert die vorgestellte Methode in einer annähernd alltagsähnlichen Hörsituation: Versuchspersonen hören, unbeschwert von einer Aufgabe, eine Auswahl von Musikstücken und anderen naturalistischen Geräuschen während ihr EEG aufgezeichnet wird. Die Analysen zeigen, dass unterschiedliche Ausprägungen von CACor während verschiedener Hörbeispiele durch akustische und musikspezifische Merkmale erklärt werden können. Eine komplemetäre Untersuchung behavioraler Messungen erlebter Spannung in der Musik gibt erste Hinweise darauf, dass die durch CACor evidente Synchronisierung des Gehirnsignals mit dem Musiksignal in Verbindung steht mit dem subjektiven Musikerleben.

Acknowledgments

I am grateful for the support of many people during the past four years. Above all, I thank my supervisors Prof. Dr. Benjamin Blankertz, Prof. Dr. Gabriel Curio and Prof. Dr. Klaus-Robert Müller who gave me the freedom and trust to pursue my scientific ideas.

My special thanks go to Prof. Dr. Gabriel Curio who helped to launch my doctoral project, who continuously supported my work with his advice and experience, who helped me to focus my thoughts in countless stimulating discussions, and, who genuinely shares my joy of investigating the workings of the music-listening brain. I thank Prof. Dr. Benjamin Blankertz for enabling me to conduct my research in his group while giving me lots of freedom and being open-minded and supportive in many ways. He was available when I had questions, and with patience and attentiveness helped me to enhance my understanding of EEG analysis. With his creativity he showed me an inspiring and open approach to scientific research and often pointed me towards interesting new research directions. I am very grateful to Prof. Dr. Klaus-Robert Müller for his long-standing support and encouragement that has accompanied my scientific life since the very early days as a HiWi at Fraunhofer First. His scientific enthusiasm is infectious and has actually kindled my first tentative interest in Machine Learning.

I thank Gerwin Schalk for giving me the opportunity to expand my scientific horizon towards ECoG data. Not only the access to a very special data set extended my scientific experience, but also the fruitful and effective way of collaborating with him, with Cristhian Potes, and other members of the Schalk lab in Albany.

Additionally, I would like to thank all past and present lab members of IDA/BBCI for generating such an inspiringly interdisciplinary and friendly atmosphere. In particular, I would like to mention Matthias Treder, Daniel Miklody and Hendrik Purwins who shared their mission to make BCI more musical with me. I'd like to thank Sven Dähne for his patience when explaining mathematical or statistical issues. I'd like to thank Felix Bießmann and Stefan Haufe for sharing their experience in EEG analysis. I'd like to thank Dominik Kühne, Imke Weitkamp and Andrea Gerdes and all the others that I may have forgotten.

My other scientific 'home' has been in the Neurophysic Group at the Campus Benjamin Franklin, a place of quiet and focused working. There, I could enjoy the exchange with colleagues like Vadim Nikulin, Tommaso Fedele, Gunnar Waterstraat, Friederike Hohlefeld and Katherina v. Carlowitz-Ghori and many others. Thanks to all of them!

My special thanks goes to Manon Grube for her careful proof-reading that helped to increase the quality of this manuscript and for taking great interest in my work.

I thank the Berlin School of Mind and Brain which provided full funding of this work. Being part of this interdisciplinary graduate school was a great experience. My research benefited greatly from the opportunity to expand my neuroscientific background in courses and lectures. In particular, M&B supported me in an optimal way in balancing family life and doctoral studies. I thank the Christiane-Nüsslein-Volhard Foundation for additional financial and non-material support.

Finally, I owe my deepest gratitude to my family, first and foremost my husband Gunter, for supporting and encouraging me in every possible way. I thank my children for being curious about my work, for asking many questions, for providing unbiased and direct judgments of the music I chose for my experiments, and, in particular, for being so effective in (temporarily) distracting me from pressing scientific problems. I thank my mother for being always ready to offer her support, in particular her childcare services.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Investigating music processing

The opportunities modern neuroscience offers for investigating the perception of music have been seized extensively in the last decades (Peretz and Zatorre, 2005). The cascade of brain processes that leads from a stream of real-world acoustic data to a personal, often emotional, listening experience can be considered as the most generic research problem in this context. Investigating how the brain processes the multiple aspects of a musical piece is bound to enhance our understanding of this fascinating process.

Brain responses to essential structural components of music have been investigated in an enormous number of studies addressing the processing of, e.g., pitch (Hyde et al., 2008, Kumar et al., 2011, Nan and Friederici, 2013, Plack et al., 2014), sensory dissonance (Daikoku et al., 2012, Perani et al., 2010, Regnault et al., 2001), timbre (Caclin et al., 2006, 2007, Deike et al., 2004, Goydke et al., 2004), melodic contour (Trainor et al., 2002), key (Janata et al., 2002), mode (Halpern et al., 2008), scale properties (Brattico et al., 2006), music-syntactic congruity (Jentschke et al., 2014, Kim et al., 2014, Koelsch and Mulder, 2002, Sammler et al., 2013, 2011) and rhythmic aspects (Abrams et al., 2011, Grahn and Rowe, 2009, Jongsma et al., 2004, Schaefer et al., 2011b, Snyder and Large, 2005), to name only a few. This has led to a large corpus of evidence on associations between practically all aspects of a music stimulus and neurophysiological phenomena. Typically, related approaches rely on carefully selected or specifically designed stimulus material that allows to examine one aspect of music in an isolated manner while controlling for other influences. By design, they do not directly address the confluence of the multitude of musical features and their intrinsic relations and contextual effects in an actual piece of music. Therefore, there is still a considerable gap to bridge between today's knowledge about the processing of isolated (musical) sounds and music listening in everyday life.

A growing scientific effort has been targeted towards investigating music processing with more naturalistic stimulus material. Over the last years only have advances in data analysis turned the study of the processing of so-called 'natural' music into a tangible topic. 'Natural' music in the present thesis denotes music that has been created as an artwork and not for experimental purposes and that is presented without manipulation and in excerpts long enough to represent the complex musical context.

There are a number of reasons for advancing from simple controlled stimuli to more naturalistic listening scenarios. In a general sense, this endeavor aims at deriving an ecologically valid picture of brain responses to music. More specifically, and viewing music perception as a special case of auditory scene analysis (ASA), it aims at encompassing general principles of auditory perception, such as segmentation, integration, and segregation (Bregman, 1994) that go beyond isolated acoustic features. Finally, it is only through ecologically valid setups that experiential and affective aspects of listening to music can possibly be accessed.

The benefits of the use of natural music have been employed in several ways: one line of research utilized music as a 'vehicle' to create different experimental conditions (e.g., mental states) and to subsequently examine how features of the brain signal differ between these conditions. Early studies delineated generic contrasts between listening to music and resting state (EEG: Bhattacharya et al. (2001a), Brown et al. (2004), Schmidt and Trainor (2001), Positron emission tomography: Lin et al. (2014)) and between listening to music and listening to non-music complex sounds (Abrams et al., 2013, Kaneshiro et al., 2008, Menon and Levitin, 2005). More specific approaches utilized natural music to study musical memories (Janata, 2009), musical preferences (Wilkins et al., 2012), familiarity (Pereira et al., 2011) and expressive performance (Chapin et al., 2010). The intensively researched field of music and emotions in particular, has used naturalistic music to distinguish happy versus sad states Brattico et al. (2011), Khalfa et al. (2005), Mitterschiffthaler et al. (2007), to characterize music-induced 'chills' (Blood and Zatorre, 2001) and emotional peak events (Salimpoor et al., 2011). Brain measures that revealed differences between conditions have been related to activation patterns of the BOLD response in fMRI, to bandpower features in EEG, and to brain connectivity measures as well as inter-subject-correlation (ISC) in both modalities. A common feature of all these approaches is that the to-be-decoded mental states are assumed to be (relatively) static within one condition.

Over the last five years efforts have been intensified to directly track the dynamics of music stimuli in the brain signal. In general, this aims at examining how brain dynamics underlying perceptual and cognitive processes unfold over time in naturalistic conditions and has also been pursued intensively in the visual (Einhäuser and König, 2010, Hamamé et al., 2014, Hasson, 2004, Hasson and Honey, 2012, Hasson et al., 2010) and audio-visual domain (Dmochowski et al., 2012, Gaebler et al., 2014, Whittingstall et al., 2010). In the music domain, several approaches that combined neuroimaging and acoustic feature extraction directly investigated the relationship between brain signals and the multi-dimensional structure of full-length pieces of natural natural music (Abrams et al., 2013). Further fMRI studies examined how continuously measured emotional responses and neural activity evolved with changes in stimulus parameters. These studies were able to link reported emotional responses to music to activity in key brain areas of affective processing (Lehne et al., 2013b) and motor brain areas (Chapin et al., 2010). However, very few studies used the electroencephalogram (EEG), which, due to its high temporal resolution, is most suitable for investigating the dynamics of music at the appropriate fine-grained time scale. Apart from the work by Mikutta et al. (2012, 2013) which related EEG bandpower fluctuations to behavioral arousal measures, there has been one novel approach to analyze continuous EEG elicited by natural music stimuli proposed by Cong et al. (2012) and applied in Thompson (2013). This approach allows the identification of EEG components that are common within a group of subjects and, subsequently, relates their time course to music features. One electrocorticographic study (Potes et al., 2012) demonstrated that the sound envelope of a piece of natural music leaves a marked

reflection in the high gamma power modulations of the ECoG. Furthermore, numerous studies used naturalistic but short excerpts (Schaefer et al., 2011a, 2009) to examine brain responses to several aspects of music.

In summary, finding and interpreting the reflection of natural music in the listener's brain signals is an ongoing, timely research endeavor. So far, there neither is a gold standard solution for analyzing data nor a clear aim of what can be learned from the results. Very few accounts of how naturalistic music may be presented in the brain signal exist. Furthermore, they cover only a small sample of musical pieces and are highly diverse in terms of recording modality, level of specificity and musical aspect of interest. Particularly scarce are studies that use EEG. In principle, electrophysiological signals are attractive for investigating music processing, as they present the opportunity to operate at a far more fine-grained time scale than, e.g., fMRI. However, there exist data analytical challenges that impede the use of EEG in combination with continuous complex stimuli. Finally, even though EEG signals seem to share some characteristics with music waveforms, it is not clear what can be learned from a reflection of music in an EEG signal. In particular, it is not known whether and how the EEG can inform about experiential aspects of listening to music.

1.2 Research questions and outline

The purpose of this thesis is to develop methodological approaches for extracting and interpreting music-related neural activity from electrophysiological signals recorded from the listening brain, and to combine these with behavioral measurements and with information extracted from the music signal. This is accomplished in three steps:

(i) We probe whether the technique of sound envelope-tracking in ECoG high gamma power can be extended towards integrating a multidimensional description of a natural music stimulus.

(ii) We introduce a multi-variate method of (scalp) EEG analysis to optimally extract stimulus-related neural activity with an enhanced signal-to-noise ratio (compared to univariate methods).

(iii) We apply the proposed method in a number of stimulation scenarios in order to demonstrate the potential of EEG recordings and appropriate analyses in the exploration of natural music processing.

Chapter 2 deals with the background and basic methods that are necessary for conducting the single steps of this multi-facetted endeavour. The first section of Chapter 2 focuses on the motivation for investigating music perception with naturalistic sounds. Next within the same chapter, the basic facts about the acquisition of the different types data to be analyzed and their characteristic properties are reviewed. Subsequently, the mathematical concepts from machine learning and statistics that are of general use are introduced. The last part of the chapter gives an overview about audio analysis of musical signals.

Chapters 3 and 4 contain the core methodological contributions of this thesis. Chapter 3 presents a re-analysis of an ECoG study that aims at finding a detailed reflection of the multi-dimensional structure of an original rock song in ECoG high gamma power.

Chapter 4 introduces an analysis pipeline for applying Linear Ridge Regression to extract brain responses to the temporal structure of music from the ongoing EEG. From the Regression models brain signatures are derived that can be interpreted neurophysiologically. We present a number of applications of this method in Chapter 5. After introducing the experimental setups and data sets in Section 5.1, in Section 5.2.1 we directly compare the brain signatures derived by use of the the proposed method to those obtained through conventionally averaging of ERPs. In Section 5.2.2 we apply our method in the context of a problem related to Auditory Stream Segregation (ASS). In Section 5.2.3 we advance towards examining EEG recordings, behavioral measures and music features in a setting where participants listen to 'real' music (full-length naturalistic pieces of music) in a 'free listening paradigm' (without a specific task). Finally, this thesis concludes with a general discussion in Chapter 6.

1.3 List of included published work

The following publications are included in this thesis.

(1) Sturm, I., Blankertz, B., Potes, C., Schalk, G., and Curio, G. ECoG high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song. *Frontiers in Human Neuroscience*, 8(798), 2014.

(2) Sturm, I., Treder, M., Miklody, D., Purwins H., Dähne, S., Blankertz, B., and Curio, G. Extracting the neural representation of tone onsets for separate voices of ensemble music using multivariate EEG analysis, in print. *Psychomusicology: Music, Mind and Brain*, 2015.

(3) Sturm, I., Dähne, S., Blankertz, B., and Curio, G. Multi-variate EEG analysis as a novel tool to examine brain responses to naturalistic music stimuli, under review. *PlosOne*, 2015.

1.4 List of all published work

(4) Blankertz, B.,Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., Maeder, C., Ramsey, L., Sturm, I., Curio, G., and Müller, K. R. The Berlin brain–computer interface: non-medical uses of BCI technology *Frontiers in Neuroscience*, 4(198), 2010.

(5) Treder, M., Purwins, H., Miklody, D., Sturm, I. and Blankertz, B. Decoding auditory attention to instruments in polyphonic music using single-trial EEG classification. *Journal of Neural Engineering*, 11(2), 026009, 2014.

1.5 Peer-reviewed conference contributions

(6) Sturm, I., Curio, G. , Blankertz, B. Single-Trial ERP Analysis Reveals Unconsciously Perceived Structures of Music In: *TOBI workshop, Graz, 2010.*

(7) Sturm, I., Blankertz, B., Curio, G. Neural repercussions of minimal music: EEG indices for conscious and 'unconscious' decisions on chord consonance In: *the Decision Neuroscience Workshop Berlin, 2011.*

(8) Sturm, I., Curio, G. , Blankertz, B. Single-trial EEG analysis reveals non-conscious processing of different dimensions of chord consonance In: *Segregation and Integration in Music and Language, Tübingen, 2012.*

(9) Treder, M., Purwins, H., Miklody, D., Sturm, I. and Blankertz, B. The musical BCI: Control by selective listening, In: *Proceedings of the Asilomar International BCI Meeting, 2013.*

(10) Sturm, I., Blankertz, B., Curio, G. Investigating the processing of original music with EEG: Brain responses to temporal musical structure In: *Summer School Music - Emotion - Learning, Rheinsberg, 2013.*

(11) Sturm, I., Blankertz, B., Curio, G. The brain's onset detector: Investigating the temporal structure of music with EEG. In: *Neuromusic, Hamilton, 2013.*

(12) Sturm, I., Blankertz, B., Potes, C., Schalk, G., and Curio, G. ECoG high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song In: *International Conference for Clinical Neurophysiology, 2014.*

(13) Sturm, I., Treder, M., Miklody, D., Purwins H., Dähne, S., Blankertz, B., and Curio, G. The polyphonic brain: extracting a neural representation of the rhythmic structure of separate voices of polyphonic music using ERPs. In: *The Neurosciences and Music V, Dijon, 2014.*

(14) Sturm, I., Treder, M., Miklody, D., Purwins H., Dähne, S., Blankertz, B., and Curio, G. Extracting the neural representation of tone onsets for separate voices of polyphonic music. In: *Cognitive Hearing for Communication, 2015.*

Chapter 2

Fundamentals

Investigating relations between music, brain data and behavioral measurements is a multidisciplinary task that integrates methods from different fields. In the following pages we briefly introduce the basic facts and techniques relevant for the steps of the typical cycle of experiment design, data acquisition and data analysis in our experimental scenario.

2.1 Experiment design: Towards ecological valid experimental paradigms

The key argument for progressing from simple stimuli and tightly controlled listening paradigms towards naturalistic listening situations lies in the increase of ecological validity which is the extent to which study results can be generalized to real-life settings. Ecological validity seems to have a value 'per se' in the sense that if perceptional processes or brain states are investigated this should resemble real-life perceptional processes as closely as possible (Abrams et al., 2013).

The most obvious way to approach this goal is to present stimulus material in the form of musical sounds as they are commonly heard in everyday life. Beyond that, the presence and nature of a task (e.g., passive vs. active listening paradigm) have an impact on the level of ecological validity of. A third influencing factor is the listening situation that is determined by the experimental setup (laboratory vs. everyday life setting) and the presentation mode (e.g., recorded sounds versus live sounds, spatially distributed or single-source sound). Currently, practically all experiments that involve brain recordings take place in a lab setting. However, on-going and future advances in wearable EEG technology will facilitate out-of-the-lab experiments, including those related to music processing in the future. Considerable efforts have in fact been made to equip a concert hall environment with EEG recording facilities for performers and audience (http://livelab.mcmaster.ca), manifesting the relevance of situational factors to the music perception research community.

The pursuit of ecologically valid listening scenarios has been driven by evidence suggesting that brain responses to naturalistic stimulation may differ from those related to controlled stimulation with simplified stimuli. In the visual domain this was demonstrated by Hasson et al. (2010) who found a greater response reliability of electrophysiological

brain activity for aspects of naturalistic visual stimuli/movies than for comparable artificial stimuli. Similar findings were obtained by Mechler et al. (1998) and Yao et al. (2007). In the music domain Abrams et al. (2013) provided direct evidence that the between-subject synchronization of a large-scale distributed brain network was significantly higher when listeners were presented with complex musical stimuli lasting minutes than when they listened to shorter pseudo-musical contexts.

Furthermore, there are also phenomena that cannot be studied without complex naturalistic stimuli. These relate to user variables, such as affective responses to music (e.g., arousal, valence, or aesthetic experience), situational aspects (immersion, engagement or listening effort) and general aspects of individual listener experience (familiarity, liking, memories and associations) (Brattico et al., 2013, Janata, 2009, Wilkins et al., 2012). Naturalistic stimuli of sufficient duration are critical to the examination of the processing of complex structural/perceptual or aspects of music (Large, 2001). Investigating the listener's ability to integrate different types of music-related information over time periods on the order of minutes into one 'coherent perceptual gestalt' (Leaver et al., 2009) obviously creates such a demand (Abrams et al., 2013). Furthermore, investigating socioculturally defined properties of musical excerpts such as stylistic or genre-related aspects is not feasible without naturalistic music stimuli, neither are complex formal properties of music involving macro-structure, complex patterns or motifs.

Finally, auditory 'decoding' approaches that primarily seek to find out how complex auditory stimuli are represented in the brain often depend on naturalistic stimuli. Advances in this field and in its techniques are the prerequisite for applications in the clinical context, e.g., for diagnostic use, in the development of assistive listening technology or brain computer interfaces (BCI).

2.2 Data acquisition

Brain processes can be examined using a number of different methodologies. Of these, electrophysiological recordings are regarded as one of the most direct and immediate correlates of neural activity (Turner, 2009). With a temporal resolution of milliseconds electrophysiological measures allow to examine music perception at high level of detail. Considering that the time scale of conscious music perception reaches a resolution of approximately 100 ms ('rhythmic ceiling' London (2012)) electrophysiological recordings can be regarded as the ideal modality for a examining the process of listening to music at a high level of detail. In contrast, recording modalities that rely on metabolic processes, such as functional magnetic resonance imaging (fMRI), have a time-resolution on the order of seconds.

A second major advantage of electrophysiological recordings is that they are obtained in silently. In contrast, in particular the fMRI brings along a scanner noise that may severely interfere with music listening (Skouras et al., 2013). Electroencephalogram (EEG) and Electrocorticography (ECoG), the two types of electrophysiological measurements that are relevant for this thesis are briefly reviewed.

The Magnetoencephalogram (MEG), which is a further electrophysiological recording modality that is highly relevant in research on auditory perception, is not discussed here.

2.2.1 EEG

General introduction to EEG The electroencephalographic signal (EEG) is an electric potential that is recorded non-invasively from the scalp. Since measured signals represent synchronous activity of large ensembles of neurons, it offers a macroscopic view on the brain. EEG signals, as mentioned before, reflect neural activity and can be recorded with high temporal resolution. However, due to the low conductivity of skull and tissue between the neural generators and sensors their spatial specificity is relatively low. Moreover, the brain signal is dampened, resulting in a relatively low strength of the brain signal of interest relative to the level of background noise, such as sensor noise.

The two most widely studied neurophysiological phenomena observed in EEG signals are event-related potentials (ERPs) and oscillations. ERPs represent synchronized activity of neurons that is time-locked to distinct external or internal events. In the recorded EEG these transient events are presented by a series of deflections or 'components' that exhibit a characteristic topographical distribution and occur with a specific latency after the evoking event. Oscillations or 'brain rhythms' represent synchronized periodic activity of ensembles of neurons and are typically examined for specific frequency bands. The power of these oscillations can be taken as an index of spatial synchronization strength. Typical research scenarios studying neural oscillations would for instance examine changes in power related to internal or external events, e.g., those preceding motor action (Pfurtscheller and Lopes da Silva, 1999), stimulus-based modulation of oscillatory power (Dähne et al., 2014b), or the interaction between brain regions (see Jensen and Colgin (2007) for an overview).

To increase the signal-to-noise ratio, ERPs and oscillatory responses are usually analysed not based on single trials, but averaged over a sufficiently large number of repetitions of the same experimental condition.

EEG in music perception research ERPs have been linked to practically every aspect of music perception, starting from the basic parameters of music, such as pitch (Hyde et al., 2008, Kumar et al., 2011, Nan and Friederici, 2013, Plack et al., 2014), rhythmic aspects (Abrams et al., 2011, Grahn and Rowe, 2009, Jongsma et al., 2004, Schaefer et al., 2011b, Snyder and Large, 2005), intensity and accentuation (Abecasis et al., 2005, Peter, 2012), and sensory dissonance (Daikoku et al., 2012, Perani et al., 2010, Regnault et al., 2001) extending also to more complex aspects, such as, timbre (Caclin et al., 2006, 2007, Deike et al., 2004, Goydke et al., 2004), melodic contour (Fujioka et al., 2004, Trainor et al., 2002), harmony (Janata, 1995), key (Janata et al., 2002), mode (Halpern et al., 2008), scale properties (Brattico et al., 2006), music-syntactic congruity (Jentschke et al., 2014, Kim et al., 2014, Koelsch and Mulder, 2002, Sammler et al., 2013, 2011) and musical phrase structure (Nan et al., 2006). In addition to their usefulness in informing about the sensory and cognitive processing of specific features of music, ERPs have been operationalized to investigate the influence of affective states (Brattico et al., 2010), musical preferences (Istók et al., 2013), attention (Loui et al., 2005), and memory effects on the perception of music (Janata et al., 2002). They have provided insights into developmental and age-related aspects of music perception (Fujioka et al., 2006, 2004), and are widely used to study the effects of short- and long term learning (Baumann et al., 2008), (Shahin et al., 2003). Furthermore, an important field of application for ERP technique is the comparative study of music and language, for an overview see Patel (2007).

EEG-measured oscillations have been linked to rhythm perception, e.g., to derive neural correlates of beat perception (Nozaradan et al., 2011, 2012, 2013) and to identify brain networks that are specific for listening to music (Bhattacharya et al., 2001b, Wu et al., 2013). Bandpower modulations have been examined with respect to the relaxing/exciting effect of music (Höller et al., 2012) and to music-induced emotional arousal (Mikutta et al., 2012, 2013). Furthermore, oscillatory brain responses have been used in the study of the processing of musical-syntactic incongruence (Carrus et al., 2011) and musical complexity (Ruiz et al., 2009, Schaefer et al., 2011a).

2.2.2 ECoG

Electrocorticography (ECoG), or intracranial EEG (iEEG) is the practice of recording electric potentials directly from the surgically exposed cortical surface. Since the 1950s (Jasper and W., 1949) it has been developed as a standard procedure for localizing epileptogenic zones and mapping cortical functions in patients prior to surgical treatment for intractable epilepsy, brain tumors, or vascular malformations of the brain (Crone et al., 2009, Roebuck, 2012).

Like scalp EEG, ECoG recordings measure the synchronized activity of ensembles of cortical neurons with excellent temporal resolution. In contrast to scalp EEG, ECoG signals are only minimally dampened by tissue and not at all by the skull. This leads to higher signal amplitudes (i.e., 50-100μV maximum compared to 10-20μV Wilson et al. (2006)) and a considerably wider range of frequencies (i.e., 0-200 Hz versus 0-40 Hz Leuthardt et al. (2004)) that can be examined at an favorable signal-to-noise level. Moreover, the impact of volume conduction is lower and thus spatial specificity (i.e., at the order of millimeters versus centimeters Crone et al. (2009)) increased. ECoG, therefore, offers a 'mesoscopic' spatial resolution between the microscopic scale of single and multi-unit brain recordings and the macroscopic view on brain responses provided by EEG (Crone et al., 2009).

These substantial benefits have led to a growing interest in ECoG in research on cortical activation related to perceptual, cognitive and motor tasks. In particular, in the domain of Brain-Computer-Interfacing (BCI) ECoG is appreciated as a minimally-invasive alternative to scalp EEG. The fact that ECoG signals have been found to contain highly specific activation patterns for actual and imagined actions, has been utilized in a number of applications that aim at decoding BCI user's intent from their brain signals (Brunner et al., 2011, Leuthardt et al., 2011, 2006, 2004, Miller et al., 2010, Schalk et al., 2008, Spüler et al., 2014, van Vansteensel et al., 2010, Wilson et al., 2006). In the field of speech perception research, ECoG has emerged as a new technique to study the functional cortical organization of speech processing (Kubanek et al., 2013, Leonard and Chang, 2014, Martin et al., 2014, Pasley et al., 2012).

ECoG offers unprecedented opportunities in challenging analysis conditions, such as those created by naturalistic listening scenarios (Brunet et al., 2014, Derix et al., 2014, Kubanek et al., 2013, Majima et al., 2014, Pasley et al., 2012, Potes et al., 2012). Here, the superior signal-to-noise ratio of ECoG is advantageous for studying the processing of complex naturalistic stimuli at the level of single stimulus presentations and single subjects, and thereby enabling a view on the spatial and temporal characteristics of processing that is detailed enough to differentiate between specific aspects of stimuli.

In spite of these highly favorable characteristics of ECoG, several obvious constraints limit the applicability of ECoG for non-clinical research purposes. Firstly, ECoG is performed on patients whose physical and cognitive conditions are impaired in different ways and whose brains may not be representative of those of the healthy population with respect to function and neuroanatomy. Secondly, clinical are the sole determinant of all ECoG measurements and their settings, i.e., the selection of patients, placement of electrode grids, measurement protocols and medication. For the investigator who endeavors to acquire data as a byproduct of the clinical procedures this means that data that provide a specific field of view onto the brain may have to be aggregated over long periods of time. Infrequently occurring opportunities to collect data, in turn, prevent to application of highly specific exclusion/inclusion criteria for an experiment. This means that subject variables are typically less balanced than in EEG studies and sample sizes naturally much smaller. However, a number of studies have shown that consistent results can be obtained despite these difficulties (Kubanek et al., 2013, Potes et al., 2014, 2012).

2.2.3 Behavioral measures of experiencing music

The focus of this thesis is on the relationship between music signals and brain signals. In order to explore the behavioral relevance of any link between stimulus and brain signal, however, the concomitant assessment of experiential aspects of music at an appropriately detailed time scale is considered crucial here and represents another pillar of this work.

Affective responses that are triggered by music are 'the' driving force that makes us listen to music, go to music performances and spend money on buying music (Panksepp, 1995, Sloboda et al., 2001). The relationships between musical content and emotional responses have been well studied, initially by describing the global affective experience for entire excerpts, e.g., by assigning linguistic labels (Hevner, 1936) or by using scales to rate aspects of emotion, e.g. Lane et al. (1997) or Gabrielsson and Lindström (2001), see Juslin and Sloboda (2010) for an overview. However, both, affective processes and music unfold over time and, therefore, lend themselves exquisitely for a comparison of their respective dynamics - an idea that has been first introduced by Nielsen (1983) who recorded subjects' perceptions of ongoing 'musical tension' with specifically designed tension tongs. Since the 1990s continuous self-report measures of experiencing music have become an established method in music psychology Schubert (2010). 'Continuous' in this context refers to sequences of observations that are taken sequentially in time without interrupting the stimulus presentation and that describe the perceived magnitude of a dimension of listening experience at series of points over time. These continuous measurements avoid the issues of a 'retrospective story-telling effect' (Plotnikov et al., 2012), duration neglect (Fredrickson and Kahneman, 1993) and selective recall (Schubert, 2010) that have been found to affect global post-presentation assessments of music experience (e.g., Gabrielsson and Lindström E. (2010), Goebl et al. (2014)). However, since continuous ratings can only record one- or maximally two-dimensional data (e.g., with the Continuous Response Dial Interface Madsen and Fredrickson (1993) or the EMuJoy system Nagel et al. (2007)), this technique is not integrated easily with common multidimensional models of emotion, e.g., Wundt's three-dimensional model (Wundt, 1913) or Russell's model of valence and arousal (Russell, 1980) (for an overview see Scherer (2000)). Therefore, it is still a central question how to capture affective responses to music in an experimentally tractable manner.

One concept that has often been applied, in particular, to elucidate the relationship between musical structure and emotion, is that of musical tension. Musical tension is best described as a continuous build-up and subsiding of excitement that a listener experiences when listening to a piece of music. As the 'dynamic ingredient of emotion' (Lehne and Koelsch, 2015), the concept of tension has its roots in Meyer's work 'Emotion and meaning in music' (Meyer, 1961). It has been the topic of numerous empirical studies (Farbood, 2012, Farbood and Schoner, 2009, Farbood and Upham, 2013, Krumhansl, 1996, Madsen and Fredrickson, 1993, Nielsen, 1983), and also in the field of music theory (Lerdahl and Krumhansl, 2007).

The psychological mechanisms underlying this phenomenon have been examined recently in Lehne and Koelsch (2015) where musical tension is integrated in a more general framework of tension and suspense in a variety of contexts, including the reception of music, film, literature and visual arts. Lehne et al. delimit as main contributors to these processes states of instability or in-certainty that are accompanied by expectation or anticipation. Both, uncertainty and discrepance between actual state and expected state create feelings of rising tension, while the reduction of uncertainty typically induces a feeling of understanding and resolution. The interplay between fulfillment and violation of expectancies can thus be regarded as important determinants of the appeal of an artwork. An essential factor in Western tonal music is the ratio (and also the transition) between moments of dissonance and irregularity and moments of consonance and stability. Notably, both, stability and prediction are entertained at several levels of structural organization of music. Notes in a tonal context for instance are perceived as more or less stable (Krumhansl and Toiviainen, 2001), and chords can be more or less consonant/dissonant. Music-syntactic structures, such as cadences, create explicit expectations about their continuation. Furthermore, rhythmic and metric features create expectation about the timing of future events and in general, the occurrence of patterns at various time scales, such as rhythmic figures, motifs or formal macro-structure, creating and leading expectations in the listener at multiple levels (Rohrmeier and Koelsch, 2012). Thus, the ebb and flow of tension as an overall experience in music listening is associated with a multitude of aspects of music. Empirical results have further shown that reported tension can be related to basic acoustic features of music, such as loudness or sensory dissonance (Farbood, 2012, Pressnitzer et al., 2000), to higher-level 'top-down' tonal expectations (Lerdahl and Krumhansl, 2007), and to formal structure of music (Krumhansl, 1996).

Taken together, musical tension is a (one-dimensional) 'proxy' of more complex emotional experiences triggered by music. As an 'important intermediate step' between the recognition of musical structure and related affective responses (Farbood, 2012) musical tension relates to the interest of the present work in the relationship between brain signal and music signal. Musical tension seems to be influenced by a range of aspects of music and therefore, as a concept, can be applied to a large variety of musical sounds. Importantly, experienced tension has been explained in terms of domain-general psychological mechanisms that evoke enjoyable or rewarding feelings in the listener. Thus, tension presents itself as a promising concept for obtaining first insights into the experiential relevance of a potential link between audio and brain signals.

2.3 Data analysis

2.3.1 Notation

In the following we denote:

- matrices by italic upper-case letters
- vectors by bold lower-case letters
- scalars by italic lower-case or upper-case letters.

Vectors are understood to be in columnar shape. For a given vector $\mathbf{x}$ the i-th entry is denoted by x_i .

In a matrix A $\mathbf{a}_i$ denotes the i-th column vector and a_{ij} the entry in the i-th row and j-th column.

In a (multivariate) time series X the value at time point t is represented by the t-th column of X.

We denote with:

- T the number of data points in time.
- N the number of dimensions, e.g., EEG channels.
- k the number of components, e.g., neural sources that are extracted from the EEG.
- $X = [\mathbf{x}_1\mathbf{x}_2 \dots \mathbf{x}_T]$ a $N \times T$ matrix of observed data and $\mathbf{x}(t)$ an observation at a single time point t.
- $\mathbf{s}(t)$ a k-dimensional vector of component activity.
- $\hat{\mathbf{s}}(t)$ a k-dimensional vector of estimated component activity.
- $\eta(t)$ a measurement of noise at time point t.
- $A = [\mathbf{a}_1\mathbf{a}_2 \dots \mathbf{a}_k]$ a $N \times k$ matrix of patterns (in the columns) that define the mapping between neural sources and electrodes in the forward model.
- $W = [\mathbf{w}_1\mathbf{w}_2 \dots \mathbf{w}_k]$ a $N \times k$ matrix of filters (in the columns) that define the decomposition of the sensor level measurement into k neural sources in the backward model.
- $C_{XX} = XX^\top$ the (auto)covariance matrix of the multivariate time series X.
- $C_{XY} = XY^\top$ the (cross)covariance matrix of two multivariate time series X and Y.
- $z(t)$ a scalar target variable, e.g., a feature of the stimulus derived by audio signal analysis.

2.3.2 A generative model of EEG

Forward and backward model In the following we introduce a standard model of EEG generation that underlies numerous techniques of multivariate EEG analysis. The core idea of this model is that signals recorded from the scalp surface represent a linear mixture of the activity of neural sources. These neural sources have distinct and specific spatial and temporal characteristics and can be thought to represent functional brain units (Dähne, 2015, Parra et al., 2005). In mathematical terms this can be formulated as

$$\mathbf{x}(t) = A\mathbf{s}(t) + \eta(t). \tag{2.1}$$

where $\mathbf{x}(t)$ is the potential of the full set of electrodes measured at time point t, $\mathbf{s}(t)$ the activity of k of sources at time point t and $A = [\mathbf{a}_1\mathbf{a}_2 \dots \mathbf{a}_k]$ a matrix $\in \mathbb{R}^{N\times k}$ that represents the individual coupling strengths of each source $\mathbf{s}(t)$ to the array of N surface electrodes. A is called the *forward model*. The term $\eta(t)$ models the contribution activity that is not explained by the source components and, thus, is treated as 'noise' that is not relevant to the investigation. Since the original neural sources $\mathbf{s}(t)$ that led to the measurements $\mathbf{x}(t)$ cannot be observed directly, the task for the use of Machine Learning here is to derive an estimation $\hat{\mathbf{s}}(\mathbf{t})$ from the data. This process is called *backward modeling* and can be formulated as

$$\hat{\mathbf{s}}(t) = W^\top \mathbf{x}(t). \tag{2.2}$$

where the matrix $W = [\mathbf{w}_1\mathbf{w}_2 \dots \mathbf{w}_k]$ represents the extraction filters that transform sensor level observations into estimated source activity (for details see Blankertz et al. (2011)). The coefficients of W determine the contribution of each recorded sensor to the estimated source time course. However, in its generic form, the task of finding W represents a so-called inverse problem that does not have a unique solution (Pascual-Marqui, 1999). For practical use, the solution has to be constrained in a suitable way.

In general, approaches for deriving extraction filters can be 'supervised' or 'unsupervised'. Supervised methods integrate external information, e.g., label information or an additional target function, into the optimization process, while unsupervised methods are based on intrinsic properties of the data without any additional information. The analysis scenarios in this thesis are typical examples of the supervised type as they aim at detecting and defining relationships between EEG data and musical stimuli. The regression technique that is applied in this thesis in order to extract a filter matrix W while integrating a representation of the stimulus structure is explained in detail in Section 2.4.1.

Filters and patterns Although the weights of W determine how the sensor level time courses are integrated to form the time course of the estimated source $\hat{\mathbf{s}}(t)$, it is important to be aware of the fact that these weights do not yield information suitable for neurophysiological or hypothesis-related interpretation (Blankertz et al., 2011, Haufe et al., 2014). Such information can only be obtained from the patterns of the forward model. The reason for this is that filters of the backward model are determined so that the signal-to-noise ratio of a signal of interest is optimal. Therefore, they depend not

only on the distribution of neural sources, but also on the distribution of noise in the data. For a detailed explanation and illustration of the difference between pattern and filter see Blankertz et al. (2011). A forward model A can be obtained from every linear backward model W using the following equation (Haufe et al., 2014):

$$A = C_{XX}WC_{\hat{S}\hat{S}}^{-1} = C_{XX}W(W^\top C_{XX}W)^{-1} \tag{2.3}$$

where $C_{\hat{S}\hat{S}}$ is the covariance matrix of the estimated sources. Further processing steps, such as source localization techniques, need to be performed on forward model patterns, not on the backward model filters.

2.4 Machine Learning and Statistics

In the following we briefly introduce the most important analytical tools that are used within this thesis. This introduction focuses on the intuitive understanding and the specific application in the present context, rather than rigorousness or completeness. However, the methods presented here are documented extensively in the literature and relevant sources referred to in each subsection. The overarching goal of this thesis is to tackle the relationship between brain signal and music signal in contexts where typical strategies to enhance the EEG's signal-to-noise ratio, such as averaging across repetitions, are not applicable.

In summary the most important questions in this context are:

- How to extract stimulus-related neural activity from the EEG.
- How to extract neural activity from ECoG that is related the processing of specific aspects of music.
- How to deal with typical problems related to the structure of music stimuli, such as auto-correlation.

2.4.1 Regression

The generic regression task is that of finding a rule, which assigns an N-dimensional data vector $\mathbf{x}(t)$ to the value of a target variable $z(t)$. This is done by expressing the target variable $Z = [z(1)z(2)\dots z(T)]$ as a linear combination of the data matrix X. This problem can be formulated as

$$Z = \boldsymbol{w}^\top X + E. \tag{2.4}$$

where the vector $\boldsymbol{w}$ contains the weights linearly combining the values of X. A solution can be derived by way of the mean squared error (MSE) between the estimated target $\boldsymbol{w}^\top \boldsymbol{x}(t)$ and the true target variable $z(t)$. The MSE is given by

$$MSE(\boldsymbol{w}^\top \boldsymbol{x}(t), z(t)) = \frac{1}{T}\sum_{1}^{T}\frac{1}{2}(\boldsymbol{w}^\top \boldsymbol{x}(t) - z(t))^2. \tag{2.5}$$

The weight matrix W that minimizes the MSE is given by

$$\boldsymbol{w} = (XX^\top)^{-1}XZ^\top. \tag{2.6}$$

This solution is called Ordinary Least Squares (OLS) (for details of the derivation see Bishop (2006), p.140). Alternatively, an equivalent solution can be derived by approaching the problem in terms of covariance between $\boldsymbol{w}^\top\boldsymbol{x}(t)$ and $z(t)$ which leads to the following objective function

$$\max_w \mathrm{Cov}(\boldsymbol{w}^\top X, Z), \;\; \text{subject to} \qquad \mathrm{Var}\,(\boldsymbol{w}^\top X) = 1. \tag{2.7}$$

In matrix notation this can be expressed as

$$\max_w \boldsymbol{w}^\top XZ^\top, \;\; \text{subject to} \qquad \boldsymbol{w}^\top C_{XX}\boldsymbol{w} = 1. \tag{2.8}$$

Equation 2.8 can be cast into the following eigenvalue problem (for details of the derivation see Borga (1998)) that can be solved efficiently using standard numerical linear algebra tools, e.g., MATLAB or R.

$$\begin{bmatrix} 0 & C_{YX} \\ C_{XY} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{w_x} \\ 1 \end{bmatrix} = \Lambda \begin{bmatrix} C_{XX} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{w_x} \\ 1 \end{bmatrix}, \tag{2.9}$$

In the framework of the generative model of EEG (see 2.3.2) the regression approach can be utilized for finding a backward model $\boldsymbol{w}$ for a set of EEG measurements $\boldsymbol{x}(t)$ under the assumption that a neural source $s(t)$ exists with activity that co-varies with an external target function $z(t)$. In the present context, the target variable $z(t)$ typically is represented by the time course of an extracted music feature.

2.4.2 Canonical Correlation Analysis

A more general method for investigating relationships between two or more time series is Canonical Correlation Analysis (CCA), first introduced by Hotelling (1936). The core assumption of CCA is that two (or more) variables follow a common underlying time line. CCA finds a subspace for each variable, such that the respective projections, called canonical components, are maximally correlated. In the simple case of two time series $X_{N1\times T}$ and $Y_{N2\times T}$ and one canonical correlation coefficient this task can be formalized into the following objective function:

$$\max_{\boldsymbol{w}_x,\boldsymbol{w}_y} \mathrm{Corr}\,(\boldsymbol{w}_x^\top X, \boldsymbol{w}_y^\top Y). \tag{2.10}$$

The approach to obtain a single canonical correlation coefficient can be generalized towards finding several pairs of orthogonal subspaces W_x and W_y and corresponding canonical components. The number of pairs of subspaces that can be found is limited by the smaller one of the number of dimensions of the variables X and Y. In the case of several

subspaces the sum of the canonical correlation coefficients is maximized and the objective function is given by

$$\max_{W_x,W_y} \text{trace}(W_x^\top XY^\top W_y), \text{ subject to } W_x^\top XX^\top W_x = W_y^\top YY^\top W_y = 1. \tag{2.11}$$

The optimal W_x, W_y can be found by transforming Equation 2.11 into the following generalized eigenvalue problem (for the complete derivation see Bießmann et al. (2011)):

$$\begin{bmatrix} 0 & C_{YX} \\ C_{XY} & 0 \end{bmatrix} \begin{bmatrix} W_X \\ W_Y \end{bmatrix} = \Lambda \begin{bmatrix} C_{XX} & 0 \\ 0 & C_{YY} \end{bmatrix} \begin{bmatrix} W_X \\ W_Y \end{bmatrix}, \tag{2.12}$$

An important property of CCA is that resulting canonical correlation coefficients and components are invariant with respect to affine transformations of the variables. The basic CCA technique described above can be further extended to the use of more than two variables (Kettenring, 1971), a technique called Multiway-CCA. One particularly suited use for Multiway-CCA in neuroscience, is its application in so-called hyperscanning settings, where brain signals of several subjects exposed to the same stimulation in parallel are integrated into one Multiway-CCA model in order to detect neural activation that is shared across subjects (Bießmann et al., 2014, Gaebler et al., 2014). Furthermore, CCA has been demonstrated to be highly effective for multimodal analyses (Bießmann, 2011, Correa et al., 2010a,b). Importantly, in high-dimensional data settings where calculating covariance matrices can be problematic, the kernelized variant (tkCCA) (Bießmann et al., 2010) of CCA is used, in an analogue way to kernelizations of other decomposition methods (Schölkopf et al., 1998).

From Equations 2.9 and 2.12 it becomes clear that Linear Regression and CCA are closely related. The essential difference between both is that CCA takes the auto-covariance C_{YY} into account, while Linear Regression does not. This difference, however, becomes relevant only if both time series are multi-variate. A detailed discussion of the relation between Regression and CCA is contained in Borga et al. (1997) and Borga (1998).

2.4.3 Regularization

Optimization procedures are prone to suffer from the over-fitting problem. Over-fitting arises if the available amount of data is small relative to the complexity of the model that is learned. In the spirit of the principle of Occam's razor (Angluin and Smith, 1983) this effect can be alleviated by constraining the complexity of the solution (Müller et al., 2001, Vapnik, 2000), a technique called regularization. In the context of Linear Regression this problem can arise if, for instance, the weight vector $\boldsymbol{w}$ is determined using the OLS solution given in Equation 2.6. Often, the matrix $XX^\top$ has a determinant which is close to zero, which renders the matrix 'ill-conditioned', so the it cannot be inverted with the necessary precision.

A remedy for this problem is 'Tikhonov regularization', also known as 'Ridge Regression'. It can be accomplished by introducing a regularization term into the objective function (here: Equation 2.5) to reduce the complexity of the solution. This can be achieved for instance with a L2-norm regularization term (also Tikhonov regularization Tikhonov

and Arsenin (1977)) that 'penalizes' the L2 (=euclidean) norm of w and, thus, enforces solutions with smaller norms. Such a regularization term can be integrated by calculating a weighted sum of $XX^\top$ and νI where νI is the Identity matrix, multiplied by the average ν of the eigenvalues of $XX^\top$). νI represents the 'ridge' that is added to the sample covariance matrix $XX^\top$, so that Equation 2.6 is replaced by

$$\boldsymbol{w} = ((1-\gamma)XX^\top + \gamma\nu I)^{-1}XZ^\top. \tag{2.13}$$

This approach is motivated by the fact that an unfavorable ratio between number of available samples and dimensions introduces instability into the estimation of covariance matrices: small eigenvalues are estimated too small and large eigenvalues too large. The described regularization (called 'shrinkage technique') counteracts this effect and can be thought of as 'shrinking' the eigenvalues towards the average magnitude. The optimal shrinkage parameter γ that determines the contribution of the identity matrix can be determined analytically (Bartz and Müller, 2013, Blankertz et al., 2011, Ledoit and Wolf, 2004, Schäfer and Strimmer, 2005).

In CCA regularization can be included in an analogue way by modifying the covariance matrices on the right hand side of the equation. In this case Equation 2.11 is replaced by Equation 2.14 where the auto-covariance matrices are expanded by a 'ridge'.

$$\begin{bmatrix} 0 & C_{YX} \\ C_{XY} & 0 \end{bmatrix} \begin{bmatrix} W_x \\ W_y \end{bmatrix} = \Lambda \begin{bmatrix} (1-\gamma_x)C_{XX} + \gamma_x\nu I) & 0 \\ 0 & (1-\gamma_y)C_{YY} + \gamma_y\nu I) \end{bmatrix} \begin{bmatrix} W_x \\ W_y \end{bmatrix}, \tag{2.14}$$

2.4.4 Temporal Embedding

Naturally, a cortical brain response tracking auditory stimulus features will not respond instantaneously, but delayed. Typical cortical responses, such as ERPs, belong to the family of mid-latency components which, typically, occur with a delay ranging from tenths to hundreds of milliseconds. Consequently, when one regresses the brain signal onto a feature extracted from the music stimulus, a time delay has to be factored in. In the event that a clear a-priori assumption about the lag of the brain response with respect to the stimulus does not exist, a 'temporal embedding' of the brain signal prior to the optimization procedure can be applied. This technique has been proposed by Bießmann et al. (2010) and allows to deal with couplings between signals with an unknown delay between them.

In order to be able to account for delays from of 1, $\dots k$ samples lag, k copies of the original dataset (here $X_{N\times T}$, a data matrix with T observations in N channels) are created. Each of these copies is shifted by one of the 1, $\dots k$ time lags. Subsequently, the shifted copies are added as 'artificial' channels to the original data set, resulting in data matrix X_{emb} of the dimensionality $N \cdot (k+1) \times (T-k)$. The structure of the resulting matrix is given in Equation 2.15 where the column vectors $\boldsymbol{x}(t)$ denote the observations on all N electrodes at time point t.

$$X_{emb} = \begin{bmatrix} \boldsymbol{x}(1) & \boldsymbol{x}(2) & ... & \boldsymbol{x}(T-k) \\ \boldsymbol{x}(2) & \boldsymbol{x}(3) & ... & \boldsymbol{x}(T-k+1) \\ ... & ... & ... & ... \\ \boldsymbol{x}(k+1) & \boldsymbol{x}(k+2) & ... & \boldsymbol{x}(T) \end{bmatrix} \tag{2.15}$$

In practice, the choice of time lags for the embedding is informed by prior knowledge about the mechanisms expected to play a role in the processing of the studied stimulus feature. The new, embedded data matrix X_{emb} can then be fed into the regression or CCA procedures described above, yielding filters of the dimension $N \times (k+1)$ that correspond to $k+1$ spatial filters for the time lags $0, \ldots k$ samples.

2.4.5 Cross-validation

Cross-validation Geisser (1993) is a model validation technique for evaluating how well an estimated predictive model generalizes to an independent dataset. This technique involves partitioning a data set into disjunct subsets, training a model on one subset (called the training set), applying the trained model on the remaining data (called the validation set or testing set) and assessing its performance by means of an adequate measure.

2.4.6 MUltiple Signal Classification (MUSIC)

The multiple signal classification (MUSIC) algorithm is a technique to localize a defined number of neural dipolar sources that correspond to observations at the sensor level.

The general idea of the algorithm was introduced by Schmidt (1986) as an extension of Pisarenko's covariance approach (Pisarenko, 1973) for the estimation of parameters of complex sinusoids in additive noise. In neuroimaging it has been applied for source localization in many variants (Mosher et al., 1992, Mosher and Leahy, 1998, Shahbazi et al., 2012).

The key idea of this approach is to reduce an observed signal of interest (e.g., a set of spatial ERP patterns) to a lower-dimensional subspace and, to then, by means of a 3D forward head model, find a set of dipoles that optimally matches this subspace.

This is done by first separating the space of multi-dimensional observations into orthogonal subspaces that represent the signal of interest and noise-only, respectively, by means of singular value decomposition (SVD) and a threshold. Then, in a scanning approach, a grid of source points covering the 3D-head model is defined and for each candidate source point the angle between projection and signal subspace is calculated in order to find the dipole location and orientation with the best fit. The scanning metric is given by the cosine of the principal angle θ between dipole and signal subspace

$$\cos^2 \theta(L, \phi) = \frac{L^\top \phi \phi^\top L}{L^\top L}. \tag{2.16}$$

where $L \epsilon \mathbb{R}^{1 \times N}$ is the forward model (defined by the 3D head model) for a point source and ϕ the matrix of eigenvectors corresponding to the set of largest eigenvalues of the covariance matrix that were selected to represent the original EEG data.

This approach can be regarded as a way to obtain a solution for a least squares dipole fitting problem with a non-convex cost function (Darvas et al., 2004) that is problematic to solve analytically.

In Chapter 4.2.3 we extend this well-established procedure for application to with spatio-temporal patterns.

2.4.7 Statistical Problems

In the present analysis scenario where, both, time series data describing naturalistic music stimuli, and representing brain signals are examined with correlation-based methods, two issues require special consideration:

1. Correlation of brain signals with a set of interrelated features.
2. Correlation between auto-correlated time series.

2.4.7.1 Correlation of brain signals with a set of interrelated features

Typically, a description of a complex auditory stimulus comprises a multitude of features that are not independent of each other, but are correlated to different degrees. Only by accounting for this correlation, one can attribute a particular brain signal to one particular music feature. The Partial Correlation Coefficient is a simple, but effective method to assess the relation between two variables while removing the influence of one ore more of other variables (Kendall et al., 1973). This post-hoc approach is a way to exert statistical control over variables in a setting where experimental control over the different aspects that are to be investigated is ruled out or incompatible with the study design. The partial correlation coefficient between two time series x and y while removing the influence from z is given by Equation 2.17.

$$r_{xy.z} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{(1 - r_{xz})^2 (1 - r_{yz})^2}}. \tag{2.17}$$

Within the framework of linear regression analysis, the partial correlation coefficient can be derived as the correlation of the residuals that are produced if the interfering variable z (that is to be eliminated) is used as a regressor to predict each of the two variables of interest x and y (Abdi, 2007). The partial correlation coefficient is related to multiple linear regression analysis (MLR) It was applied in (Schaefer et al., 2009) to decompose EEG responses into components of evoked response that can be linked to specific aspects of music stimuli. Importantly, the partial correlation coefficient differs from the semi-partial correlation/regression coefficient of the multiple linear regression framework in that: The partial correlation coefficient eliminates the influence of the interfering factor from both variables of interest, not only from one (in the framework of MLR: from the regressor). As a consequence, using the partial correlation coefficient,

shared variance that does not cover a large proportion of the total variance, but may still reflect specific relations, is also detected. In a different context, partial correlation has been applied previously in connectivity analysis of EEG recordings: In Marrelec et al. (2006) it was used to identify connections between brain areas while accounting for the effects of volume conduction between electrodes.

2.4.7.2 Correlation between auto-correlated time series

It is important to realize that, typically, brain signals as well as time courses of features extracted from a music audio signal both comprise a high degree of autocorrelation as successive sampling points are not independent of each other. This fact violates the assumptions that underlie the standard tests for significance of correlation. This problem can be accounted for in different ways: Pyper and Peterman (1998) for instance propose a method that, given two (autocorrelated) time courses u and v of length N, estimates the effective degrees of freedom $\hat{df}$ needed to calculate the p-value for $r(u, v)$ based on the auto-correlation r_{uu} and r_{vv}

$$\hat{df} = \frac{1}{\frac{1}{N} + \frac{2}{N} \cdot \sum_{j=1}^{N} \frac{N-j}{N} r_{uu}(j) r_{vv}(j)}. \tag{2.18}$$

In the music domain this method has been applied in Alluri et al. (2012).

A different strategy can be pursued by applying randomized permutation tests with surrogate data as proposed in Theiler et al. (1992). If, for instance, the significance of $r(u, v)$ is assessed, a surrogate target function for v can be generated. This is done by transforming the time course v into the frequency domain, randomly permuting its phase spectrum, and reconstructing the time domain signal using the original spectral amplitudes and the permuted phases. This surrogate function $\tilde{v}$ can be correlated with the original u. Repeating this process a number of times results in a distribution of correlation coefficients for the surrogate data. This gives an estimate of how likely a correlation coefficient of the magnitude of the correlation observed between u and the original v was if the signals had the same auto-correlation properties, but no relationship between them, i.e. if the null-hypothesis was true.

2.4.8 Audio analysis

The earlier parts of this chapter introduced the idea to utilize target functions derived from the music audio signal in order to develop backward models that extract stimulus-related neural activity from the EEG as one key part of this thesis. Obviously, this approach relies critically on the selection and extraction of relevant features from the music signal.

Quantitative (automated) analyses of music signals have a long tradition in the field of Music Information Retrieval (MIR). They are at the basis of typical MIR tasks, such as content-based audio tagging, music-related search engines, audio classification and prediction of genre, style, or composer. In music perception research the interest in describing variables of human perception in terms of musical features has grown substantially in the last decade. In particular, the relationship between stimulus properties and affective

responses has been explored in numerous studies (for an overview see Gabrielsson and Lindström E. (2010)). Thus, automated analysis of music signals has become part of the basic methodological inventory of researchers, a development that has been aided by a range of freely available tools for audio analysis, such as the MIRtoolbox Lartillot et al. (2008b), and a number of active communities.

However, the abundance of available features poses the problem of identifying an appropriate description of an audio signal. Typically, investigations on the neural processing of music are primarily interested in basic variables of music, such as loudness, pitch, rhythm, tempo, timbre, articulation (Deutsch, 2013). These, however, are complex perceptual categories, each of which encompass multiple aspects. Unfortunately, for hardly any of these perceptually defined aspects of music a 'gold standard' exists how to represent it by features extracted from the audio signal.

In the following the most important general considerations that may guide the design of an audio analysis in this context are discussed. Then, the most relevant domains of features are described. The focus lies on automated analysis of music stimuli in the form of audio waveforms, neither considering symbolic descriptions of music, such as the MIDI format or score representations, nor representations that include meta-data.

Following Mitrović et al. (2010) audio features suggested in the literature can be characterized formally with respect to their domain, their time scale, their semantic meaning and with respect to the presence or nature of an underlying model. The domain of an audio feature describes the basic property of the data represented by that feature. For example, a feature in the time domain, such as the Zero Crossing Rate, describes aspects of the waveform (in the time domain), while a feature in the frequency domain, e.g., the spectral centroid, relates to an aspect of the signal's spectrum. Important domains are the time domain, the frequency domain, the cepstral domain and the modulation frequency (Mitrović et al., 2010).

Furthermore, audio features can be extracted at different time scales. Typically, a distinction is made between short-term features (also intra-frame features) extracted for time frames of up to approximately 100 ms (e.g., spectral centroid), long-term features (also inter-frame features) that describe audio signals over time intervals of up to several seconds (e.g., fluctuation features) and global features that are present across the entire audio signal (e.g., key or mode).

An audio feature's semantic meaning indicates whether it is a physical audio feature or whether it directly relates to human perception. Physical features, such as, e.g., sound pressure level, describe the underlying audio data in mathematical or statistical terms. They are transparent with respect to their extraction process, but in general do not directly relate to human perception. Perceptual features, naturally, are also derived via mathematical descriptions of audio data, but typically are combinations of several physical features that have been established as reflecting perceptual categories. For instance, pulse clarity, an audio feature that combines several spectral physical features, has been introduced and evaluated by Lartillot et al. and is regarded to reflect the salience of the pulse in music Lartillot et al. (2008b). However, even though a number of perceptual audio features exist in the literature, it is sometimes not clear how far the perceptual relevance generalizes across a variety of stimuli.

There are three basic strategies for selecting features that describe an audio signal: (i)The analyses can be based on physical audio features that serve as a technical proxy

for perceptual categories that, in principle, is based on assumptions or best practice. (ii) Alternatively, established perceptual features can be used, but then it has to be considered whether their perceptual relevance generalizes. (iii) The perceptual relevance of a given feature for a given stimulus set can be utilized, after being proven explicitly (see for instance Alluri et al. (2012) and Cong et al. (2012)), which, requires considerable experimental effort.

Finally, features may be further characterized through the presence and nature of an underlying model. Models may be psychoacoustic, such as Zwicker's loudness model (Zwicker and Scharf, 1965) that takes into account the transfer function of the peripheral auditory system for transforming sound pressure level in loudness. Audio features can alternatively also be derived through models from music theory, e.g., to derive a transcription from audio signal into chords symbols (Mauch and Dixon, 2010).

For a comprehensive overview about feature categorization and a taxonomy of 77 features with references to the audio retrieval literature see Mitrović et al. (2010).

Chapter 3

Finding reflections of an original rock song in ECoG high gamma activity

3.1 Introduction

In Chapter 1 we have seen that the number of neuroscientific studies that let their subjects listen to natural music is still small. Out of these, only few used the electroencephalogram (EEG), which, in principle, is a highly attractive modality for investigating the dynamics of music on a fine-grained time scale, due to its high temporal resolution. In Chapter 2 we have highlighted the additional benefits of electrocorticographic recordings, but also pointed out that the availability of this kind of data is limited. To our knowledge there exists only one data set of ECoG recordings where patients were presented with a full length naturalistic music stimulus. This data set has been subject to several investigations. A first example how the time course of sound intensity of a naturalistic music stimulus can be tracked in ECoG features was provided by Potes et al. (2012). Specifically, this study revealed that high-gamma band (70-170 Hz) ECoG activity in the superior temporal gyrus as well as on the dorsal precentral gyrus is highly correlated with the time course of sound intensity in a continuous stream of natural music. A subsequent study by Kubanek et al. (2013) found that high-gamma ECoG activity also tracks the temporal envelope of speech and compared it to the activations related to music. This analysis revealed different levels of specificity in an auditory network constituted by the auditory belt areas, the superior temporal gyrus (STG) and Broca's area. Recently, a new analysis of the same data set identified spatial and causal relationships between alpha and gamma ECoG activity related to the processing of sound intensity (Potes et al., 2014).

Considering that sound intensity (a technical proxy for perceived loudness) was tracked in ECoG features with significant robustness, the same data set appeared highly promising for a further investigation that takes into account the variety of features available in this natural music stimulus, a rock song. Therefore, the goal of the present follow-up analysis was to explore whether music-related variables other than sound intensity can be tracked in ECoG and, if so, how respective areas of cortical activation compare to those associated with the processing of sound intensity in Potes et al. (2012). Furthermore,

ECoG recordings were available where the same subjects listened to natural speech. This represented an opportunity to compare the processing of audio features in two different categories of sounds.

Because a naturalistic music stimulus contains different perceptual dimensions that are intrinsically related, it was a critical challenge to differentiate these in the brain response. In addition to the feature of sound intensity that was investigated in the previous studies, we chose four features that relate to different aspects of music. These include the moment-to-moment distinction vocals on/off, a continuous measure of harmonic change probability, a measure related to timbral aspects (spectral centroid), and a rhythm-related measure (pulse clarity).

3.2 Methods

3.2.1 Subjects and data collection

We analyzed data from ten subjects (for patient's clinical profiles see Sturm et al. (2014)). These 10 subjects included seven of the eight subjects who were analyzed in Potes et al. (2012) where patients with epilepsy (4 women, 4 men) were instructed to listen attentively (without any other task) to a single presentation of the rock song "Another Brick in the Wall - Part 1" (Pink Floyd, Columbia Records, 1979) while ECoG activity was recorded. In all patients in the present analysis the electrode grid was in the left hemisphere. None of the subjects had a history of hearing impairment. After removal of channels containing environmental or other artifacts the number of implanted electrodes left for analysis ranged between 58 and 134 channels. Grid placement and duration of ECoG monitoring were based solely on the requirements of the clinical evaluation without any consideration of this study. Each subject had postoperative anterior–posterior and lateral radiographs, as well as computer tomography (CT) scans to verify grid locations. The song was 3:10 min long, digitized at 44.1 kHz in waveform audio file format, and binaurally presented to each subject using in-ear monitoring earphones (12 to 23.5 kHz audio bandwidth, 20 dB isolation from environmental noise). ECoG signals were referenced to an electrocorticographically silent electrode (i.e., a location that was not identified as eloquent cortex by electrocortical stimulation mapping), digitized at 1200 Hz, synchronized with stimulus presentation, and stored with BCI2000 (Schalk et al., 2004, Schalk and Mellinger, 2010). In addition, we analyzed data from the same subjects where they listened to the presentation of four narrated stories that are part of the Boston Aphasia Battery (Goodglass et al., 1983) (for details see Kubanek et al. (2013)).

3.2.2 Extraction of ECoG features

Our analysis focused on the high-gamma band. ECoG activity in the high gamma band has generally been associated with functional activation of the cortex in different domains (Crone et al., 2006). We extracted ECoG high-gamma power using the same method as in (Potes et al., 2012): high-gamma (70-170 Hz) amplitudes were extracted by first applying a 0.1 Hz high-pass filter and then a common average reference (CAR) spatial filter to the ECoG signals. For every 50 ms window, we estimated a power spectrum from the time-series ECoG signal using an autoregressive (AR) model. Spectral magnitudes

were averaged for all frequency bins between 70 and 115 and between 130 and 170 Hz (omitting line noise at 120 Hz).

3.2.3 Audio analysis

Audio feature selection From the large number of potential features that characterize a music audio signal, we chose a set of five features that capture salient dynamic features of the stimulus and cover a broad spectrum of structural categories of music. Since the results of Potes et al. (2012) revealed a strong correlation of ECoG high-gamma power fluctuations with the sound intensity of the continuous music stimulus, sound intensity was chosen as first feature. It is a temporal feature that can be extracted directly from the raw audio signal and can be considered as an approximate measure of loudness. The second feature was the logarithmic spectral centroid, which is perceptually related to the complex property of timbre. More specifically, it has been related to perceived brightness of sound in Schubert et al. (2004) and to perceived pitch level in Coutinho and Cangelosi (2011). The third feature was probability of harmonic change, which relates to higher-level musical structure, i.e., to harmonic progression and musical syntax. Pulse clarity as fourth feature indicates how easily listeners perceive the underlying rhythmic or metrical pulsation of a piece of music. This feature has been introduced and perceptually validated in Lartillot et al. (2008b) and since then has been used in numerous studies (Alluri et al., 2012, Burger et al., 2013, Eerola et al., 2009, Higuchi et al., 2011, Zentner, 2010). Since an essential characteristic of the music stimulus is the presence of song with lyrics, the fifth feature, vocals on/off, captures the change between purely instrumental passages and passages with vocal lyrics content. In summary, we chose a description of the audio signal that relates to important basic variables of the perception of music: loudness, timbre, and rhythm. With harmonic change, it encompasses also an abstract high-level property related to the rules of Western major-minor harmony. Finally, with vocals on/off, it allows also to address the impact of vocals with lyrics in music. For comparison, in a complementary analysis, the identical analysis was applied to the sound files of the speech stimuli.

Audio feature extraction Sound intensity was calculated in Matlab (The MathWorks Inc., Natick, Massachusetts). Vocals on/off was determined manually. All other features were extracted using freely available software (see below). We used the first 125 seconds of Pink Floyd's The Wall - part 1 in the analysis since the last minute of the song is an instrumental afterlude passage with considerably less variation, in particular without any vocal parts. The five features were calculated as described in the following paragraphs:

Sound intensity The sound intensity of the audio signal was calculated as the average power derived from 50 ms segments of the audio waveform overlapping by 50%. The resulting time course was downsampled to match the sampling rate of 20 Hz of the extracted ECoG high gamma power.

Vocals on/off The presence of vocals was annotated manually in the audio file. This annotation resulted in a binary function that contained the value 1 for passages with lyrics and 0 otherwise. In the music stimulus there are seven passages with vocal lyrics with average duration of 4.22 s ($\pm$0.77) that are separated by at least 5 s of purely

instrumental music. In a complementary analysis, we applied a similar procedure to the speech stimuli. Here, 0 was assigned to passages of silence within the story that exceeded the duration of 400 ms, such as pauses between sentences or phrases, while 1 denoted ongoing speech. In the speech stimulus the duration of speech passages was shorter (mean duration 1.65 s ±0.55) and vocals on/off changes occurred more frequently (30 changes in 100s). In both stimuli the analyzed data start with the first tone of the song or with the first sentence of the narration, respectively, not including a silent pre-stimulus period.

Spectral centroid The centroid of the log-scaled frequency spectrum was calculated for 50% overlapping windows of 50 ms using the implementation in the MIRtoolbox (Lartillot et al., 2008b). The spectral centroid is the amplitude-weighted mean frequency in a window of 50 ms. It is an acoustic measure that indicates where the 'mass' of the spectrum is located. The log-scaled centroid was downsampled to match the sampling rate of 20 Hz of the extracted ECoG high gamma power.

Pulse clarity Pulse clarity was calculated for windows of 3 s with a 33% overlap using the MIRtoolbox (Lartillot et al., 2008b), then interpolated to match the ECoG sampling frequency of 20 Hz. Pulse clarity is a measure of how strong rhythmic pulses and their periodicities can be perceived by the listener. It is based on the relative Shannon entropy of the fluctuation spectrum (Pampalk et al., 2002) and has been perceptually validated as being strongly related to listener's perception of the degree of rhythmicity in a piece of music in Lartillot et al. (2008a).

Harmonic change The harmonic change function measures the probability of a harmonic change and detects chord changes. We derived this metric using the Queen Mary plugin for the sonic visualizer (del Bimbo et al., 2010), which implements an algorithm that was proposed and validated on a selection of rock songs in Harte et al. (2006). The algorithm comprises a segmentation of the audio signal into 50 ms windows, spectral decomposition of each window, assignment of chroma and a tonal centroid to each window. After that, the tonal distance between consecutive frames is calculated based on a hypertoroid model of tonal space proposed by Chew (2000).

Figure 3.2 gives a visual representation of each stimulus' spectrogram, an annotation of lyrics and chords or text, and the time courses of the five extracted music features for a 12 s-segment of the song.

3.2.4 ECoG data analysis

3.2.4.1 Partial correlation

The five features that we used to describe the music stimulus are not independent of each other, but are correlated with each other to variable degrees (see Figure 3.1). As described in Chapter 2.4.7 we calculate the Partial Correlation Coefficient in order to examine how much each of the five features of music contributes to the sensor-level ECoG recordings in a manner that is independent from the remaining four features.

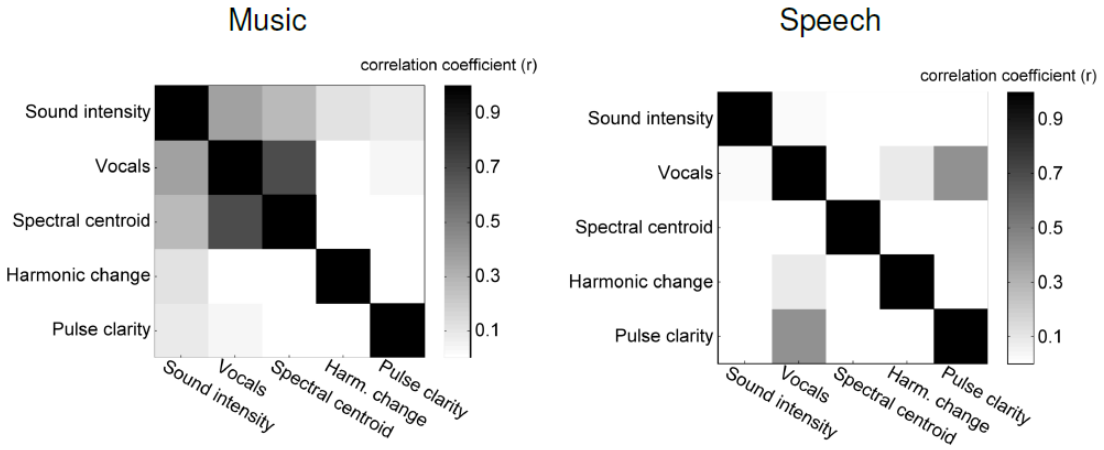

FIGURE 3.1: Correlation between five stimulus features: left: music stimulus, right: speech stimulus.

To account for autocorrelation in both, the extracted music features and the ECoG high gamma time courses, we assessed the significance of the partial correlation coefficients by applying randomized permutation tests with surrogate data (see Chapter 2.4.7). For each music feature, we generated a surrogate target function, calculated the partial correlation coefficient between the ECoG feature and a set of surrogate target functions. We repeated this process 1000 times, which resulted in a distribution of correlation coefficients for the surrogate data.

The resulting p-values were corrected for multiple comparisons within all electrodes (false discovery rate (FDR), $q<0.05$). We then plotted the negative logarithm of the corrected p-values for each electrode on each subject's brain model as an indicator of how much brain activity at a particular site was related to a specific acoustic feature. Since we did not observe negative correlation coefficients, there was no need to distinguish between negative and positive correlation.

3.2.4.2 Latency of brain response

Naturally, one would expect that a cortical brain response that tracks features of an auditory stimulus will not respond instantaneously, but delayed. Accordingly, we examined the channel-wise partial correlation coefficients with time lags up to 300 ms. However, this resulted in cross-correlation sequences that varied only on a very small scale over time and were not conclusive with respect to an optimal time lag, suggesting that a time lag between stimulus and brain response may be evened out by our sampling rate of 20 Hz. For instance, selecting a biologically plausible time lag of 100 ms, based on Kubanek et al. (2013) where the optimal (averaged) time lag for tracking the speech envelope ranged between 86.7 ms and 89.9 ms, had only an marginal effect on the significance of correlation coefficients, although the magnitude of correlation coefficients varied slightly (but not systematically). An overview of the group-level results for different time lags is depicted in Figure 3.8. On these grounds it would have been arbitrary to define a fixed time lag for the analysis and, moreover, a chosen time lag would not have been informative. Therefore, we decided to calculate instantaneous correlation coefficients in the present analysis, using this is a neutral or 'null' hypothesis given that no significant

estimate of a biologically plausible time lag was obtainable for this data set. For a detailed analysis of latencies, in particular with respect to differences in the processing of different aspects of music, as suggested in Schaefer et al. (2011a), our approach is not appropriate since the dependencies between the five features play a role in calculating the partial correlation coefficients for one music feature and the ECoG signal. This could be a topic for a future investigation, for instance applying methods proposed in Bießmann et al. (2010) or Power et al. (2012).

3.2.4.3 Calculating group-level results

Since these measures of significance cannot be directly averaged across subjects, to examine the topographical distribution of significant correlations at the group-level, we visualized the results as following: for each subject, we determined electrodes with significant correlation and projected their positions onto the MNI brain. To detect activated electrodes in similar regions, each of these electrodes was represented by a round patch of activation with radius 10 mm (called 'activation radius' in the following) centered around its position. These representations were added up for the 10 subjects, resulting in a map showing the topographical overlap of the presence of significant correlation within the group of subjects. Values range from zero (no significant correlation in all ten subjects) to ten (significant correlation in all ten subjects).

Choosing an activation radius that corresponds to the inter-electrode distance (here: 10 mm) is a common practice for visualizing ECoG activations (Kubanek et al., 2013, Potes et al., 2014, 2012). The (standard) inter-electrode distance of 10 mm is based on the estimated spatial resolution of ECoG that takes into account volume conduction within the cortical tissue (Mullen et al., 2011). Yet, it has to be kept in mind that the activation radius does not necessarily need to be the same as the inter-electrode distance, but that it is a parameter that can be adjusted. In general, choosing an activation radius is a trade-off between spatial specificity and sensitivity with the respect to the detection of common activation patterns between subjects. This means that, e.g., in a group of patients with similar coverage of a certain cortical area, a smaller activation radius may be useful to detect common foci of activation with high spatial specificity. If grid placement varies more between subjects, a larger activation radius, e.g., corresponding to the inter-electrode distance, might ensure that shared activations at a larger topographical scale become visible.

Since grid placement was determined by clinical requirements and, consequently, varied between patients, we needed to account for the fact that the maximal number of subjects who can contribute to the group-level overlap of activation also varies between brain regions. Therefore, we determined the group-level overlap of grid coverage on the MNI brain, referred to as grid coverage index in the following, for all electrodes. Using the grid coverage index, a normalized group-level overlap in a specific cortical area can be obtained by dividing the (unnormalized) group-level overlap by the grid coverage index for each vertex. However, even the normalized group-level overlap values cannot be used for inferring group-level statistics, for instance to assess differences between brain areas. Nonetheless, this does not affect the primary goal of the present analysis, which is to explore potential differences in one location between features and also between the conditions music and pure speech. For distinct foci of high degree of group-level overlap, we

determined representative coordinates on the MNI brain manually, and derived the corresponding Brodmann areas using the Talairach Atlas daemon[1]. Owing to the variance introduced by the projection of each subject's individual brain onto the MNI brain and to the blurring effect that the above mentioned procedure of determining group-level overlap may cause, this procedure yields only an approximate localization of cortical activation. Notwithstanding, on the scale of the Brodman area, this level of exactness appears appropriate for comparing the present results with the literature.

3.3 Results

3.3.1 Stimulus features

Figure 3.1 shows a confusion matrix. For each element in this matrix, the brightness gives the correlation between two particular music features. In the music stimulus vocals on/off is strongly correlated with spectral centroid (r=0.69) and intensity (r= 0.37), which confirms the necessity for calculating partial correlations. Figure 3.2 gives a visual representation of each stimulus' spectrogram, an annotation of lyrics and chords or text and the time courses of the five extracted music features for a 12s-segment as well as the time course of ECoG high gamma power, measured at one representative electrode in one subject.

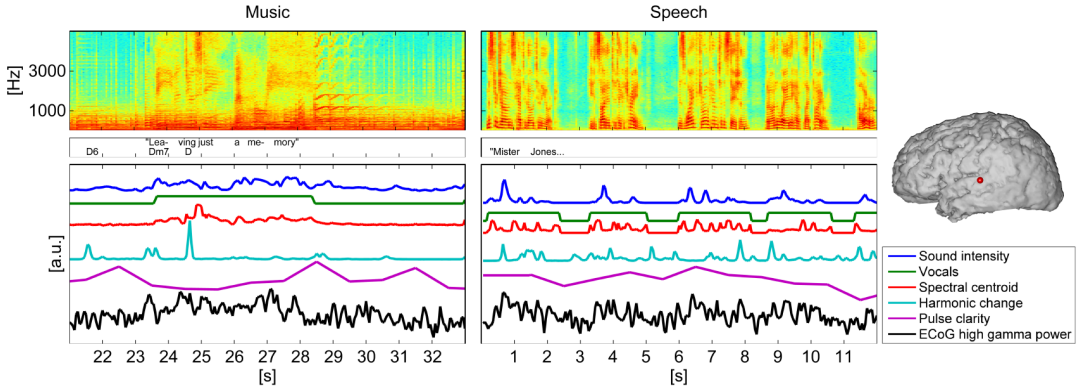

FIGURE 3.2: Spectrogram of a segment (12 seconds) of the music/speech recording, lyrics/text and chord annotations and time courses of the five analyzed features. For comparison with the time course of the music features the time course of ECoG high gamma power, measured at one representative electrode of subject S5 was added below. The location of the electrode is indicated on the brain model on the right panel.

3.3.2 Grid coverage

Figure 3.3 documents the overlap of grid coverage (grid coverage index) within the group of patients. The regions covered in all of the ten subjects comprise the posterior part of the superior temporal gyrus and the ventral parts of the precentral and postcentral gyri.

[1]http://www.talairach.org/daemon.html

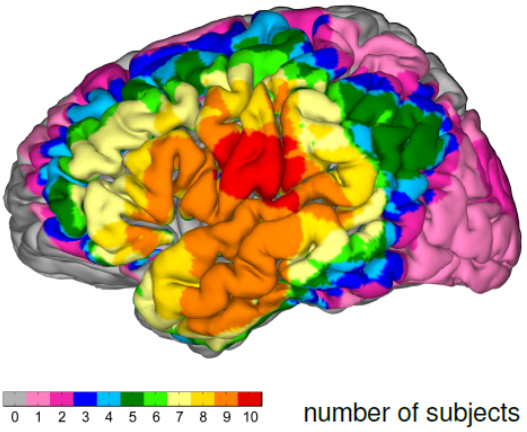

FIGURE 3.3: Grid coverage index: Overlap of grid coverage on MNI brain.

3.3.3 Single-subject results

Figure 3.4 shows the significance values of partial correlation of ECoG high-gamma features with each of the five music features for each individual patient. Significant high-gamma correlations with vocals on/off are present in 9/10 of the subjects, and exceed in spatial extent those of all other acoustic features. In all of these nine patients, significant positive correlations are present in auditory areas around the Sylvian fissure. In addition, significant correlation in an isolated area at the dorsal precentral cortex is present in three subjects (S3, S5, and S9). Compared to the effect related to vocals on/off, correlation with sound intensity (after calculating the partial correlation and thereby rendering it now independent from fluctuations in the other four acoustic parameters, including vocals on/off) is low, reaching significance only in subject S4, S5, S7 and S10) and is detected only in a smaller region on the posterior Superior Temporal Gyrus (pSTG). Correlation with spectral centroid is significant only in subject S5 and S10 and distributed similarly to the feature vocals on/off, but spatially less extended. For harmonic change, significant correlation is present in four subjects (subject S3, S5, S9 and S10) on the posterior STG and in subject S3 in frontal areas. The correlation with pulse clarity reaches significance in only one subject (S6) in a small region on the precentral cortex.

Figure 3.5 depicts the cortical distribution of significant partial correlation of ECoG high-gamma features with each of the five acoustic features for the natural speech stimuli at the level of each individual patient. Differing from the music condition, the feature that is reflected most consistently within the group is sound intensity with significant correlation in 6/10 subjects (S1, S2, S3, S4, S5, S9, and S10). In all of them, the focus of correlation is located on the pSTG. For the feature spectral centroid, significant correlations are present only in three subjects on the superior and medial temporal gyrus.

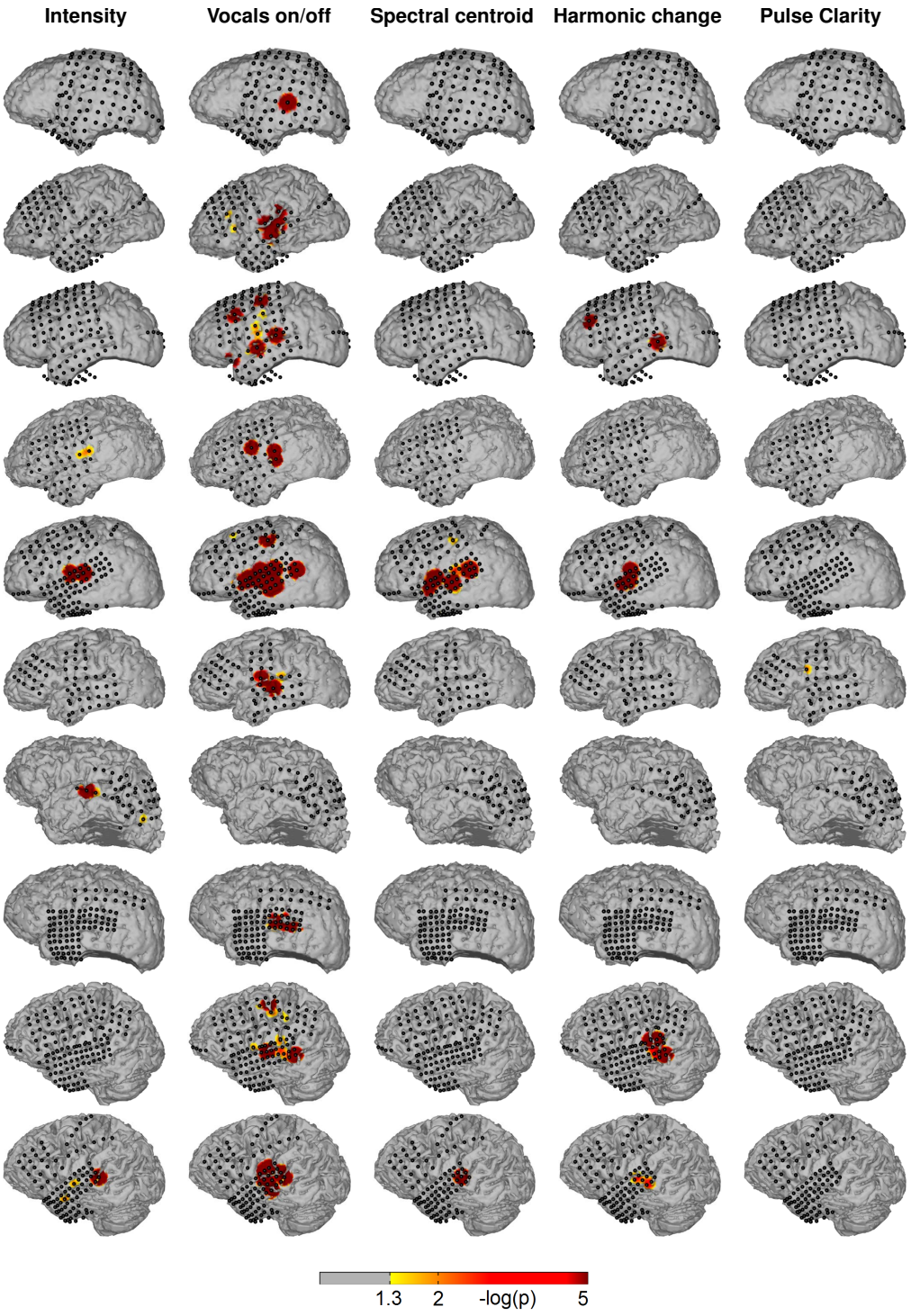

FIGURE 3.4: Single subjects (individual brain models), music condition: Cortical distribution of significant correlation with each of the five acoustic features after removing the influence of the remaining four features by calculating partial correlation coefficients. A value of 2 corresponds to a p-value of 0.01. Correlation coefficients determined as significant by permutation tests ranged between r=0.07 and r=0.26.

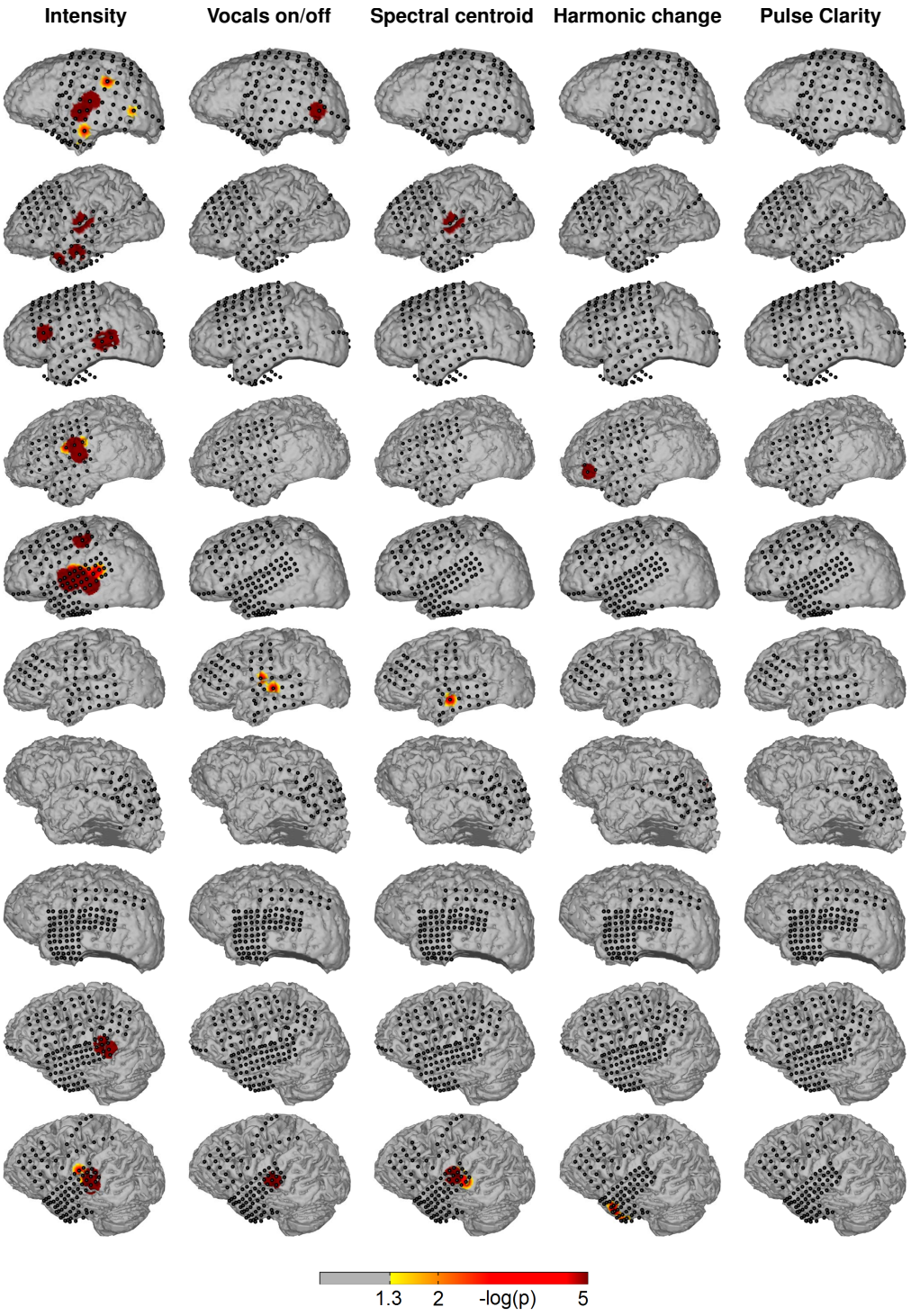

FIGURE 3.5: Single subjects (individual brain models), speech condition: Cortical distribution of significant correlation with each of the five acoustic features after removing the influence of the remaining four features by calculating partial correlation coefficients. A value of 2 corresponds to a p-value of 0.01. Correlation coefficients determined as significant by permutation tests ranged between r=0.06 and r=0.16.

3.3.4 Group-level results

Figure 3.6 directly compares the group-level overlap of significant 'standard' correlation (Pearson's correlation coefficient, top row) of high-gamma ECoG features with each of the five music features with that of the partial correlation coefficient (middle row). In general, at a descriptive level, the similarity between cortical overlap patterns mirrors the correlation matrix of the music features in so far as they mainly document the interdependence of musical features rather than allowing to differentiate between processing of specific dimensions of music. The middle row of Figure 3.6 gives the group-level overlap of significant correlation of high-gamma ECoG features with each of the five music features after the influence of the remaining four other features has been removed by calculating partial correlations (see Section 2.4.7). The highest degree of overlap is present in the feature vocals on/off with significant correlation of high-gamma power with vocals on/off in more than seven subjects around the Sylvian fissure, covering the posterior and middle part of the superior temporal gyrus and of the middle temporal gyrus. The point of most consistently detected activations in the present group of subjects is the posterior part of the superior temporal gyrus (9/10 subjects). Furthermore, overlap of significant correlation is present in the precentral gyrus in three subjects. For all other features, the group-level overlap is considerably less: for sound intensity, there is a common focus of activation in the anterior peri-Sylvian area in three patients. Locations of significant correlation for harmonic change vary along the STG, amounting to a number of three overlapping subjects at maximum. Significant correlation with spectral centroid is distributed around the Sylvian fissure, however with minimal inter-individual overlap.

The bottom row of Figure 3.6 shows the group-level overlap of significant correlation for complementary analysis of speech-only stimuli. The overlap of significant correlation with sound intensity is distributed around the Sylvian fissure with highest values on the middle part of the STG, corresponding to the respective location in the music condition, but with five contributing subjects, compared to three subjects in the music condition. However, for all other features the degree of overlap does not exceed two subjects in any location. Figure 3.7 shows the group-level overlap depicted in Figure 3.6, normalized with respect to the grid coverage index depicted in Figure 3.3. We included only cortical sites with a minimum grid coverage of 2 subjects. This representation demonstrates that the characteristic patterns of the group-level overlap representation (Figure 3.5) do not merely reflect the distribution of the grid coverage index, but that the distribution of significant correlation has features that are consistently present in a large proportion of the subjects in which grid coverage is given.

Figure 3.8 shows the group-level overlap of significant correlation for delays of 0, 50, 100, 150, 200, 250 and 300 ms of the time course of the ECoG high gamma power and the music features.

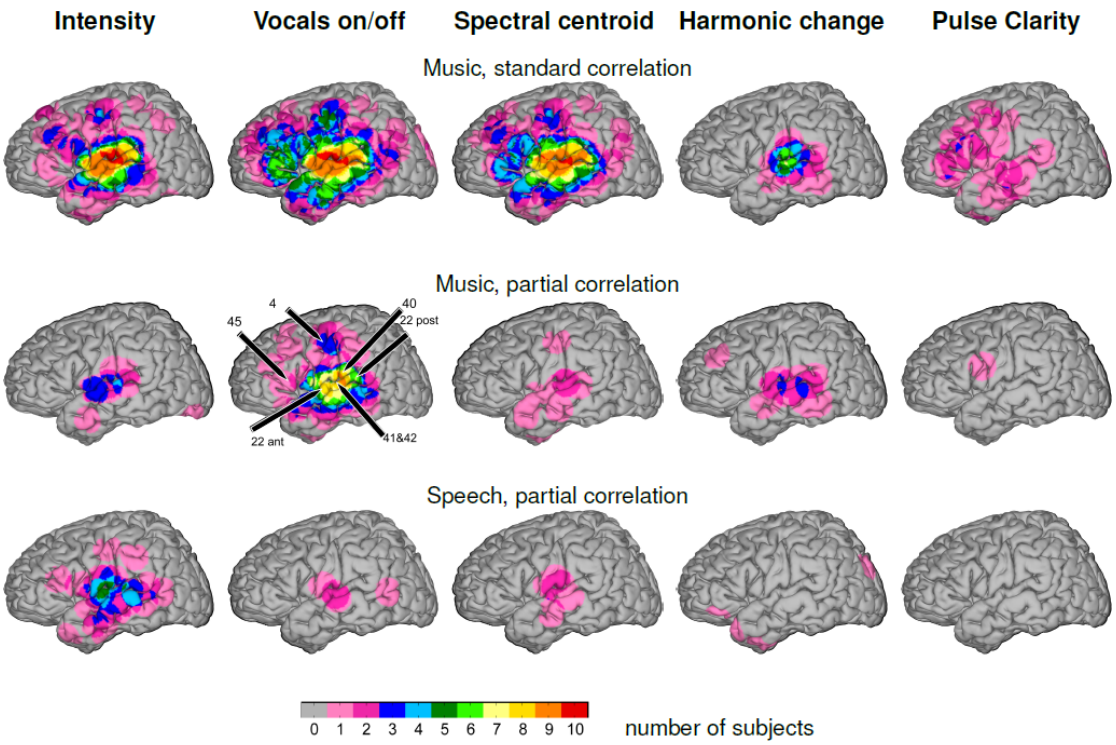

FIGURE 3.6: Number of participants with effects visualized on the MNI brain. The color code indicates the degree of group-level overlap. Top: Music, 'standard' correlation. Middle: Music, partial correlation. Bottom: Speech, partial correlation.

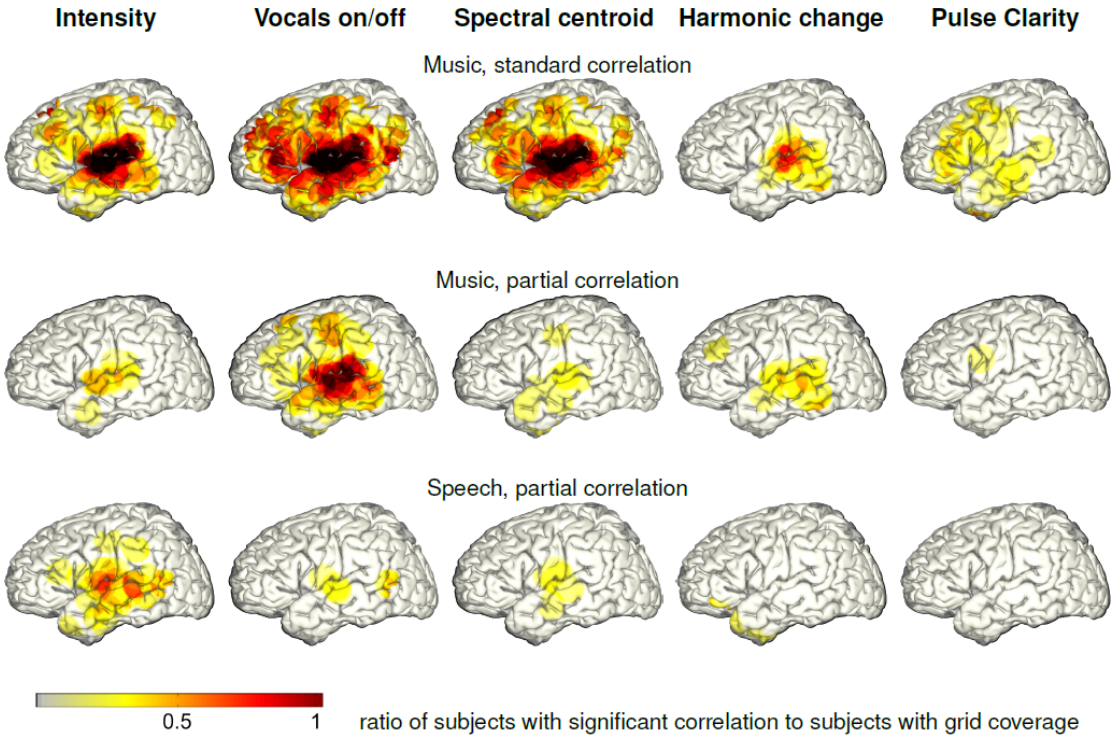

FIGURE 3.7: Number of participants with effects visualized on the MNI brain, **normalized with respect to the grid coverage index**. Top: music, 'standard' correlation. Middle: music, partial correlation. Bottom: speech, partial correlation.

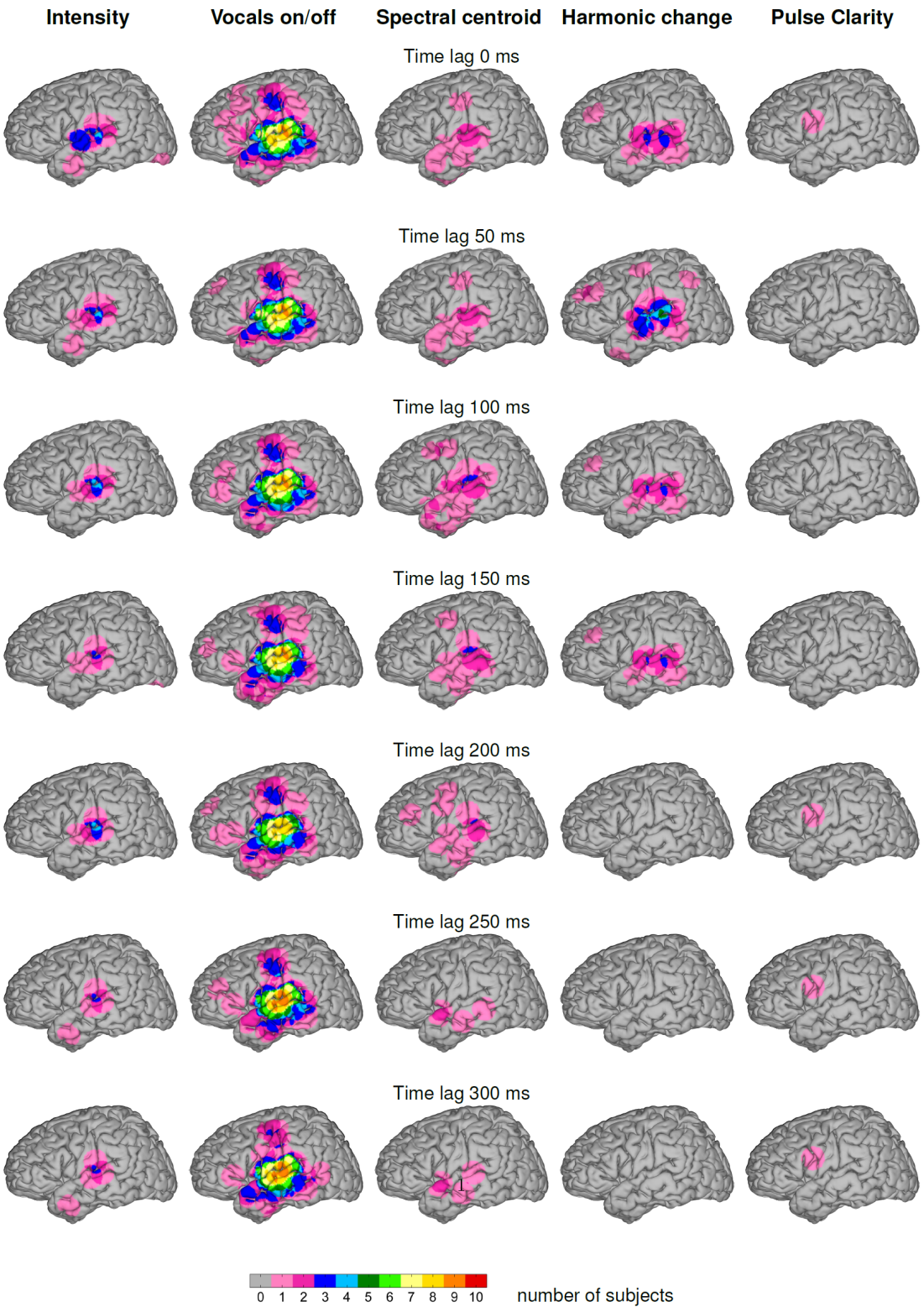

FIGURE 3.8: Number of participants with effects visualized on the MNI brain: Partial correlation coefficient for different time lags between stimulus and time course of ECoG high gamma power.

3.4 Summary and Discussion

In this chapter we have presented an approach that exemplarily demonstrates what limits of differentiation and specificity can be reached in an analysis situation where, on the one hand, electrocorticographic recordings offer high temporal resolution, a high level of spatial specificity and exquisite signal-to-noise-ratio, but, on the other hand, only a very small amount of (single-presentation) data is available to examine the processing of a complex naturalistic stimulus in detail. With partial correlation we have proposed a simple, but effective method that helps to delineate the cortical topography of the processing of aspects that make up a naturalistic music stimulus.

3.4.1 Neuroscientific results

The present study examined the unique relations between ECoG high-gamma band power and five features of music. Our results demonstrate that in this example of a rock song, the change between purely instrumental passages and those with vocal lyrics content is the compositional feature that exerts the most marked effect on the electrocorticographic brain response in the high-gamma frequency band. Furthermore, distinct cortical patterns of significant correlation with the features sound intensity, spectral centroid and harmonic change were present in single subjects.

The core region of high group-level overlap of significant correlation was located along the middle and posterior superior temporal gyrus (including Brodman areas 22, 41 and 42, see Figure 3.6, second row, second column). In three subjects, significant correlation was also present on the dorsal precentral cortex (BA 4) and in two subjects on the inferior frontal gyrus near Broca's area (BA 45). Considering that the partial correlation approach has removed the influence of the co-fluctuating four other factors, the remaining significant correlation could be related linguistic-semantic aspects of speech or to the presence of the human voice that has been found to effect ECoG gamma activity even for (unintelligible) reversed speech in Brown et al. (2013). The topography of the speech-related neural activity during listening to music is in line with Merrill et al. (2012), Brattico et al. (2011) and Sammler et al. (2010). Beyond the impact of vocals on/off, a specific reflection of the features spectral centroid and harmonic change is present on the STG, mostly on its posterior part. This particular area has been related to auditory processing in general, but specifically also to frequency (Liebenthal et al., 2003), pitch (Patterson et al., 2002), timbre (Alluri et al., 2012) and harmonic tones (Zatorre and Krumhansl, 2002).

A complementary analysis of ECoG recordings of stimulation with natural speech (without music) showed that the feature vocals on/off was reflected much less than in the music condition, while, contrastingly, reflections of sound intensity were consistently present on the middle part of the STG. This finding agrees with the essential role of the sound envelope in speech understanding that has been established by clinical results (Drullman et al., 1994, Lorenzi et al., 2006, Rosen, 1992, Zeng et al., 1999). It also suggests that, if speech-related content is embedded as song in music, the well-known impact of the sound envelope may be overruled by the change from instrumental to vocal/lyrics sound. Regarding the comparison between stimulus conditions (processing pure speech versus processing sung speech in instrumental music) this result demonstrates a difference in the relative importance of stimulus features. This is a finding that may be interpreted

as indicating the presence of different (global) listening modes for music with song and pure speech. At a more local level (within the music condition) our results seem to point in a similar direction: The present analysis included a broad range of musical features. Of all these (and also of those examined in preliminary stages) the feature that is reflected strongest in the brain response is the binary feature vocals on/off. This means that the feature that distinguishes predefined categories 'instrumental' and 'instrumental plus song' explains modulations of brain activity better than any direct acoustic property of the stimulus, such as, e.g., concomitant spectral contrasts that would have been captured by the feature Spectral Centroid. In general, this finding agrees with the long tradition of categorical perception of sound (Sundberg et al., 1991) or of listening modes (see Tuuri and Eerola (2012) for an overview). The present results provide an important new insight since they demonstrate that transitions between different listening modes occur even during one stimulus presentation. Specifically, our results may be interpreted as demonstrating that even the short vocals passages that are embedded in a rich acoustic arrangement suffice to put the brain in a mode that differs from the listening mode for instrumental music. This (putative) change of mode may be owed to the strong functional significance of speech-related content. Such an effect may be enhanced by the effort of understanding sung speech embedded in instrumental music.

The present results differentiate further the pioneering work of Potes et al. (2012) where ECoG high-gamma features were found to trace sound intensity in the posterior STG and in the inferior frontal gyrus. The present follow-up analysis helps to attribute this effect mainly to the presence of vocal speech in the stimulus and, with spectral centroid and harmonic change, we identified two further aspects specific for music that have an impact on the high-gamma ECoG response in some subjects. Notwithstanding, in these single subjects, these effects are highly significant and derived from one presentation of the stimulus. The present results complement those of Kubanek et al. (2013) where (based on the same data set and high-gamma ECoG activity) a specificity of STG and IFG for speech-related processing was suggested. The present results help to further elucidate the previous results in so far as they demonstrate that not only the sound envelope is encoded weaker in the music condition than for pure speech content, but that the alternating presence/absence of vocals is represented predominantly. To the growing corpus of highly heterogeneous evidence that sheds light on the neural processing of natural music, the present findings contribute that in addition to the alpha and theta frequency bands, which have been found to reflect dynamic rhythmic features of music (Cong et al., 2012), the high-gamma band is a suitable starting point for a differential analysis of the processing of higher-level auditory features. Furthermore, our findings emphasize the context-dependence of the processing of aspects of naturalistic auditory stimuli.

3.4.2 Methodology

In the present analysis, we approach a typical problem that arises when assessing the relationship between brain recordings and natural auditory stimuli. We address the problem of non-orthogonal features describing a natural auditory stimulus that complicates analyzing the relation between brain signals and stimulus features. Given the fact that this is an intrinsic problem when studying the relationship between brain signals and the structure of music, surprisingly few approaches in the literature deal with it. While in the domain of behavioral music psychology correlation coefficients for interrelated

features of music sometimes are reported 'as they are' for descriptive reasons (Farbood and Upham, 2013) (while acknowledging the problem), in the typically high-dimensional feature spaces of brain data appropriate measures of statistical significance are indispensable. In Alluri et al. (2012) principal component regression modeling has been proposed: an interrelated multi-dimensional description of a music signal is transformed into a lower-dimensional space of uncorrelated components. These are subsequently perceptually evaluated and result in a orthogonal set of music descriptors with perceptual labels. Although this method was applied successfully in a series of investigations (Alluri et al., 2012, Toiviainen et al., 2014), it is not clear whether it achieves interpretable results with other stimuli. Besides that, it requires additional experimental effort for the perceptual validation.

Here, operating on the original features, we demonstrate that partial correlation as a simple post-hoc approach provides a sensitive method to identify highly specific indices of the processing of auditory information. Partial correlation takes advantage of the fact that in naturalistic music the correlation between different music features varies during the course of a stimulus. If partial correlation is viewed as correlation between residuals (see 2.4.7), the present approach can be understood as identifying unique variance of a music feature in the time course of the stimulus and, then, probing the presence of a correlated brain response. In the present context, owing to the ECoG's characteristics of offering both high temporal and spatial resolution, this achieves an extreme level of specificity. In contrast to numerous approaches that assume inter-individual consistent spatial distribution of neural activity, e.g., by averaging the EEG time course across subjects (Schaefer et al., 2009) or by selecting components that are common to the majority of subjects (Alluri et al., 2012, Cong et al., 2012) it operates on single-subject ECoG recordings of single stimulus presentations. Moreover, it differentiates between single aspects of the music signal. The proposed analysis scheme is efficient and applicable in scenarios with naturalistic auditory stimulation, and, in principle, also for averaged data.

3.4.3 Limitations

Obviously, there are limitations of what can be achieved with this approach. First of all, there are the general limitations of electrocorticography in non-clinical research, such as the fact that epilepsy patients may not be comparable in brain function and anatomy to the healthy population. Furthermore, the number of patients analyzed here is small and their grid coverage differed. Important issues, such as hemispheric specialization for speech and music, cannot be addressed with the present data set of left-hemispheric recordings. Furthermore, information about the patients' music preference, cultural background and musical training that could give valuable clues for interpreting interpersonal differences is not available in this follow-up analysis. However, our analysis is an example of what information can be gained within these limits and contributes to the growing body of methodological approaches for research on the processing of natural music.

Partial correlation, proposed here as one solution for inter-dependence of stimulus features, has detected specific reflections of music features in the ECoG high-gamma response. However, it has to be kept in mind that this method gives a *differential* picture of each music feature's impact on the brain response, not a comprehensive decomposition of the brain signal. It shows cortical reflections that are unique to this feature beyond

all others in the feature set. Thus, for a given feature, the portion of independent variance from the other features is crucial for the detectability of its reflection in the brain response.

Naturally, when comparing two different stimuli, such as in our case in the speech and music condition, the individual interdependence of stimulus features is not the same, nor can the stimulus features themselves be balanced between both stimuli. Our results, therefore, have to be regarded as highly specific cortical imprints of two different, naturally unbalanced examples of natural auditory stimulation from two sound categories, not as general findings on the processing of music or speech. Nonetheless, the present differentiated picture of brain responses at the level of single subjects and a single presentation is a valuable complement of the recent series of investigations in natural music processing research.

Chapter 4

From single notes to continuous music: extracting cortical responses to note onsets in music from the continuous EEG

In the previous chapter we have seen that electrocorticography allows to derive an extremely detailed picture of the cortical processing of aspects of music even in the case of single presentations of only a short segment of an original rock song. However, the availability of this kind of data is very limited. A much more widely-used recording modality is the scalp EEG, that, due to its relative low cost and effort, is a popular multi-purpose tool in auditory research. As described in Chapter 1, the signal-to-noise-ratio in EEG is considerably lower. Therefore, extracting stimulus-related activity requires more elaborate methods of data analysis.

In this chapter we present an approach that aims at extracting cortical responses to note onsets in music from the continuous EEG. We use the extracted continuous brain responses to assess the brain-stimulus synchronization with a measure we call Cortico-Acoustic Correlation (CACor).

The core idea of this approach is to (i) select a feature of the audio waveform that is reflected well by cortical evoked responses, and to (ii) train a spatio-temporal filter that maximizes the correlation between an extracted EEG projection and this music feature. This results in a one-dimensional EEG projection that tracks the note onset structure of music.

In the following chapter we first motivate our focus on onset responses in Section 4.1. In the same section, we give an overview on the state-of-the-art of related methods, in particular of methods from the speech processing domain where the reconstruction of the sound envelope is a popular task. After that, in Section 4.2 we describe the steps of analysis and provide scenarios for application.

4.1 Brain responses to note onsets

Music has been defined as 'organizing sound in time' by modernist composer Edgar Varèse (Varèse and Chou, 1966). From a general physiological perspective the strong timing mechanisms of music have been recognized to critically contribute to music being a highly 'mediating stimulus' that engages human behavior and brain function in multiple ways (Thaut, 2005). The low-level elements of the rhythmic structure in music are distinct auditory events, such as note onsets. These serve as acoustic landmarks that provide auditory cues that underlie the perception of more complex phenomena such as beat, rhythm, and meter (Jackendoff and Lerdahl F., 2006).

Single sound events with a distinct onset are known to elicit evoked responses in the brain, such as the P1-N1-P2 complex (Näätänen and Picton, 1987). Thus, music can be hypothesized to also organize our brain activity in time: a sequence of tone onsets in music can be assumed to be echoed by a sequence of event-related potentials (ERPs) in the brain signal.

The P1-N1-P2 complex is a sequence of 'obligatory' auditory ERPs that index the detection of sudden changes in the energy or in spectral properties of the auditory input (Näätänen and Picton, 1987, Winkler et al., 2009). P1, N1 and P2 are assumed to reflect different neural generators and functional processes, but typically occur together (Crowley and Colrain, 2004). The P1-N1-P2 complex has been found to be susceptible to changes in a variety of aspects of auditory stimulation and, in general, is a sensitive tool for gaining insights into auditory processes. In the music domain onset-related ERP responses have been studied in numerous contexts, e.g., to investigate aspects of timbre perception (Meyer et al., 2006, Shahin et al., 2005) and rhythm (Schaefer et al., 2009). In the domain of speech processing the P1-N1-P2 has been found to reflect many of the spectral and temporal cues contained in spoken language Hertrich et al. (2012), Ostroff et al. (1998), Whiting et al. (1998), Woods and Elmasian (1986). Moreover, the P1-N1-P2 complex as a central auditory representation of sound has been linked to behavioral function, as reflected by speech comprehension, lexical ability and auditory discrimination (Sabisch et al., 2006, Santos et al., 2007, Tremblay et al., 2001, Yoncheva et al., 2014). Importantly, the P1-N1-P2 complex has been found to be influenced by subject-individual variables, such as maturation, learning and memory (Baumann et al., 2008, Fujioka et al., 2006, Shahin et al., 2003, 2008, 2010, Trainor et al., 2002, Tremblay and Kraus, 2002, Tremblay et al., 2001, 2014), and to situational factors, such as attention or arousal (Tremblay et al., 2001). These properties have established the P1-N1-P2 complex as a versatile tool in a wide range of applications in clinical and non-clinical contexts (Billings et al., 2011, Campbell et al., 2011, Martin et al., 2008, 2007).

Taken together, note onsets are basic constituents of the musical surface (Lerdahl and Jackendoff, 1983) and seem to contribute essentially to the effect of music. They are well reflected in the EEG and, in general, related ERPs, such as P1-N1-P2 complex, seem to be sensitive to a range of acoustic, user-related or situational factors. For examining the relations between stimulus structure, brain signal and experiential aspects of music, note onsets, thus, can be considered as a good starting point.

The classical way to examine cortical onset responses, such as the P1-N1-P2 complex, is to present a high number of identical stimuli and subsequently to average across these presentations, thereby enhancing the signal-to-noise ratio in order to make ERPs visible. Obviously, this technique puts constraints on the complexity and duration of the

material that can be presented and is of limited use in the case of full-length pieces of naturalistic music. Beyond averaging techniques, few attempts have been made to track the processing of a complex naturalistic music stimulus in the continuous EEG (Cong et al., 2012, Schaefer et al., 2011a, Thompson, 2013). In general, these approaches identify stimulus-related components of brain signals by looking for consistently occurring patterns/discriminating features and, subsequently, relate these to a range of music features. They, therefore, can be thought to be 'unsupervised' with respect to assumptions about contributing aspects of music.

In a classification approach Schaefer et al. (2011a) decoded to which of seven 3s-fragments of original music participants were listening to from their EEG. Using Singular Value Decomposition (SVD) they identified a N1-like EEG component from the Grand Average EEG as most discriminative component that was correlated with the sound envelope. Cong et al. (2012) and Thompson (2013) applied ICA-related techniques to extract a general mapping between musical audio and the subject's electrical brain activity. Their results revealed that EEG components predicting the envelope of the stimulus waveform are a common feature in a set of EEG recordings of either a group of subjects that listens to the same piece of music (Cong et al., 2012) or of one subject that listens to a variety of musical excerpts (Thompson, 2013). These results support the idea that the sound envelope (that contains information about the onset structure) consistently is reflected in the EEG.

In the domain of speech processing cortical onset responses that reflect changes in the waveform envelope (termed Envelope Following Responses, EFRs), have been a target of interest for a long time (Aiken and Picton, 2006, Kuwada et al., 1986, Purcell et al., 2004).

Approaches for finding a mapping between brain response and stimulus envelope can be distinguished into *forward* mapping approaches and *backward* mapping approaches. Common to both is that the mapping between stimulus envelope and brain signal is 'learnt' via a least squares solution that corresponds to linear regression (see Chapter 2). *Forward* mapping approaches learn the convolution model that transforms the envelope of the input signal into the measured brain signal by minimizing the error between estimated EEG signal and true EEG signal. This, usually, is done for single sensor signals or after reducing the multivariate brain signal to a single source. The resulting convolution model can be interpreted as an estimation of the impulse response of the auditory system. Forward mapping approaches have been applied in Aiken and Picton (2008), Ding and Simon (2012b), Kerlin et al. (2010), Koskinen et al. (2013), Lalor et al. (2009), Mesgarani and Chang (2012), Power et al. (2012, 2011) and Golumbic et al. (2013).

Backwards mapping approaches work in the reverse direction and learn a mapping from the neural data back to the stimulus. The learned model, also termed 'spatio-temporal decoder' (O'Sullivan et al., 2014) represents the transformation of the sensor level measurements to the stimulus' envelope, i.e., it 'learns' how each sensor contributes to the stimulus reconstruction. Furthermore, there exist techniques to additionally take into account and optimize a time lag between stimulus and brain response. The simple linear regression models can be further extended to include (non-linear) models of the auditory system, such as the characteristics of the basilar membrane (Biesmans et al., 2015). *Backwards* mapping approaches were first applied to reconstruct the envelope of various simple and complex sound features from the invasive brain recordings of animals (Mesgarani et al., 2009, Rieke et al., 1995, Stanley et al., 1999). From human invasive brain

recordings (ECoG) highly precise reconstructions of natural speech were obtained with backwards mapping techniques (Golumbic et al., 2013, Mesgarani and Chang, 2012, Pasley et al., 2012). Using magnetoencephalography (MEG) this approach has been shown to be highly sensitive to selective attention in a multi-speaker environment (Ding and Simon, 2012a) and even in single-trial EEG (O'Sullivan et al., 2014).

However, an approach that is specifically dedicated to directly extracting onset responses to natural music stimuli from the continuous EEG has (to our knowledge) not been proposed yet. Therefore, it is not known what can be gained from neural accounts of stimulus processing that are similarly detailed as those derived in the aforementioned studies from the speech processing domain.

One goal of this thesis is to probe whether the idea of a sequence of onset with a cortical echo holds for complex naturalistic stimuli, such that an EEG component can be extracted that follows the sequence of onsets that constitutes a piece of music. If so, in a second step we explore whether this technique is a tool for systematically investigating the physiological/cortical reflection of music in relation to behavioral measures of conscious experiencing music.

In the following we propose a method to obtain a representation from the ongoing EEG that reflects brain responses to the sequence of tone onsets that constitutes a natural piece of music. The core idea is to find a backward model that reduces multi-channel EEG measurements to that component of the signal that 'follows' the stimulus' sequence of onsets. This is done by regressing the EEG signal onto a target function extracted from the audio signal that represents the onsets best. The workflow (summarized in Figure 4.1) can be divided into four modules for (i) preprocessing EEG and audio data, (ii) calculating spatio-temporal regression filters for optimally extracting EEG features, (iii) applying the derived filters to new data in order to extract EEG projections, and (iv) transforming the spatio-temporal filters into a representation suitable for neurophysiological interpretation. The extracted EEG projections in (iii) can be utilized to examine brain-stimulus synchronization as described in Chapter 4.2.4. A description of the analysis steps in pseudocode is given in Figure 4.2 at the end of Section 4.2.3.

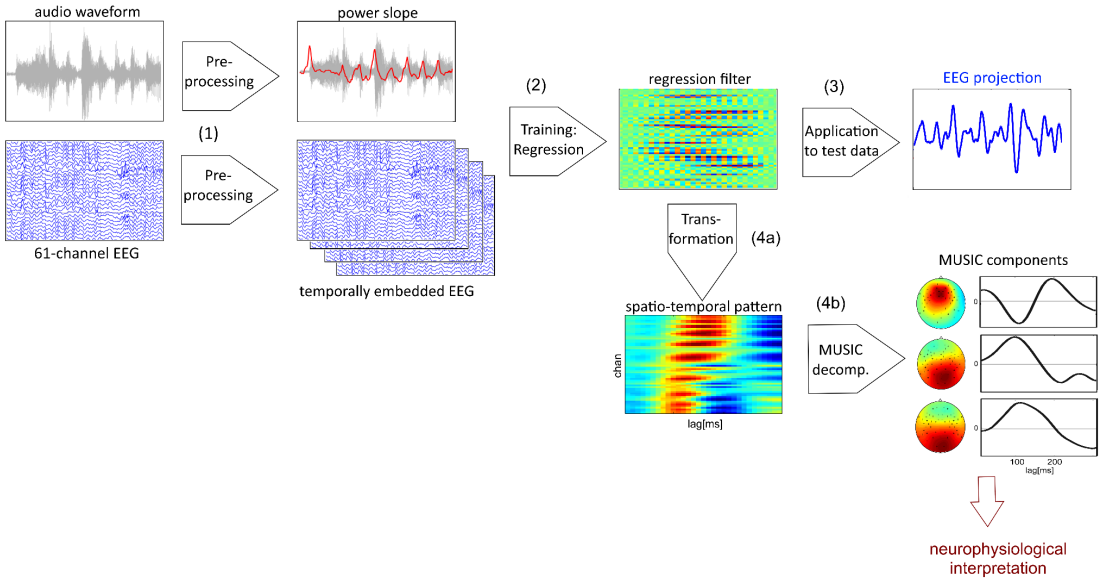

FIGURE 4.1: EEG feature extraction. (1) In the first step of the analysis the 61-channel EEG signal (after generic preprocessing) is temporally embedded and the power slope of the audio signal is extracted. In the training step (2) the embedded EEG features are regressed onto the audio power slope (Ridge Regression). After that (3) the resulting spatio-temporal filter (regression weight matrix) reducing the multichannel EEG to a one-dimensional projection is applied to a new presentation of the same stimulus. The regression filter can be transformed (4a) into a spatio-temporal pattern that indicates the distribution of information which is relevant for the reconstruction of the audio power slope. This spatio-temporal pattern, in turn, can be (4b) decomposed into components (derived with the MUSIC-algorithm) which have a scalp topography and a temporal signature. The EEG projections obtained in (3) subsequently are examined with respect to Cortico-Acoustic correlation (CACor).

4.2 Analysis pipeline

4.2.1 Preprocessing

4.2.1.1 EEG

A priori, it is not clear by how much the EEG response we are interested in lags behind the presented stimulus. Therefore, a temporal embedding of the EEG signal is performed, as described in Section 2.4.4 before training the regression model. For examining cortical onset responses, such as P1-N1-P2 which is a so-called 'midlatency range' component, brain responses within a latency of 0 to 300 ms can be considered. For 61-channel EEG recordings this results in a data matrix $\boldsymbol{X_{1891\times T}}$ corresponding to $61 \cdot 31 = 1891$ 'channels' and T observations (sampled in steps of 10 ms).

4.2.1.2 Audio data

For an optimal extraction of onset-related cortical activity it is crucial to identify the target function that captures note onsets best. In general, note onsets are indicated by amplitude modulations of the sound envelope, therefore related approaches often use the sound envelope or its logarithm (Lalor et al., 2009, Power et al., 2012, 2011). According to our experience, however, the *first derivative* of the sound envelope (in the following denoted 'power slope') represents best the intensity changes that are expected to trigger ERP responses. This choice of a target function can be motivated by the brain's sensitivity to change and has been supported by the importance of 'auditory edges', both for the perception of 'attack' of musical sounds (Gordon, 1987) and also for speech understanding (Doelling et al., 2014).

For an audio waveform with a typical audio sampling rate of 44.1 kHz the audio power slope is extracted by segmenting the audio signal into 50% overlapping time windows of 50 ms width and then calculating the average power of each window. Subsequently, the resulting time course is smoothed using a Gaussian filter of three samples width and the first derivative is taken, yielding the power slope. The extracted power slope is then re-sampled to match the sampling frequency of the EEG.

4.2.2 Training of spatio-temporal filters

To learn the linear relationship between brain signal and stimulus we train a Linear Ridge Regression model that is described in Chapter 2.4.1 to optimize the correlation between the temporally embedded EEG and the power slope of the audio signal. To avoid overfitting this is done in a cross-validation approach 2.4.5, where training and evaluation of the model are performed on separate portions of data. Specifically, in the training step a filter $\boldsymbol{w}$ is learned that optimizes the correlation between the filtered (embedded) EEG and the audio power slope z. Subsequently, this filter $\boldsymbol{w}$ is applied to the (embedded) EEG features of the test set, resulting in a one-dimensional EEG projection. This EEG projection is evaluated with respect to its correlation with the power slope of the test data. We call the resulting correlation coefficient Cortico-Acoustic Correlation (CACor).

In general terms this can be described as learning a linear mapping between stimulus and brain response that is assumed to be relatively stable within the training data. The training step produces a spatio-temporal weight matrix that indicates how each of the 'channels' of the embedded data contributes to predicting the target function optimally at each time lag. In the evaluation step this weight matrix is applied to new EEG data. The result is a time course that predicts the target function of the new data. The goodness-of-fit between the predicted target function and the true target function indicates whether the assumption of a generalized mapping between EEG and stimulus holds.

The core question this approach attempts to answer is whether a learned brain signal-stimulus mapping generalizes to new data. Depending on the design of the leave-one-out-pattern of cross-validation this question can be posed in different flavors:

- Within-presentation cross-validation (where the stimulus was divided into segments) indicates the presence of a stimulus-general mapping in face of a (typically) physically varying sound pattern.
- Between presentation cross-validation (training and testing on several presentations of the same stimulus) emphasizes a consistent (situation-independent) reflection across presentations.
- Between-subjects cross-validation (training and testing on presentations of the same stimulus, but of different subjects) attempts to abstract from subject-individual differences between mappings.
- Between-stimulus cross-validation (training and testing on presentations of different stimuli) is interested in very general processing of music, independent of specific stimulus characteristics.

4.2.3 Visualization of patterns/neural sources: extension of Multiple Signal Classification (MUSIC)

The spatio-temporal regression filters that result from the training step described in Section 4.2.2 are matrices of the dimensionality N channels $\times T$ time lags which correspond to a temporal sequence of spatial filter topographies. However, these filters are not appropriate for neurophysiological interpretation. Instead, the filter maps need to be transformed into patterns, which specify how the activity of the respective sources in the brain is projected on the scalp (forward linear model of EEG, see (Chapter 2, Section 2.3.2)). These patterns allow to examine how the information that is used to reconstruct the power slope of the stimulus is distributed in space (on the scalp) and time (relative to the stimulus).

Interpreting the matrix representation of this information is not very intuitive. Therefore, it would be desirable to distill a representation from the spatio-temporal regression patterns that is reduced in dimensionality and that has a form that allows for a better comparison with conventional ERP components, e.g., consisting of one or more 'components' with a spatial pattern and a time course. A standard method for examining the characteristics of a matrix is to examine its eigenvectors and -values. Eigenvectors, however, are orthogonal to each other. It is, however, questionable that neural sources and their physiological patterns should obey such an orthogonality constraint. Therefore, the

direct result of such a factorization will hardly yield a good estimate of related ERP-like components.

In the following we propose an approach that applies and extends the Multiple Signal Classification (MUSIC) algorithm (Mosher et al., 1992) that is described in Chapter 2.4.6 in more detail in order to derive set of not (necessary) de-correlated spatial and temporal patterns that represent the regression patterns.

In a nutshell, the MUSIC algorithm factorizes the spatio-temporal pattern matrices using Singular Value Decomposition (SVD) and reduces them to a subspace covering a predefined portion of the variance. A least-squares-based scanning procedure is performed to find dipoles that have, according to a multi-shell spherical head model, produced these patterns. This results in a set of (potentially) non-orthogonal spatial MUSIC components for a set of orthogonal SVD patterns. Since we are interested not only in spatial patterns, but also in their temporal evolution within the range of time lags under consideration, we propose an additional step to the MUSIC algorithm that extracts time signatures corresponding to the spatial patterns in our specific scenario. This is explained in the following:

From a regression filter matrix W we derive a spatio-temporal pattern A_{orig} using Equation 2.3. Our goal is to extract a set of corresponding spatial and temporal components from the spatio-temporal pattern A_{orig} that describes A_{orig} in a physiologically interpretable way. The first step of the MUSIC algorithm is to reduce A_{orig} to A_{red} contained in a lower-dimensional subspace by performing singular value decomposition (SVD). The reduced version A_{red} can be represented in tyical SVD form

$$A_{red} = USV. \tag{4.1}$$

where U denotes the matrix of left-singular vectors, S a rectangular diagonal matrix containing the singular values and V the matrix of right-singular vectors. The left-singular vectors U can be multiplied by the singular values S and interpreted as (weighted) spatial patterns $A_s = U * S$ of A_{red}. The right-singular vectors V can be interpreted as temporal patterns A_t. Accordingly, the factorization of A can also be written as

$$A_{red} = A_s A_t. \tag{4.2}$$

The second step of the MUSIC algorithm finds a set of spatial patterns M_s that optimally represent the spatial dimensions of A_{red}. Therefore the spatio-temporal A_{red} can be approximated by the product of M_s with unknown time courses M_t (see Equation 4.3).

$$A_{red} \approx M_s M_t. \tag{4.3}$$

The MUSIC algorithm has extracted M_s by a linear transformation B from A_s (4.4), therefore A_{red} can be also be approximated by

$$A_{red} \approx A_s B M_t. \tag{4.4}$$

and

$$M_s = A_s B. \tag{4.5}$$

B can be determined by solving equation 4.5 which yields:

$$B = (A_s^\top A_s)^{-1} A_s^\top M_s. \tag{4.6}$$

Now, M_t can be calculated by setting equal 4.2 and 4.4 and by multiplying both sides with $B^{-1}(A_s^\top A_s)^{-1} A_s^\top$.

$$A_s A_t = A_s B M_t \tag{4.7}$$

$$B^{-1} A_t = M_t \tag{4.8}$$

M_t represents the temporal patterns of the non-orthogonal components M_s derived by the MUSIC algorithm.

Algorithm 1 Pseudocode for deriving a CACor coefficient from EEG data (training and test set) and the respective audio waveforms.

Require: EEG data: $X_1 \in N \times T_1$ (training data) , $X_2 \in N \times T_2$ (test data), after generic preprocessing as described in Section 4.2.1, Audio data: wavfiles y_1 and y_2 representing the respective stimuli, sampled at 44.1 kHz and in mono format, maxlag: maximal time lag of EEG with respect to stimulus in samples, e.g. $maxlag = 30$ for EEG data sampled at 100 Hz. 1. **Extract audio power slope**

for $i = 1$ **to** 2 **do**
  segment y_i into num_win windows of 50 ms.
  for $w = 1$ **to** num_win **do**
    $ps_i(w) \leftarrow$ power(data in window num_win)
  end for
  smooth ps_i with a gausswin of $width = 3$ samples
end for

1. **Temporal embedding of EEG**: Note that temporal embedding should only be applied to continuous data. If several epochs of EEG data are to be concatenated to form the training/test set, then temporal embedding has to performed on each epoch prior to concatenation.

for $i = 1$ **to** 2 **do**
  $X_{embi} = X_i(:, 1 : end - maxlag)$
  for $m = 1$ **to** $maxlag$ **do**
    $X_{embi} \leftarrow \begin{bmatrix} X_{embi} \\ X_1(:, m+1 : end - maxlag + m) \end{bmatrix}$
  end for
end for

3. **Train Ridge Regression model on training data**

$W \leftarrow$ ridge regression (X_{emb1}, ps_1)

4. **Apply Ridge Regression model on test data**

$\hat{ps_2} \leftarrow W^\top(X_{emb2})$

CACor $\leftarrow corr(\hat{ps_2}, ps_2)$

Algorithm 2 Pseudocode for visualizing patterns and sources with the extended MUSIC algorithm.

Require: Spatial filter $W \in N \times maxlag + 1$ obtained by training a Ridge Regression model, matrix of EEG data X.

1. **Derive pattern from filter.**

$A_{orig} \leftarrow C_{XX}W(W^\top C_{XX}W)^{-1}$

2. **Factorize A_{orig} and reduce its dimensionality**

$A_{orig} \leftarrow USV$ where S contains the eigenvalues on the diagonal

$A_{red} \leftarrow U\tilde{S}V$ where $\tilde{S}$ contains a reduced set of eigenvalues.

3. **Extract the spatial pattern M_s of a neural source with the MUSIC algorithm**

$M_s \leftarrow MUSIC(A_{red})$

4. **Extract a time course M_t for M_s.**

$B \leftarrow (A_s^\top A_s)^{-1}A_s^\top M_s$.

$M_t \leftarrow B^{-1}A_t$.

FIGURE 4.2: Algorithm 1: Pseudocode for calculation CACor coefficient. Algorithm 2: Pseudocode for calculating spatial and temporal MUSIC components from a regression filter.

4.2.4 Application scenarios: Cortico-Acoustic-Correlation (CACor)

The time course that is obtained by applying a trained regression model to new data is a projection of the original EEG and, consequently, has the format of a one-dimensional EEG channel. In the following we briefly introduce three basic analysis techniques that are based on these projections.

1. **Event-related analysis.** The extracted EEG projections can be treated in an analogue way to an 'ordinary' EEG channel, e.g., it can be subject to event-related analysis techniques of single or averaged onset events. Since the onset structure of the stimulus should be reflected with enhanced signal-to-noise ratio, this will mitigate the demand for a high number of stimulus repetitions that is typical for event-related analysis techniques.

2. **Global Cortico-Acoustic Correlation.** The correlation coefficient between the extracted EEG projection and the audio power slope of the stimulus, termed Cortico-Acoustic Correlation (CACor), can be regarded as a measure of brain-stimulus synchronization. Its significance can be determined in order to assess whether the onset structure of the stimulus is significantly reflected in the brain signal. In principle, CACor coefficients can be compared between presentations, subjects or stimuli.

3. **Time-resolved Cortico-Acoustic Correlation.** CACor can also be examined in a time-resolved manner by calculating the correlation between EEG projection and audio power slope not for the entire length of a stimuli, but for (overlapping) consecutive time windows. The resulting time course of CACor coefficients informs about the dynamics of the brain-stimulus synchronization within the presentation of a stimulus, an aspect that is particularly important in the context of naturalistic music.

4.2.5 Significance of CACor

When Cortico-Acoustic Correlation coefficients are interpreted, it is important to recognize that both, the EEG signal and the audio power slopes contain serial correlation. In Chapter 2.4.7 the two most widely used methods to correct for this have been described: Pyper er al.'s method for estimating the effective degrees of freedom (Pyper and Peterman, 1998) and a permutation testing approach with surrogate data as proposed in Theiler et al. (1992).

However, in the present context, both methods have their drawbacks. Pyper et al.'s method 'punishes' cross-correlation between two time series in order to avoid inflated correlation coefficients - a technique that also reduces the sensitivity. Permutation tests as proposed in Theiler et al. (1992), in principle, can be regarded as more sensitive, since they demonstrate that two signals with a given auto-correlation structure need a particular phase configuration to be correlated with a certain magnitude. However, in our previous experience permutation tests in combination with temporal embedding can be problematic: If the spectrum of the target function that is to be permuted is dominated by frequencies with cycles that are shorter than the time window for the embedding, a phase-scrambled version of this function can be learned equally well as the original, since

the temporal embedding allows to adjust for the phase scrambling. If this is the case, the test is fully insensitive.

The exact nature of the interplay between the window length for the embedding and the structure of the target function has not yet been investigated systematically. However, in the following Chapter we have included some first comparisons between both methods.

4.2.6 Discussion and Future Work

The most basic approach for examining the relationship of EEG signals with an external target variable is to determine the correlation between brain signal and target variable for each single sensor. The technique we have proposed here alleviates three typical problems of this basic approach: it enhances the low signal-to-noise ration of single-sensor measurements, it projects out effects of an uneven distribution of noise between sensors, and it does not require multiple testing corrections.

However, there exist alternative spatial/spatio-temporal filtering methods for finding multi-variate components in the EEG signals that reflect stimulus properties. A popular approach is Independent Component Analysis (ICA) (Ziehe et al., 2003, 2004, Ziehe and Müller, 1998, Ziehe et al., 2000) where EEG is decomposed into statistically independent components. ICA belongs to the family of Blind Source Separation techniques (BSS), since it decomposes the EEG signal guided only by statistical properties of the data. This means that a relationship with stimulus features can be established only post-hoc, as, e.g., in Cong et al. (2012). Therefore, ICA does not represent the first choice in the endeavour to integrate EEG data analysis and audio analysis.

In the context of decomposition methods for EEG data that integrate external (multimodal) information the technique of Source Power Co-Modulation (SPoC) has to be mentioned. In the most simple case SPoC derives spatial filters that relate power modulations of neural sources to a target function. SpoC has been demonstrated to be highly efficient in a number of neuroscientific studies (Dähne et al., 2013, 2015, 2014a,b), and, in general, would be highly suited for application in the present scenario. In the present, very specific, task of finding the brain sources that track tone onsets, however, preliminary analyses did not suggest that oscillatory power is an informative feature. Instead, the (raw) broadband EEG signal seemed to capture the 'landscape' of onsets that constitute a piece of music best. This suggests that, not power, but phase information is the neural phenomenon that is critical for following the rhythmic structure with high temporal precision. Yet, exploring whether SpoC can help to detect power modulations in the brain signal that relate to different aspects of music would be an interesting task for the future.

In the proposed method we integrate spatio-temporal EEG features, i.e., data points representing the full set of electrodes within a time window 300 ms after a given point in time, to predict the target function the audio power slope at this point. We do this by vectorizing the temporal dimension (time lag) of this data (see 'Temporal Embedding' Chapter 2, Section 2.4.4). Although this is a well-proven technique to extract correlated components with unknown delay in multi-modal settings (Bießmann et al., 2014), it has to be kept in mind, that the vectorization may destroy some of the spatio-temporal dependencies within the data. A better preservation of these structures can be achieved with a discriminative approach (Tomioka and Müller, 2010) where in a prediction task

spatial and temporal EEG filters are (truly) jointly optimized. Importantly, feature learning, feature selection, and feature combination are addressed through regularization. Different regularization strategies together with a visualization technique that is related to that described in Section 4.2.3 provide complementary views on the distribution of information relevant for the discrimination. A direct comparison of the method proposed in this thesis with the discriminative approach could give insights into the possible impact of the vectorization of spatio-temporal features in the future.

Chapter 5

Applications

In the previous chapter we have described a method to extract a neural 'echo' of a stimulus' sequence of note onsets from the continuous EEG. In the following we evaluate this method in applications that attempt to (i) directly compare how extracted neural signatures relate to ERPs that are obtained with conventional averaging techniques. Furthermore, we (ii) apply the proposed method to learn about of auditory stream segregation in the context of a multi-voiced semi-musical stimulus. Finally, we (iii) explore whether significant CACor can be detected in a naturalistic 'free listening paradigm' where participants listen to excerpts of recorded music without a specific task. We ask how the presence of significant CACor relates to stimulus properties and behavioral measures of experiencing music.

5.1 Experiments and data sets

This section gives an overview over the data sets used for experimental evaluation.

5.1.1 Dataset 1: The Music BCI

This data set contains EEG recordings from a study that proposed a 'musical' brain computer interface application (Treder et al., 2014) where participants listened to short clips of a complex semi-naturalistic, ensemble music stimulus. In the music clips of 40 s duration three musical instruments (drums, keyboard, and bass) were presented, each playing a (different) sequence of a repetitive standard pattern, interspersed by an infrequent deviant pattern. As an ensemble, the instruments produced a sequence resembling a minimalistic version of Depeche Mode's 'Just can't get enough' (1980s Electro Pop). The experiment consisted of presentations of the ensemble clip in which the instruments played together and solo clip presentations for each instrument. During the ensemble presentations participants were instructed to attend to a target instrument and to silently count the number of deviant patterns in this instrument.

5.1.1.1 Participants

Eleven participants (7 male, 4 female), aged 21-50 years (mean age 28), all but one right-handed, were paid to take part in the experiment. Participants gave written consent and the study was performed in accordance with the Declaration of Helsinki.

5.1.1.2 Apparatus

EEG was recorded at 1000 Hz, using a Brain Products (Munich, Germany) actiCAP active electrode system with 64 electrodes, placed according to the international 10-20 system. One of the 64 electrodes was used to measure electrooculogram (EOG). Active electrodes were referenced to left mastoid, using a forehead ground. Music stimuli were presented using Sennheiser PMX 200 headphones. The audio signal was recorded as an additional EEG channel.

5.1.1.3 Stimuli

Stimuli consist of 40-seconds music clips in 44.1 kHz mono WAV format, delivered binaurally, i.e., listeners were presented with the identical audio stream at each ear. The ensemble version of the clip is composed of three overlaid instruments, each repeating 21 times the respective bar-long sound pattern depicted in Figure 5.1, once in a while interrupted by a deviant bar-long pattern. In the following, the term 'single trial' denotes a single presentations of one of these 40s-long music clips. Deviants are defined by a single tone or a whole sequence of tones deviating from the standard pattern. The stimulus represents a minimalistic adaptation of the chorus of 'Just can't get enough' by the Synth-Pop band Depeche Mode. It features three instruments: drums consisting of kick drum, snare and hi-hat; a synthetic bass; and a keyboard equipped with a synthetic piano sound. The keyboard plays the main melody of the song. The relative loudness of the instruments has been set by one of the authors such that all instruments are roughly equally audible. The tempo is 130 beats-per-minute.

In the original experiment two different kinds of musical pieces were tested: in addition to the 'Just can't get enough' adaptation (music condition SP) a stimulus resembling a jazz-like minimalistic piece of music (music condition J) was presented. This jazz-like piece of music was in stereo format, i.e., left ear and right ear were stimulated with different streams. The present analysis focused on utilizing continuous onset-related brain responses for the investigation of stream segregation.Therefore, the jazz-like stereo stimulus which introduced additional spatial cues for stream segregation was not appropriate.

According to the pattern of standard and deviant, 10 different music clips were created with variable amounts and different positions of the deviants in each instrument. Additionally, solo versions with each of the instruments playing in isolation were generated.

5.1.1.4 Procedure

Participants were seated in a comfortable chair at a distance of about 60 cm from the screen. Instruction was given in both, written and verbal form. They were instructed to sit still, relax their muscles and try to minimize eye movements during the course of

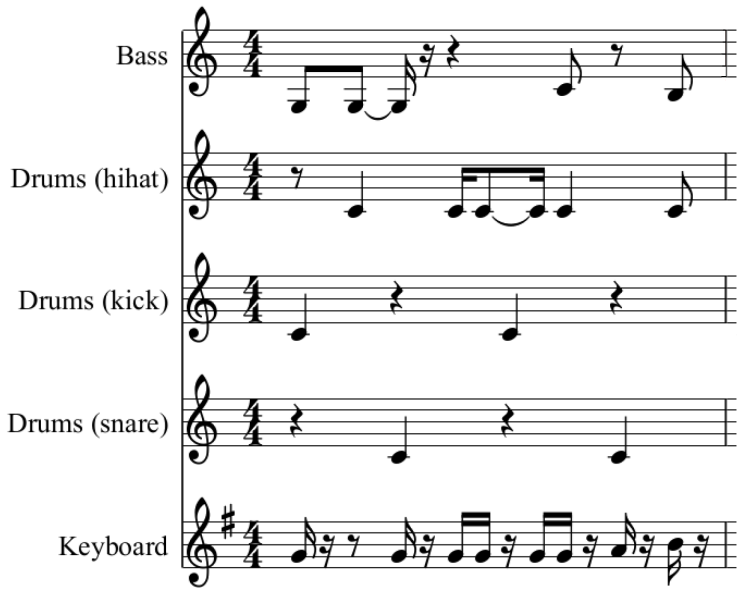

FIGURE 5.1: Score of ensemble version stimulus. Drums, although consisting of three instruments, are treated as one voice in the analysis. One (out of 63) music clips of 40 s duration consists of 21 repetitions of the depicted one-bar pattern. In addition, 14 solo clips were presented for each instrument.

a trial. Prior to the main experiment, participants were presented the different music stimuli and it was verified that they can recognize the deviants. The main experiment was split into 10 blocks and each block consisted of 21 40s-long music clips (containing 21 bars each). All clips in a block featured one music condition: Synth-Pop(SP), Jazz(J), Synth-Pop solo(SPS), or Jazz solo(JS). The solo clips were identical to the mixed clips except for featuring only one instrument. Within one block the 21 music clips were played according to a randomized playlist containing the ten clips that differed with respect to the position of deviant patterns. Each of the three instruments served as the cued instrument for 7 clips within a block. The music conditions were presented in an interleaved order as: SP, J, SPS, JS, SP, J, SPS, JS, SP, J. In other words, there were 3 blocks with ensemble presentations (= 63 clips, 21 for each target instrument) and 2 solo blocks (= 42 clips, 14 for each instrument) for each music condition; only conditions SP and SPS are part of the present analysis. Each trial started with a visual cue indicating the to-be-attended instrument. Then, the standard bar-long pattern and the deviant bar-long pattern of that particular instrument were played. Subsequently, a fixation cross was overlaid on the cue and after 2s, the music clip started. The cue and the fixation cross remained on the screen throughout the playback and participants were instructed to fixate the cross. To assure that participants deployed attention to the cued instrument, their task was to count the number of deviants in the cued instrument, ignoring the other two instruments. After the clip, a cue on the screen prompted participants to enter the count using the computer keyboard. After each block, they took a break of a few minutes.

5.1.1.5 Generic EEG pre-processing

The EEG data was lowpass-filtered using a Chebyshev filter (with passbands and stopbands of 42 Hz and 49 Hz, respectively) and then downsampled to 100 Hz. Since electrodes F9 and F10 were not contained in the head model used in the MUSIC algorithm (see 4.2.3) they were not considered in the analysis. Furthermore, the EOG channel was discarded. This left 61 channels. In order to remove signal components of non-neural origin, such as eye artifacts, muscle artifacts or movement artifacts while preserving the overall temporal structure of clips we separated the 61-channel EEG data into independent components using the TDSEP algorithm (Temporal Decorrelation source SEParation, Ziehe et al. (2003, 2004, 2000)). ICA components that were considered as purely or predominantly driven by artifacts based on visual inspection of power spectrum, time course and topography (see also McMenamin et al. (2011, 2010)) were discarded and the remaining components were projected back into the original sensor space.

5.1.2 Dataset 2: Listening to natural music

In this experiment data related to a set of nine music and non-music sound clips was obtained in three steps that are summarized in Figure 5.2. First, in a behavioral experiment 14 participants gave continuous ratings of perceived musical tension while listening to eight of the nine sound clips. Then, in an EEG experiment the brain signals of nine subjects were recorded while they listened three times to all nine stimuli. Finally, a set of nine acoustic/musical features was extracted from the waveform of each of the nine stimuli. The details for each of these procedures are given below.

5.1.2.1 Participants

Participants EEG experiment Nine participants (6 male, 3 female), aged 24-44 years (mean age 30), volunteered to take part in the experiment. All reported having normal hearing and no history of neurological disorder. The subjects differed with respect to their musical education and practice: two of them reported intensive musical training of more than 15 years and on more than one instrument, five of them modest amounts of musical training (mean: 7 years) and two of them no musical training beyond obligatory lessons at school. Subjects completed a questionnaire about their musical activities and preferences. Participants gave written consent and the study was performed in accordance with the Declaration of Helsinki. The study protocol was approved by the Ethics Committee of the Charité University Medicine Berlin.

Participants behavioral experiment In a separate experiment 14 new participants (9 male, 5 female, 12 right-handed, 2 left-handed) volunteered to take part in the behavioral experiment. Their musical experience ranged from no musical training beyond obligatory courses at school to more than 15 years of musical training (mean 6 years).

5.1.2.2 Apparatus

Brain activity was recorded with multi-channel EEG amplifiers (BrainAmp hardware, BrainProducts, Germany) using 63 Ag/AgCl electrodes (mounted on a Fast'n'Easy cap,

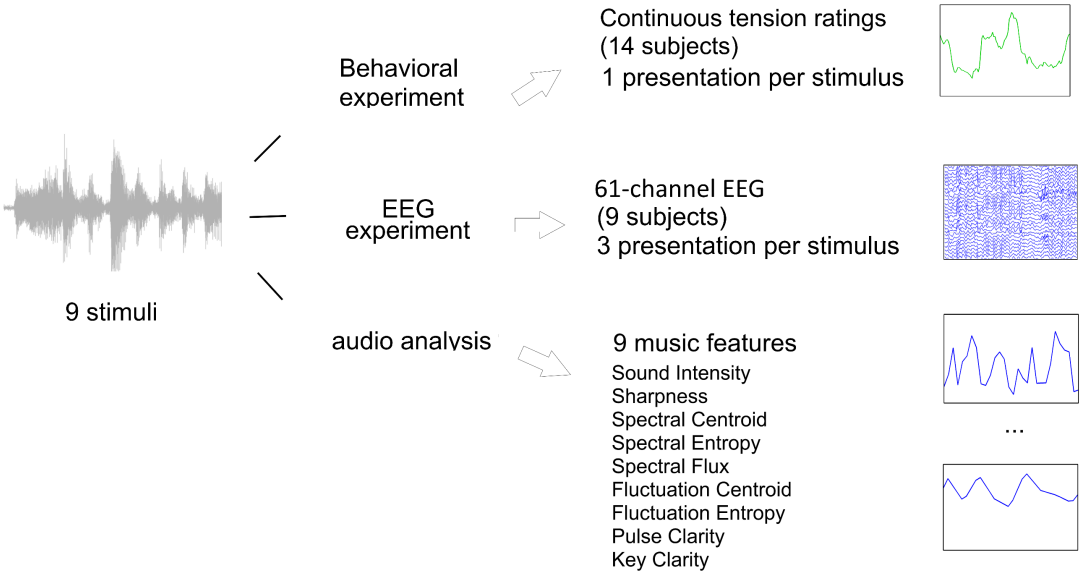

FIGURE 5.2: Overview of experiments and data types. In the present study data related to the processing of a set of music and non-music sounds was obtained in three steps. In a behavioral experiment 14 participants gave continuous tension ratings while listening. This resulted in one time course of tension ratings for each stimulus and subject. In an EEG experiment the brain signals of nine subjects were recorded while they listened to the stimuli. Stimuli were repeated three times for each stimulus, resulting in 27 EEG recordings for each stimulus. From the waveform of each stimulus a set of nine acoustic/musical features was extracted from each stimulus.

Easycap, Germany) in an extended 10-20 system sampled at 1000 Hz with a band-pass from 0.05 to 200 Hz. All skin-electrode impedances were kept below 20 kΩ. Additionally, horizontal and vertical electrooculograms (EOG) were recorded. Signals of scalp electrodes were referenced against a nose electrode. Music stimuli were presented in mono format using Sennheiser PMX 200 headphones. The audio signal was recorded as an additional EEG channel for accurate synchronization.

In the behavioral experiment tension ratings were recorded using a custom-made joystick that was operated with the thumb of the dominant hand. A small spring integrated into the joystick allowed to indicate the build-up of tension by pushing the joystick upwards and decreasing tension by releasing the joystick. The joystick position was sampled at 50 Hz.

5.1.2.3 Stimuli

Nine stimuli from different sound categories were presented in the experiment (see Table 5.1 for more detailed information): (1) Badinerie by J.S. Bach, (2) The Four Seasons, Spring, by A. Vivaldi, (3) The Four Seasons, Summer, by A. Vivaldi , (4) Etude op. 12, No. 10, by F. Chopin, (5) Prelude op. 32, No. 5, by S. Rachmaninov, (6) Theme of Schindler's List, by J. Williams, (7) an isochronous sequence of major triad chords with root notes on all tones of the chromatic scale (chord duration 350 ms, including rise and fall times of 17.5 ms, interonset interval (IOI) 420 ms, after 7-11 repetitions of a chord change to a new chord in random manner), (8) Jungle noise and (9) instrumental noise.

name	composer	title	dur.	Recording/source	description
Chord sequence			3' 8"	generated in Matlab	Sequence of major triad chords on all tones of the chromatic scale. After 7-11 repetitions of a chord change to a new chord in random manner
Bach	J.S.Bach	Badinerie	1' 20"	T. Koopman Amsterdam Baroque Orchestra Image Ent.,1995	From Orchestral Suite No. 2 BWV 1067 for Flute, strings and basso continuo
Vivaldi, Spring	A. Vivaldi	The Four Seasons, op. 8 Spring	3' 12"	P. Schoeman, 2010 London Philharmonic Orchestra X5 Music Group	Concerto for Violin and Orchestra
Vivaldi, Summer	A. Vivaldi	The Four Seasons, op. 8 Summer	2' 49"	A.Loveday, Acad. of St. Martin-in-the-fields Decca, 1969	Concerto for Violin and Orchestra
Chopin	F. Chopin	Etude op. 12, No. 10	2' 39"	V. Horowitz Sony 1997	Piano solo
Rachmaninov	S.Rachmaninov	Prelude op. 32, No. 5	3' 44"	E. Gilels, live recording, Moscow, 1967	Piano solo
Williams	J. Williams	Theme of Schindler's List	3' 31"	I. Perlman, MCA Records, 1994	Violin and Orchestra
Orchestra			1' 04"	http://www.youtube.com/ watch?v=IslDWrmOieE	Sound of an orchestra tuning in before a performance
Jungle			1'11"	http://www.youtube.com/ watch?&feature=related v=vli6cmmvXkg	Animal/nature sounds from the jungle

TABLE 5.1: Dataset 2: Stimuli

5.1.2.4 Procedure

EEG experiment The study aimed at approximating listening situations that resemble those of everyday life. Consequently, we pursued a 'free listening paradigm': participants were not given any special task and were just asked to listen in a relaxed manner and with closed eyes during the presentation of the musical stimuli. The main experiment was split into three blocks. In each block the nine stimuli were presented in a different order that was designed such that two piano pieces, two violin pieces or two non-musical stimuli never occurred in direct succession.

Behavioral experiment In the behavioral experiment subjects were given a short introduction to the concept of tension in music. Then, they were instructed as following:

"In the following you listen to eight excerpts, divided by short breaks. Six of them are musical pieces, two are non-musical. Your task is to indicate continuously with the joystick how you experience the evolution of tension of each piece of music. Please start each piece with the joystick at zero position. If you experience an increase in tension in the music, move the joystick up. If you experience a decrease in tension, release the joystick towards the zero position. You will have the opportunity to practice this before the beginning of the experiment. These tension ratings reflect your individual experience. Therefore it is not possible to do this right or wrong."

After one practice trial with a different music stimulus, each music stimulus was presented once while joystick movements were recorded. Since the stimulus *Chord sequence*, due to its simplicity and repetitiveness, was not assumed to give rise to the perception of tension in the listeners it was not part of the behavioral experiment.

5.1.2.5 Audio analysis

Power slope For each stimulus the power slope was determined by segmenting the audio signal into 50% overlapping time frames of 50 ms width and then calculating the average power of each window. Subsequently, the resulting time course was smoothed using a Gaussian filter of three samples width and the first derivative was taken. Then, the extracted power slope was interpolated to match the sampling frequency of the EEG.

Extraction of musical features We chose a set of nine musical features that cover a broad spectrum of timbral, tonal, and rhythmic categories of music. Sound intensity, which can be considered as an approximate measure of loudness, is a stimulus feature that influences a variety of brain responses (Doelling et al., 2014, Mulert et al., 2005, Näätänen and Picton, 1987). Sharpness, defined as the mean positive first derivative of the waveform envelope (Doelling et al., 2014), has been found to be an important cue for cortical tracking of the speech envelope (Ding and Simon, 2014, Doelling et al., 2014). Furthermore, with Spectral centroid, Spectral entropy and Spectral flux we included a set of three spectral features that describe pitch- and timbre-related aspects of sounds. The fluctuation spectrum of an audio signal contains the periodicities contained in a sound wave's envelope. For rhythmically regular musical sounds peaks in the fluctuation spectrum correspond to beat-related frequencies. The fluctuation spectrum can be further characterized by the Fluctuation centroid that indicates where the 'center of mass'

of the spectrum is and by Fluctuation entropy which is a measure of rhythmic complexity. Pulse clarity is a composite feature that indicates how easily listeners perceive the underlying rhythmic or metrical pulsation of a piece of music. It has been introduced and perceptually validated in Lartillot et al. (2008a) and since then has been used in numerous studies (Alluri et al., 2012, Burger et al., 2013, Eerola et al., 2009, Higuchi et al., 2011, Sturm et al., 2014, Zentner, 2010). Key clarity is a feature that estimates the salience of key.

Sound intensity and Sharpness were calculated in Matlab (The MathWorks Inc., Natick, Massachusetts). All other features were extracted using the MIRToolbox (Lartillot et al., 2008b). Sound intensity and the three spectral features were calculated for time frames of 50 ms overlapping by 50%. Sharpness, Fluctuation centroid and Fluctuation entropy, Pulse clarity and Key clarity were determined for time frames of 3 s with a 33% overlap. Additionally, for each stimulus a global description was obtained by taking the mean of each music feature in order to derive a rough estimation of the specific characteristics of each stimulus.

5.1.2.6 Generic preprocessing of EEG data

The EEG data was lowpass-filtered using a Chebyshev filter (with passbands and stopbands of 42 Hz and 49 Hz, respectively) and downsampled to 100 Hz. Since electrodes A1 and A2 were not contained in the head model used in the MUSIC algorithm (see Section 4.2.3) they were not considered in the analysis. In order to remove signal components of non-neural origin, such as eye, muscle or movement artifacts while preserving the overall temporal structure of the music-related EEG responses we separated the 61-channel EEG data into independent components using the TDSEP algorithm (Temporal Decorrelation source SEParation, Ziehe et al. (2004), with time lags of tau=0, ... 990 ms). ICA components that were considered as purely or predominantly driven by artifacts based on visual inspection of power spectrum, time course and topography (see also McMenamin et al. (2011, 2010)) were discarded and the remaining components were projected back into the original sensor space.

5.2 Studies/Analyses

5.2.1 Evaluation: Neural representation of note onsets: Extracted components versus averaging-derived ERPs

In Chapter 2.4.6 we have shown that an extended variant of the MUSIC algorithm transforms spatio-temporal regression filters into a set of scalp topographies and corresponding time signatures. It is, however, not clear how such neural representations relate to known ERPs that are derived by averaging techniques. Before we advance to naturalistic music scenarios with complex stimuli (where averaging techniques are not feasible) we apply our method in a case of a simple, repetitive stimulus that offers an opportunity for direct comparison with the results of conventional ERP analysis.

5.2.1.1 Methods

For this analysis EEG data from experiment 2 (Section 5.1.2) was used. For each subject EEG recordings for the three presentations of the stimulus *Chord Sequence* were concatenated. The stimulus *Chord Sequence* is an isochronous sequence of chords (for details see Section 5.1.2.3) that are equally structured (in terms of music theory) but have different root notes.

ERPs To obtain an ERP representation for tone onsets the continuous EEG data was segmented into epochs of 300 ms length, starting at tone onset, a baseline of -50 to 0 ms pre-stimulus was taken and the average was calculated. This technique cannot be applied to any other of the stimuli, owing to the high variability of onset characteristics in naturalistic sounds.

Regression/MUSIC components To obtain MUSIC components of the concatenated stimulus presentations for each subject, we performed a temporal embedding from $0, \ldots 250$ ms and trained a Ridge regression model (see Section 4.2.2) to predict the corresponding power slope. We reduced the spatio-temporal pattern obtained from the regression filters to a subspace containing 98% of the variance and derived three MUSIC components with a spatial and a temporal pattern each (see Section 4.2.3).

The resulting brain signatures were compared for single subjects.

5.2.1.2 Results

Panel (a) of Figure 5.3 shows the scalp topography and time course of the ERPs that were derived by classical averaging techniques from the EEG data of a single subject (S2) for the stimulus *Chord sequence*. The upper part shows the scalp topography of the time interval of 160-180 ms after chord onset. Below, the time course of EEG channel Fz for the first 300 ms after chord onset is depicted. In panel (b) the spatial (top) and temporal (bottom) dimension of one of the MUSIC components for the same subject are shown. Both scalp topographies show a positive fronto-central distribution. In the time course of the ERP a negative deflection at 100 ms is followed by a peak at 170 ms, while the temporal dimension of the MUSIC component is represented by a negativity at 110 ms that is followed by a positive peak at appr 200 ms. Panel (c) shows the Grand Average ERP scalp topography for the nine subjects along with respective time courses (grey) and the mean thereof (black). Panel (d) allows to directly compare the (z-score-transformed) time courses in (a) and (b).

5.2.1.3 Discussion

In this section we have compared averaging-derived ERPs and regression-derived MUSIC components for a simple stimulus consisting of isochronous chords. The application of the classical averaging technique produced a N1-P2 complex as a reaction to chord onsets, which is the expected auditory event-related potential indexing detection of the onset of auditory stimuli (Crowley and Colrain, 2004, Näätänen and Picton, 1987).

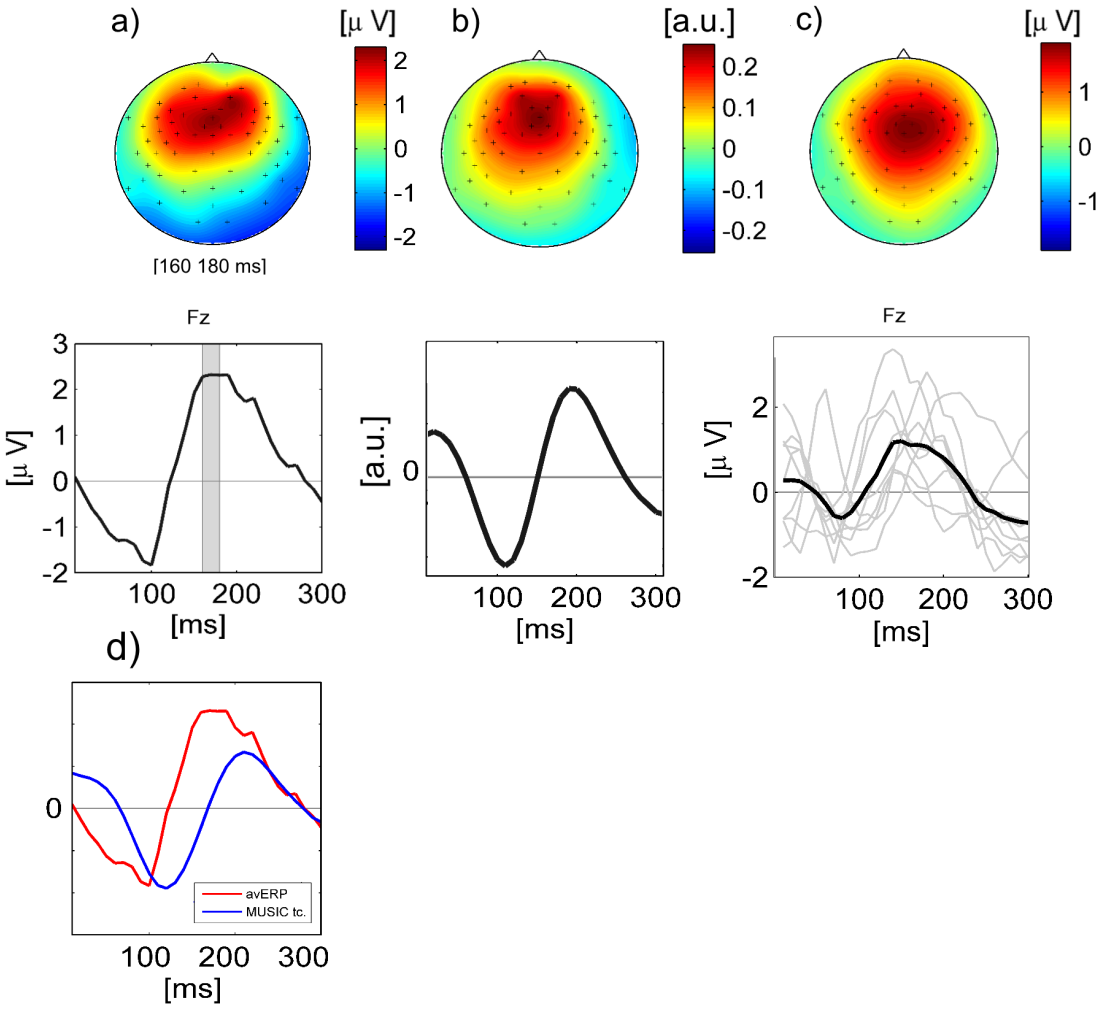

FIGURE 5.3: ERPs and MUSIC components. (a): Scalp topography (top) and time course (bottom) of ERP (single subject) derived by averaging channel Fz for all tone onsets of the stimulus *Chord sequence*. The scalp topography corresponds to the shaded interval of 160-180 ms in the time course. (b): Spatial (top) and temporal (bottom) dimension of MUSIC component derived from spatio-temporal patterns for the same subject. c) Top: Grand Average scalp topography obtained with classical ERP analysis. Individual scalp patterns that are contained in the Grand Average were determined by taking the mean scalp pattern across a 20 ms window that corresponds to the maximum amplitude of each subject's onset ERP at channel Fz. The time windows of maximum amplitude were determined by visual inspection and ranged between 130 and 250 ms after tone onset. The bottom part of (c) shows the individual time courses (grey) as well as the average time course (black). (d): ERP time course (red) and temporal dimension of MUSIC component (blue) from (a) and (b), z-score-transformed and shown in one plot for direct comparison.

The derived spatial and temporal MUSIC patterns resembled the N1-P2 signature, suggesting that the latency and the spatial distribution of information utilized in the optimization process can be related to the characteristics of typical onset-related event-related potentials. When interpreting these time courses, one has to recognize that they differ from averaged ERPs (even though they are on the same time scale), as they represent the weighting of the corresponding spatial component over time and, thus, rather resemble a convolution model or FIR filter than an ERP time course. Nonetheless,

time lags with large weights in principle can be compared to latencies of canonical ERP components.

The comparison of the time courses of MUSIC components and ERPs reveals that both have a similar shape, but also that negative and positive peaks occur approximately 20 ms later in the MUSIC component than in the ERP. If ERPs of tone onsets are averaged, the time point of the onset has to be defined, either in the experimental protocol or based on the audio waveform. Estimating with appropriate precision, however, the time point when the onset of a sound is perceived, is not a trivial task, in particular for natural sounds. This means that the absolute latency of averaging-derived ERP components cannot be regarded as very reliable information. In contrast, the latency of a MUSIC component seems to be more reliable in this context, as it has a fixed reference point.

In summary, our results demonstrate that EEG projections that 'follow' the power slope of the audio signal can be traced back to generic onset-related ERPs and, thus, represent a physiologically plausible fundament for further applications of this method.

5.2.2 Extracting the neural representation of tone onsets for separate voices of ensemble music

5.2.2.1 Introduction

Natural 'soundscapes' of everyday life, e.g., communication in a noisy environment, challenge our proficiency in organizing sounds into perceptually meaningful sequences. All the more music might spark our processing capabilities as it provides acoustic scenes with a large number of concurring sound sources. Yet, when listening to music we are able to organize the complex auditory scene into streams, segregate foreground and background, recognize voices, melodies, patterns, motifs, and switch our attention between different aspects of a piece of music. Auditory stream segregation (ASS), the perceptional process which underlies this capability, has fascinated researchers for many years, resulting in numerous studies exploring its mechanisms and determinants. In a nutshell (for a detailed review see Moore and Gockel (2002)), the segregation of a complex audio signal into streams can occur on the basis of many different acoustic cues (van Noorden, 1975); it is assumed to rely on processes at multiple levels of the auditory system; and it reflects a number of different processes, some of which are stimulus-driven while others are of more general cognitive nature, i.e., involving attention and/or knowledge Bregman (1994). Electrophysiological indices of auditory stream segregation have been detected in several approaches (Sussman, 2005, Sussman et al., 2007, Winkler et al., 2005, Yabe et al., 2001); for an overview see Snyder and Alain (2007)). One line of research focused on the Mismatch Negativity (MMN) as neural index for a distinct perceptional state of stream segregation by constructing tone sequences such that only a perceptual segregation into two streams would allow a MMN-generating sound pattern to emerge. Following a similar principle, neural steady-state responses were found to reflect the formation of separate streams (Chakalov et al., 2013) in MEG. Using EEG an influence of frequency separation of consecutive tones on the N1-P2 complex amplitudes was reported (Gutschalk et al., 2005, Snyder et al., 2006). Critically, this trend correlated with the perception of streaming in individual participants; a similar effect was reported for the N1 component. This suggests that the amplitude of early auditory ERP components like the N1-P2 complex can inform about the perceptional state with respect to segregation/-coherence of complex auditory stimuli. Since the N1-P2 complex as a sensory-obligatory

auditory-evoked potential can be utilized without imposing a complex structure, e.g., an oddball paradigm, on the stimulus material, it may be promising for investigating ASS in more naturalistic listening scenarios.

In the domain of speech processing the 'human cocktail party problem' represents a well-researched instance of ASS. In particular, cortical responses that follow the waveform envelope (termed Envelope Following Responses, EFRs) are widely used tool for investigating the neural representation of speech streams, and its modulation by attentional state (Aiken and Picton, 2008, Ding and Simon, 2012a, Golumbic et al., 2013, Kerlin et al., 2010, Lalor and Foxe, 2010, Lalor et al., 2009, Mesgarani and Chang, 2012, O'Sullivan et al., 2014). In the domain of music processing a marked reflection of the sound envelope has been detected in the EEG signal of short segments of naturalistic music (Schaefer et al., 2011a). Unsupervised approaches (Cong et al., 2012, Thompson, 2013) have confirmed that note onsets leave a reflection in the listener's EEG consistently across subjects and stimuli. However, these cortical reflections have not been investigated in detail for longer musical contexts and, in particular, an analogue to the'cocktail party' problem in speech processing has not been investigated specifically, even though composing music from several 'voices' is a common musical practice.

The N1-P2 response as a stimulus-driven sensory component varies as a function of the physical properties of the sound like its frequency (Dimitrijevic et al., 2008, Pratt et al., 2009) or spectral complexity (Maiste and Picton, 1989, Shahin et al., 2005). Considering these general characteristics, it is an interesting question whether in a music-related scenario where perception of separate streams is highly likely, Envelope Following Responses can be utilized to extract a neural representation related to these streams from the brain signal.

In principle, this task combines two so-called inverse problems that do not have a unique solution: (1) We have a number of sound sources that produce a mixed audio signal. From the mixed signal it is not possible (without further assumptions) to infer the original configuration of sources. This audio signal is assumed to result in stimulus-related neural activity in the listener. (2) What we record in the listener's EEG is a mixture of stimulus-related neural activity, unrelated neural activity, and non-cerebral noise. Inferring these sources from the EEG signal, the so-called inverse problem of EEG generation, is likewise a problem without unique solution.

In the present analysis we aim in a first step to learn a solution for the second of these inverse problems, to extract stimulus-related activity from the EEG in the case of a solo stream. Subsequently, we apply the derived solution in a scenario with mixed sound sources. We explore in how far the stimulus-related activity related to the solo stream can be extracted from the EEG of the mixed (ensemble or multi-voiced) presentation.

We re-analyze a data set from a study proposing a 'musical' brain computer interface application (Treder et al., 2014) that is described in Section 5.1.1. The original analysis showed that P3 ERP components to deviant patterns in the target instrument sufficiently differ from those in the non-target instruments to allow to decode from the EEG signal which of the instruments a subject is attending to. These results can be considered as a proof-of-concept that our capability of shifting attention to one voice in a multi-voiced stimulus may be exploited in order to create a novel music-affine stimulation approach for use in a brain-computer interface.

In contrast to the previous analysis that focused solely on P3 responses to deviations in the patterns, here, we propose to exploit the fact that *all* note onsets in a music clip should evoke ERP responses. Therefore, the sequence of onset events that constitutes each instrument's part should elicit a corresponding series of ERP events in the listener's EEG. Since onset characteristics critically contribute to an instrument's specific timbre (McAdams et al., 1995) and onset-triggered ERPs are known to be responsive to subtle spectral and temporal changes (Meyer et al., 2006) it can be assumed that the properties of this ERP response might differ for musical instruments with different tone onset characteristics. We extract this sequence of ERPs from the single-trial EEG by training a Linear Ridge Regression model with temporal embedding to optimize the relation between the power slope of the solo audio signal and the concomitant EEG. We (i) explore whether such a spatio-temporal filter obtains EEG projections from the solo-instrument trials that are significantly correlated with the sequence of onsets of the respective solo music clip; and we (ii) probe (by correlation measures) whether these filters trained on the solo trials can be used to reconstruct a representation of this solo voice from the EEG of participants listening to the ensemble version clips. Finally, we (iii) test whether the reconstruction quality increases if participants focus their attention on the respective instrument.

5.2.2.2 Methods[1]

After generic pre-processing (see Section 5.1.1) we performed a temporal embedding from $0, \ldots 250$ ms on the EEG data and extracted the power slope of all stimulus waveforms as described above (Section 5.1.2.5).

In the first stage of the analysis, regression filters that maximize the correlation between EEG and audio power slope were determined for the solo clips of the three instruments for each subject separately. In a leave-one-clip-out cross-validation approach (to avoid overfitting, see Section 2.4.5) a one-dimensional EEG projection for each of the 14 music clips for an instrument was derived. Then, the corresponding CACor coefficient and its significance were determined. Additionally, the CACor coefficient for the mean EEG projection and the audio power slope was determined for each subject and instrument, and also that for the Grand average of all EEG projections. Significance of CACor was determined with Pyper et al.'s method (see Section 2.4.7). In order to account for the repetitiveness of the music clips, we considered the cross-correlation for all possible time lags within a music clip which drastically reduced the effective degrees of freedom. The original and estimated effective degrees of freedom for the Grand Average correlation coefficients are given in the bottom line of Table 5.8. The correlation coefficients of the subject-individual mean EEG projections were corrected for multiple testing for N=11 subjects with a Bonferroni correction. Significance of correlation was determined to the level of $\alpha = 0.05$.

In the second stage of the analysis we applied the regression filters derived from the solo presentations to the EEG responses of the ensemble version stimuli. This was done for each subject and each instrument separately, resulting in three one-dimensional EEG projections for each ensemble version clip per subject. These projections were averaged

[1] In the three analysis scenarios described in Section 5.2.1, Section 5.2.2 and Section 5.2.3 the same regression method is applied. Therefore, parts of Section 5.2.1.1, Section 5.2.2.2 and Section 5.2.3.2 are very similar. The detailed descriptions of the regression analysis in each scenario were included for completeness and accuracy.

across the 63 ensemble version clips for each subject (separately for the instruments) as well as across all subjects (Grand Average).

To decompose the spatio-temporal regression patterns we concatenated all 14 solo clips for each instrument and subject, performed a temporal embedding from $0, \ldots 250$ ms, and trained a Ridge regression model to predict the corresponding power slope. We reduced the spatio-temporal pattern obtained from the regression filters to a subspace containing 98% of the variance and derived three MUSIC components with a spatial and a temporal pattern each.

5.2.2.3 Results

Solo stimulus presentations Figure 5.4 shows examples of the EEG projections that reconstruct the audio power slope; for illustration purposes these were collapsed across 11 subjects, 14 clips for each instrument and 21 bars in each clip. A comparison of the EEG-reconstructed power slope (red line) with the audio power slope (blue line) shows how the onset events in the audio signal are accompanied by peaks in the brain signal.

Table 5.2 gives the percentage of individual solo clips (14 for each instrument) with significant CACor. Note that this measure relates to the significance of single presentations of clips of 40 s duration and was derived without any averaging of EEG data. Table 5.3 shows the magnitude of correlation of the averaged EEG-projections (for the 14 solo presentations of each instrument) with the audio power slope for single subjects, revealing significant correlation in 7/11 subjects for drums, in 9/11 subjects for bass, and in 8/11 subjects for keyboard. The bottom line of Table 5.3 shows that taking the mean of all subjects' EEG projections (Table 5.3, bottom line 'GA') produces time courses that are significantly correlated with the original audio power slopes for all three instruments with magnitude of correlation r=0.60 for drums (p=0.00014, effective degrees of freedom: 34), r=0.52 for bass(r=0.52, p=0.00011, effective degrees of freedom: 48) and r=0.54 for keyboard (p=0.0000004, effective degrees of freedom: 72). Note that the original number of degrees of freedom of 3968 was drastically reduced by Pyper et al.'s method (Pyper and Peterman, 1998) that was applied to account for serial correlation in both time courses. All power slopes in Figure 5.4 are scaled for illustrative purposes. The absolute values of the audio power slopes for the three instruments are depicted in Figure 5.5, indicating differences in amplitudes and rise times.

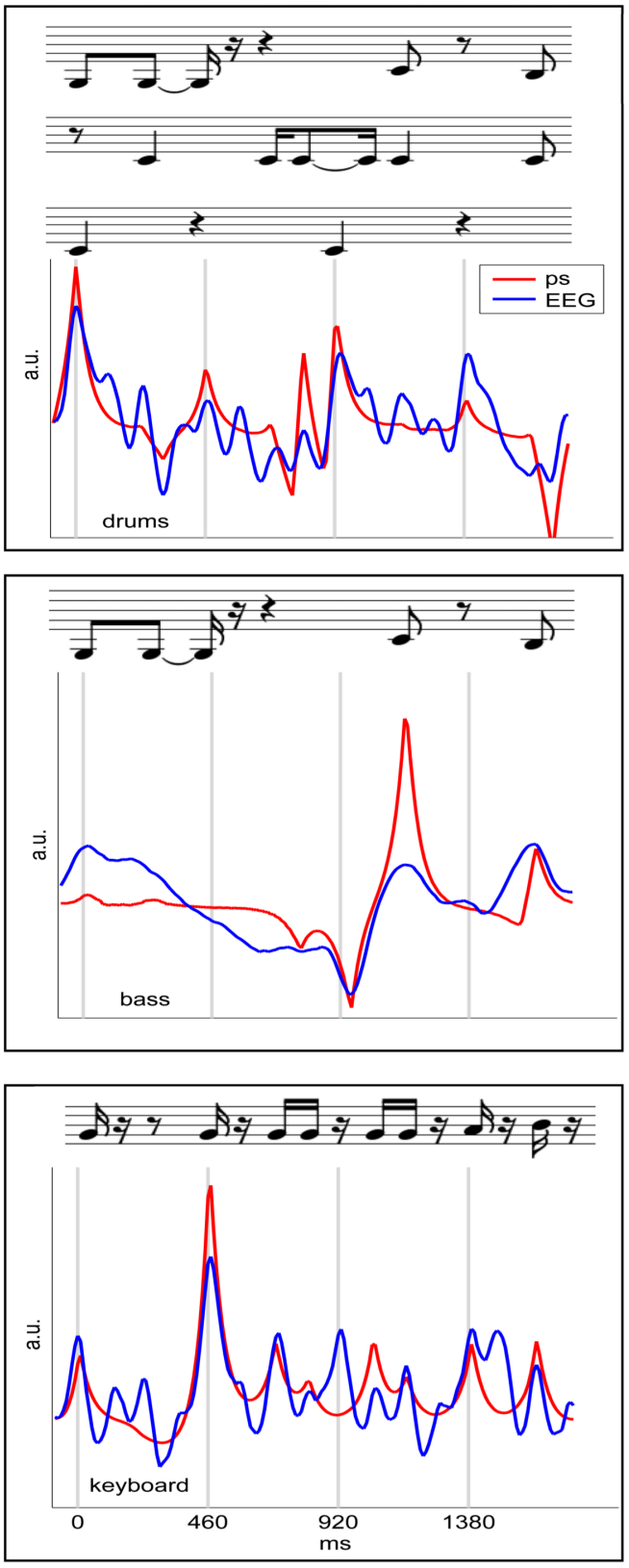

FIGURE 5.4: Solo clips: Grand Average (11 subjects) of extracted EEG projection (blue line) and audio power slope (red line), averaged across bars. The light grey vertical lines indicate the beats of the four-four time.

subject	drums	bass	keyboard
S1	100	75	67
S2	0	36	14
S3	31	100	21
S4	93	64	29
S5	57	36	64
S6	43	0	7
S7	57	79	21
S8	71	79	21
S9	71	57	50
S10	50	64	57
S11	29	64	7

TABLE 5.2: Solo presentations: Percentage of 14 solo clips that were reconstructed with significant correlation from the EEG for the three instruments.

subject	drums	bass	keyboard
S1	0.43	0.34	0.32
S2	0.23	0.26	0.21
S3	0.26	0.49	0.25
S4	0.52	0.39	0.17
S5	0.27	0.28	0.34
S6	0.22	0.13	0.08
S7	0.33	0.42	0.23
S8	0.35	0.45	0.24
S9	0.38	0.40	0.32
S10	0.32	0.33	0.30
S11	0.28	0.38	0.12
GA	0.60, p=0.00014	0.52, p=0.00011	0.54, p=0.0000004
df(corr.)	34	48	72
df(orig.)	3968	3968	3968

TABLE 5.3: Solo clips: Correlation between EEG-reconstructed power slopes (averaged across 14 music clips) and audio power slope for single subjects and the three instruments. Significance of correlation was determined taking into account the effective degrees of freedom and applying a Bonferroni correction for N=11 subjects. Shaded cells indicate significant correlation at the level of alpha=0.05. GA: Grand average over 11 subjects.

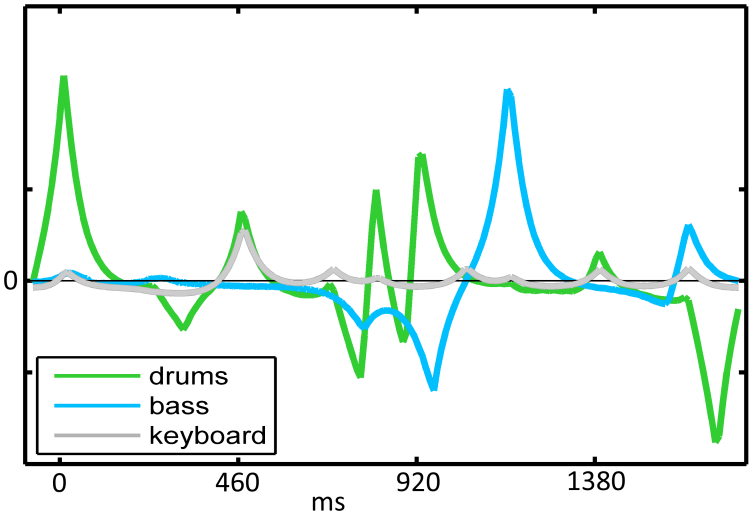

FIGURE 5.5: Audio power slopes of solo stimuli, displayed with identical scale. Amplitudes range between -8.8 and 11.2 for drums, between -5.9 and 10.5 for bass and between -0.7 and 2.8 for keyboard.

Decomposition of regression patterns Figure 5.6 shows the result of decomposing the spatio-temporal patterns with the MUSIC algorithm for one representative subject (see Section 4.2.3). In all three instruments a fronto-central scalp topography is present, resembling the topography of the N1/P2 complex. This scalp pattern is consistent for the three instruments. Furthermore, it is present in 4/11 subjects for drums, in 6/11 subjects for bass and in 5/11 subjects for keyboard. Its evolution over time differs, showing a change from positive to negative weights with extrema at 40 ms and 210 ms time lag for drums, broadly spread negative weights between 0 ms and 220 ms for bass, and a time evolution with two distinct positive peaks at 50 ms and 150 ms for keyboard. In contrast to the spatial patterns, the extracted time courses vary considerably between subjects.

Ensemble stimulus presentations Applying the three regression filters (trained on the solo stimulus presentations for the three instruments) to the EEG of the ensemble stimulus presentation extracts an EEG projection that is significantly correlated with the solo audio power slope of each instrument in 3/11 subjects for drums, in 2/11 subjects for bass, and in 9/11 subjects for keyboard (Table 5.4). In one of the subjects EEG projections significantly correlated with all three solo power slopes could be derived in parallel from the (same) EEG of the ensemble presentation, in 3/11 subjects the audio power slopes of two instruments in parallel, in 5/11 subjects for one instrument, and for 2/11 subjects for none of them. The EEG Grand Average of the ensemble presentation (11 subjects, EEG projections of 63 music clips each) is significantly correlated with the audio power slope of a solo instrument only for keyboard ($r=0.45$, $p=0.001$, effective degrees of freedom 88).

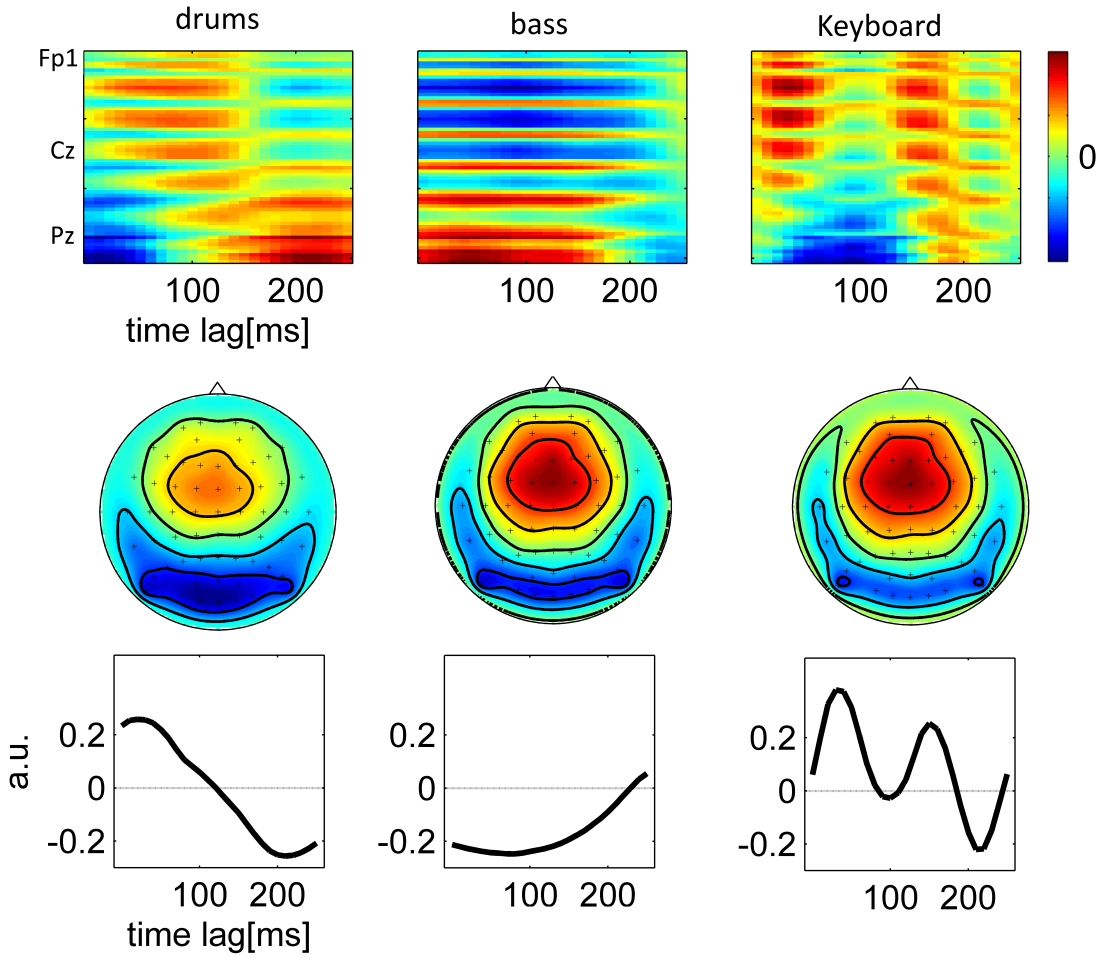

FIGURE 5.6: Spatio-temporal regression patterns and extracted MUSIC components for a representative subject. Top: Regression patterns with electrodes on the y-axis and time lag on the x-axis. Electrodes are arranged from occipital (low y values) to frontal ()high y-values). Middle: scalp pattern of first extracted MUSIC component. Bottom: time course of first extracted MUSIC component.

Specificity of reconstruction Since the solo power slopes are correlated with each other to different degrees as well as with the audio power slope of the ensemble version stimulus (Table 5.5), there is no straightforward way to estimate whether the EEG projections extracted by the instrument-specific filters are indeed specific for the instrument. To learn about the specificity, we put forward the null hypothesis that the instrument-specific filter extracts a representation of all onsets of the ensemble version stimulus. We compare Fisher-z-transformed correlation coefficients between EEG projections derived by the instrument-specific filter and solo audio power slopes to those between the same EEG projections and ensemble version audio power slopes in a paired Wilcoxon signed rank test. Figure 5.7 shows that for keyboard in all but one subject the EEG projection is more highly correlated with the keyboard audio power slope than with the ensemble version audio power slope, resulting in a significant difference between the distributions of correlation coefficients at group level ($p=0.002$). For drums and bass there were no significant differences.

subject	drums	bass	keyboard
S1	0.36	0.22	0.38
S2	-0.13	-0.06	0.25
S3	-0.07	-0.14	0.16
S4	0.0	-0.11	0.35
S5	-0.23	-0.06	0.47
S6	0.01	-0.12	0.25
S7	-0.01	0.23	0.20
S8	0.09	0.0	0.12
S9	-0.12	-0.09	0.36
S10	0.2	0.08	0.25
S11	0.26	0.09	0.20
GA	0.04	0.01	0.45, p=0.0001 df(corrected)=69

TABLE 5.4: Polyphonic clips: Correlation between instrument-specific power slopes reconstructed from the EEG of the polyphonic presentation (averaged across 63 music clips) and audio power slope of the respective single instrument for all 11 subjects and the three instruments. Significance of correlation was determined by estimating the effective degrees of freedom and applying a Bonferroni correction for N=11 subjects. Shaded cells indicate significance of correlation at the level of alpha=0.05.

r	bass	keyboard	ensemble
drums	-0.15	0.24	0.48
bass			-0.05
keyboard	0.06		0.26

TABLE 5.5: Correlation between audio power slopes of solo and ensemble version stimuli.

Effect of attention When listening to the 63 ensemble version clips subjects were instructed to focus on a specific instrument before each clip, resulting in 21 trials of an 'attended condition' and 42 trials with an 'unattended condition' for each instrument. We tested whether the correlation between the EEG-reconstructed instrument-specific audio power slope and the respective audio power slope significantly differed between these two conditions by performing a random partition test with 1000 iterations. For single subjects a significant increase in correlation was present for drums in one subject (S1), for bass in two subjects (S5, S11), and for keyboard in five subjects (S6, S7, S8, S9, and S10). Within the group of subjects a significant effect of attention was present for keyboard ($p = 0.001$).

Behavioral performance The behavioral performance differs for the three instruments (see Table 5.6) with highest counting accuracy for keyboard (Grand Average: 74%

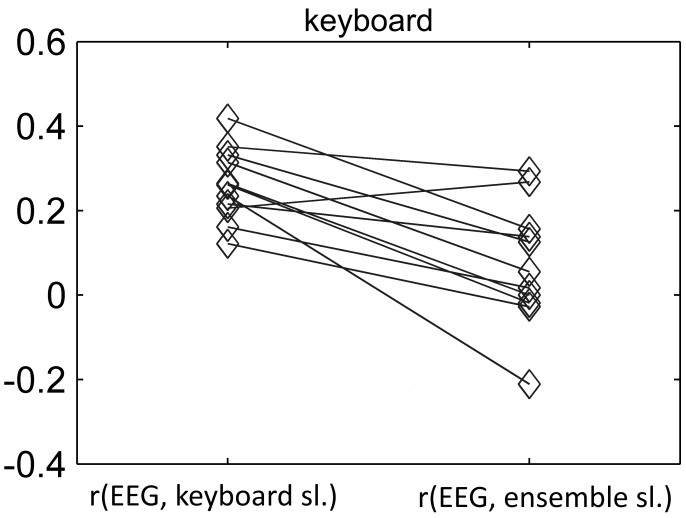

FIGURE 5.7: The EEG-reconstructed keyboard power slope extracted from the EEG of the ensemble presentation by applying the keyboard-specific filter is correlated higher with the solo keyboard audio power slope than with the ensemble version audio power slope.

correctly counted deviant stimuli), second highest accuracy for drums (71%) and lowest for bass (63%) (for details see Treder et al. (2014)).

subject	drums	bass	keyboard
GA	0.71	0.63	0.74

TABLE 5.6: Behavioral performance: Percentage of correctly counted trials (Grand Average).

5.2.2.4 Discussion

In this chapter we have applied a regression-based stimulus reconstruction method in order to extract a neural 'echo' of the sequence of note onsets that constitutes a musical stimulus from the listener's EEG. We have demonstrated that the proposed approach allows to robustly track the onset sequence of three monophonic complex music-like stimuli. Moreover, if the characteristics of a naturalistic complex sound pattern can be encoded by such a model, in principle this can be applied to extract an EEG representation of the respective sound pattern even if it is embedded into an ensemble of musical voices. Thus, our approach can provide a neural representation that parallels the separate streams a listener perceives.

Related methods The proposed application of Linear Ridge Regression with the audio power slope as a target function extends a range of approaches from the domain of speech processing where Envelope Following Responses (EFRs) have been extracted from continuous EEG and MEG with a variety of methods (see Section 4.1). The method used here belongs to the family of regression-based stimulus reconstruction methods.

In particular, the proposed method is related to the reverse correlation approach of O'Sullivan et al. (2014) since we regress EEG onto a sound envelope-related target function and operate on single trials. We extend O'Sullivan's technique by introducing an analytical way to estimate the shrinkage parameter. Importantly, we provide a way to transform the regression filters into a format that is neurophysiologically interpretable.

Physiological plausibility The extracted MUSIC components revealed a scalp pattern that was consistent between subjects and instruments. This common scalp pattern is reminiscent of a N1-P2 complex, as described in Chapter 5.2.1. The temporal dimension of the extracted components of the three instruments is much more variable. As such, the range where the extracted time courses peak is in line with the optimal time lag of cross-correlation between brain signal and sound envelope of 180 ms reported in Aiken and Picton (2006) and with results of O'Sullivan et al. (2014). In the present stimuli, however, note onsets occur in quick succession, such that the window of 0 to 250 ms time lag of the regression model potentially covers more than a single onset/ERP component. This means that the regression model not only might 'learn' latency and spatial distribution of onset-related brain responses, but could be sensitive also to the rhythmic structure of the stimulus sequence.

Auditory stream segregation The study goal was to approach the two-fold inverse problem of reconstructing (known) sound sources that create a mixed sound signal from the EEG signal of an individual who listened to this mixed signal. This enterprise capitalized on the assumption that the brain performs auditory scene analysis and creates a representation of these single sources. In the present scenario the listener was presented

with a sound scene that is stylistically relatively close to real music and, therefore, may invoke our natural abilities to stream music. The present stimulus provides a whole range of spectral, timbral and rhythmic cues on several time scales and these occur both, sequentially and simultaneously, promoting the segregation into streams. In the present scenario, thus, users were expected to perceive separate streams, and this assumption was confirmed by the behavioral results.

The present results are a proof-of-concept that a neural representation of such a stream can be extracted from the EEG, at least for one of the sound sources, here for the melody instrument keyboard.

This is an encouraging first result. For a continued application of the proposed technique in this line of research, however, an important issue has to be addressed. One general strength of multi-variate spatial filtering methods for EEG is that filters are determined not only to enhance a signal of interest, but also to project out stimulus-unrelated noise in an optimal way (Blankertz et al., 2011). In the present analysis, however, we first train the regression filters on solo presentations where no additional auditory input is present. After that, we apply them in a situation where the solo stream is part of the mixed stimulus. This means that in the training step the influence of the other instruments (that is to be projected out later) cannot be 'learnt' by the regression model. Therefore, the current application of the method may not make full use of the benefits of the spatial filtering technique. For future investigations it might be an interesting topic to probe whether the EEG representations of auditory streams can be reconstructed better if training data is used where other sources of sound are present.

Our results represent a link to the great number of studies that investigate the human 'cocktail party problem' by examining cortical activity that tracks the sound envelope of speech (for an overview see Ding and Simon (2014)) in multi-speaker environments. These have demonstrated that Envelope-Following-Responses (EFRs) can be utilized to decompose the brain signal into representations of auditory streams. Moreover, selective attention leads to an enhanced representation in the attended stream while the to-be-ignored stream is suppressed (Ding and Simon, 2012a,b, Horton et al., 2013, Kerlin et al., 2010, Kong et al., 2014, O'Sullivan et al., 2014, Power et al., 2012). Our results demonstrate that such an approach can be (at least partially) successfully applied in a music-related context where (to our knowledge) neural correlates of auditory stream separation (ASS) have not yet been investigated for continuous naturalistic stimuli. The effect of attention on the reconstruction quality for the melody instrument keyboard is in good accordance with an enhanced representation of an attended auditory stream in speech processing.

Critically, however, our stimulation scenario differs in some important points. In contrast to typical 'cocktail party' situations, the voices that constitute the present multi-voiced stimulus are more strongly correlated and do not compete, but are integrated into an aesthetic entity. Furthermore, subjects were presented the same multi-voiced stream at both ears, while multi-speaker paradigms typically make use of a spatial separation of streams. Our results show that in absence of spatial cues and with a high coincidence of onsets between streams still at least two neural representations of streams could be extracted in parallel for some subjects. The time signatures that we derived from the regression filters suggest that such neural representations depend on differences in the shape of the time course of related ERPs.

Our results contribute to the domain of auditory ERP-based BCI where ERPs like the N1 and P2 or P3 have been exploited in order to decode the user's target of attention from the EEG (Hill et al., 2012, Höhne et al., 2011, Schreuder et al., 2010, 2011, Treder and Blankertz, 2010, Treder et al., 2014), to name only a few. In particular, our results may contribute new insights with respect to so-called 'streaming' paradigms (Hill et al. (2004), Kanoh et al. (2008), for details see Hill et al. (2012)) that do not necessarily rely on the distinction of target vs. non-target auditory events, but on the attended stream. Our results confirm that such applications may in principle be designed without an oddball paradigm and based on more naturalistic stimuli.

We found that the voice of the melody keyboard was the only one that was successfully reconstructed from EEG of the ensemble presentation, while all solo voices were reconstructed similarly well if instruments played alone. First of all, this reveals that the instrument with the lowest level of sound intensity in the solo part (see Figure 5.5) is encoded strongest in the brain response. This suggests that sound intensity, the physical basis of perceived loudness, is not the critical feature for this aspect of processing, but that other characteristics of the keyboard part lead to its distinct representation in the EEG of the ensemble presentation. This finding is in line with a high-voice superiority effect for pitch encoding that has been demonstrated by means of the Mismatch Negativity (MMN) in Fujioka et al. (2005), Marie and Trainor (2012, 2014) and that can be explained by the two-tone masking effect (for a summary see Trainor (2015)): when a low-pitched and a high-pitched tone are presented together, the harmonics of the higher pitched tone tend to mask the harmonics of the lower pitched tone. In the present stimulus, instruments play their notes mostly simultaneously and, consequently, the high-pitched keyboard masks the other instruments.The high-voice superiority effect is consistent with the musical practice of putting the melody line in the highest voice and has been supported by concomitant behavioral observations of superior pitch salience in the high voice (Crawley et al., 2002, Palmer and Holleran S., 1994). Our findings demonstrate the physiological reality of this phenomenon in a naturalistic listening scenario.

Limitations The results presented here show that multivariate methods of EEG analysis can achieve considerable advances. On the one hand, previous results on the processing of tone onsets have been transferred to more complex stimulation scenarios, on the other hand, complex challenges like the reconstruction of streams can be approached. Notwithstanding, several issues call for further exploration.

When evaluating correlation-related results in this scenario one has to keep in mind that the audio power slopes of all instruments and the ensemble version audio power slope are not independent of each other, but correlated to different degrees. This makes a comparison of correlation coefficients difficult; the periodic nature of the stimuli adds further limitations. Consequently, differences in absolute correlation coefficients are hard to interpret. Therefore, the present analysis was based on significance measures taking into account differences in the periodicity of the signals (see 2.4.7). One possible concern is that the differences in reconstruction quality between keyboard and the other two solo instruments in the ensemble condition might just reflect the relations between the respective audio power slopes, more specifically, that the higher fidelity of the EEG-reconstructed keyboard slope is due to its relation to the ensemble version audio power slope. While such effects are inherent in this context and cannot be ruled out completely, two points argue in favor of a genuine instrument-specific EEG-based representation of

the keyboard's part in the ensemble condition: First, the correlation of the (original) slope of drums with the ensemble version slope is much higher than that of the (original) keyboard slope (see Table 5.5), but its reconstruction quality is poor in most subjects. Second, the EEG-reconstructed keyboard slope in all but one subjects is more similar to the original keyboard slope than to the ensemble version audio power slope, suggesting that this reconstruction indeed is specific for the keyboard part.

The stimulus sequence contains infrequently occurring deviant sound patterns in each instrument's part. These trigger a P300 component which is the key EEG feature in the operation of the original 'musical' BCI application. The present analysis uses only time lags between 0 and 250 ms and, consequently, should not make direct use of the 'strong' P300 component. Even though P3 to deviants may be picked up by our spatio-temporal filter, its reflection in the EEG projection will not be in 'sync' with the audio power slope and will rather lead to lower correlation with the power slope. However it cannot be completely ruled out that the processing of deviants influences also the earlier components. Since deviants occurred only infrequently, a possible influence would be 'diluted' strongly. Still, at this point, no strong claim can be made whether this approach can be transferred to a truly oddball-free, even more naturalistic paradigm and whether, in particular, the effect of attention is detectable in this case.

Even though the proposed method produces EEG-projections for single trials (given that training data of the same stimulus are available), a considerable part of the present effects was detected in averaged EEG projections. This means that, in a more general sense, the present approach can be regarded as an effective preprocessing step that exploits the wealth of the multivariate EEG in order to enhance the signal-to-noise-ratio and, thus, enables to extract stimulus-related activity from brain signals in far more complex stimulation scenarios. Moreover, the regression-derived patterns represent a kind of group average across the set of training data and, thus, cannot be regarded as single-trial results. In the present analysis the stimuli used for training the regression models were repetitions of one rhythmic pattern. This is not a prerequisite for applying Linear Ridge Regression, but most probably was beneficial for the 'learning processes' of the regression model. In principle, however, if an onset sequence has fairly stationary characteristics, e.g., timbre and attack, the brain response to these onsets should be extractable even in the absence of a strongly repetitive structure as in the present stimuli. This hypothesis could be addressed in future experiments.

5.2.2.5 Conclusion

In summary, the proposed regression-based multi-variate method of EEG analysis represents a promising advance towards solving a musical variant of the human 'cocktail party problem'. Because of its versatility and simplicity, we advocate its use for investigating auditory stream segregation in naturalistic listening scenarios.

5.2.3 'Real' music, free listening: brain responses to note onsets in naturalistic music stimuli

In this chapter we take a step further towards a truly 'free listening scenario' and apply the proposed approach in a setting where participants listen to excerpts of commercially recorded music without a specific task. We examine brain-stimulus synchronization (measured by CACor) for a range of musical and non-musical sounds at the level of single subjects and single presentations and explore how the presence of significant CACor relates to stimulus properties. Finally, we are interested in the behavioral/experiential relevance of CACor.

5.2.3.1 Introduction

Reflections of the sound envelope in the brain signal seem to delineate a general relationship between EEG and music stimulus that can be detected at subject-group level and across a range of music stimuli (Cong et al., 2012, Schaefer et al., 2011a, Thompson, 2013). In the last chapter we have seen that complex semi-naturalistic music stimuli leave a distinct reflection in the EEG signal. Applying Linear Ridge Regression with the audio power slope as target function, this reflection can be extracted with high fidelity from single stimulus presentations. At group level the extracted reflection is detailed enough to distinguish single streams that are part of a mixed sound signal, and sensitive enough to reveal an influence of attentional state. Furthermore, we have shown that this link between brain signal and music relies on N1-P2 ERP responses to note onsets.

Given the susceptibility of the onset-related N1-P2 response to a variety of stimulus-related, subject-related and situational influences (Baumann et al., 2008, Choi et al., 2014, Fujioka et al., 2006, Meyer et al., 2006, Näätänen and Picton, 1987, Schaefer et al., 2009, Shahin et al., 2003, 2008, 2005, 2010, Trainor et al., 2002, Tremblay and Kraus, 2002, Tremblay et al., 2001, 2014, Winkler et al., 2009), it is an interesting question whether such an imprint of the stimulus on the brain signal can be utilized to obtain valuable information about aspects of music perception. In order to gain a first insight into this question we explore the relationship between EEG signal and stimulus envelope for a range of music and non-music stimuli. Subsequently, we relate our findings to acoustic factors and behavioral measures of experiencing music. A prerequisite for such an analysis is that envelope-following responses can be derived for single presentations and single subjects, and that the experimental paradigm approximates a natural listening situation as far as possible.

We analyze EEG recordings of subjects (N=9) who listened to nine stimuli from different sound categories (mostly natural music) and examine the resulting Cortico-Acoustic Correlation (CACor), the correlation between a regression-derived EEG projection and an audio power slope at the level of a single subject and a single presentation. In addition, we extract EEG projections based on inter-subject-correlation. From the set of stimuli we derive, both, global and continuous stimulus descriptions obtained by acoustic waveform analysis. We record continuous subjective tension ratings in a separate behavioral experiment.

Our goal is to (i) probe whether significant CACor is present in EEG recordings where complex naturalistic (non-repetitive) music stimuli were presented, (ii) how CACor differs between stimuli, subjects and presentations and whether determinants of these differences can be identified, (iii) how CACor relates to behaviorally reported measures of experiencing music, and (iv) whether CACor can be detected in a 'less supervised' way where EEG projections are not explicitly optimized to match the audio power slope.

5.2.3.2 Methods

This analysis was performed on dataset 2 that is related to a set of nine music and non-music sound clips. The data set (after generic pre-processing) comprises 61-channel EEG data (three presentations per stimulus and subject), a set of nine extracted audio features for each stimulus and continuous tension ratings of 14 subjects. For details of the data acquisition see Section 5.1.2, for an overview about the different types of data see Figure 5.2.

Preprocessing of tension ratings The continuous tension ratings for each stimulus were averaged for the 14 subjects. When examining the relationship between tension ratings and stimulus parameters it has to be considered that typically ratings lag behind the stimulus. We utilize the influence of sound intensity on tension ratings that has been reported in the literature (Farbood, 2006, Farbood and Schoner, 2009, Lehne et al., 2013b) to determine the optimal time lag between tension ratings and stimulus. We calculate the cross-correlation between the Grand Average z-score-transformed tension ratings and sound intensity for time lags from 0 to 3s in steps of 10 ms and identify an optimal lag for each stimulus. The resulting time lags ranged from 760 to 1110 ms for the music stimuli, for the non-musical stimuli cross-correlation sequences did not reveal a clear peak. For these stimuli, in accordance with the literature (see above), we set the time lag to 1000 ms. All results related to tension ratings in the following were obtained after correcting for these time lags and re-sampling to the sampling rate of the time-resolved CACor of 3.33 Hz.

In addition to group averages, collections of tension ratings can also be examined with respect to the degree of inter-subject coordination a stimulus exerts on the ratings. Following the framework of activity analysis (Upham et al., 2012) we determined for each of the 50%-overlapping 1000 ms time frames the percentage of increasing tension ratings (out of the 14 ratings of all subjects) and that of decreasing ratings (activity index). Subsequently, we evaluated whether these proportions are significant by applying a permutation-based testing framework (for details see Farbood and Upham (2013) and also Section 2.4.7). This resulted in a set of time windows with significantly coordinated rating activity for each stimulus. This can be used to determine a global Coordination Score for each stimulus that indicates how much significantly coordinated rating activity (either increasing or decreasing) occurs in the course of the stimulus. This measure allows to compare stimuli with respect to their 'power' to coordinate behavioral responses.

Analysis of Cortico-Acoustic Correlation After generic pre-processing (see Section 5.1.2) we performed a temporal embedding from $0, \ldots 300$ ms on the EEG data and extracted the power slope of all stimulus waveforms as described above (Section 5.1.1).

We trained a Linear Ridge regression model to optimize the correlation between the temporally embedded EEG of single subjects and the power slope of the audio signal. This is done in a leave-one-presentation-out cross-validation approach: for each stimulus and each subject a spatio-temporal filter was trained on two concatenated presentations (training set) and applied to the remaining third presentation (test set), so that each presentation was once the test set. This resulted in a one-dimensional EEG projection for each stimulus presentation, and, accordingly, in 27 EEG projections for one stimulus (for nine subjects and three presentations each). This set of EEG projections served as basis for examining the relation between brain signal and onset structure at several levels.

Transformation and decomposition of spatio-temporal filters To obtain MUSIC components we concatenated the EEG recordings of all three presentations of each stimulus and subject, performed a temporal embedding from $0, \ldots 300$ ms and trained a Ridge regression model to predict the corresponding power slope. We reduced the spatio-temporal pattern obtained from the regression filters to a subspace containing 98% of the variance and derived 2-4 MUSIC components with a spatial and a temporal pattern per subject and stimulus.

The second column of Figure 5.3 shows an example of one component of such a decomposition. Comparing these decompositions (containing 2-4 single components each) between subjects is difficult, since no canonical representation can be derived. Visual inspection of this large collection of scalp patterns suggested the presence of a scalp pattern that was consistent between subjects and stimuli and resembled the scalp topographies of the onset-related ERPs for *Chord sequence*. In the decomposition this spatial component occurred either as a first, second or third component. We extracted a reference pattern from the onset-related ERPs obtained by classical ERP analysis for the stimulus *Chord sequence* (see Section 5.1.2.3) that is shown in Figure 5.3. This was done by averaging the scalp topography of all subjects within a 20 ms time window that corresponds to the maximum amplitude of each subject's onset ERP at channel Fz. The time windows were determined manually and ranged between 130 ms and 280 ms after tone onset. Subsequently, the distance between all spatial components and this reference was calculated and the most similar pattern for each subject and stimulus was selected for further comparison. In 43% (38%, 18%) of the components the selected pattern occurred in the first (second, third) MUSIC component.

Cortico-Acoustic Correlation coefficients for single presentations We calculated the CACor coefficient between EEG projection and the audio power slope for each single presentation (nine stimuli, nine subjects, three presentations per stimulus and subject) and assessed its significance (details as described below). This aims at assessing whether the onset structure of the stimulus is significantly reflected in the brain signal. In addition, the Grand Average of all 27 EEG projections per stimulus was calculated and a group level CACor coefficient was derived for each stimulus.

CACor score profiles and music feature profiles Subsequently, the relation between CACor, behavioral results and stimulus properties was compared between stimuli: Based on the significance of CACor coefficients for each single presentation a global

CACor score was calculated for each stimulus that specifies in how many of the 27 presentations a significant CACor is present. These scores can be summarized in a CACor score profile for the nine stimuli. In a similar fashion, a Coordination score profile is constructed from the Coordination scores derived in the behavioral experiment (see Section 5.2.3.2). Additionally, for each of the nine acoustic/musical properties that were obtained in the audio analysis a profile that describes the magnitude of the respective stimulus feature for all stimuli was constructed. The pairwise correlation between CACor score profile, Coordination score profile and music feature profiles was calculated.

Time-resolved CACor The dynamics of the brain-stimulus synchronization within each stimulus were examined by calculating a group-level CACor in a time-resolved manner: For each stimulus the Grand Average of all 27 EEG projections (three for each of the nine subjects) was segmented into 90% overlapping time frames of 3 s duration. Subsequently, CACor was calculated for each time frame, resulting in a time course of correlation coefficients that has a sampling rate of 3.33 Hz. Correlation coefficients between this time course and the time courses of the nine acoustic/higher-level music features were determined. In order to be able to do this, all music features were resampled to match the sampling rate of 3.3 Hz. In an analogue manner, the relation between mean tension ratings and music features was determined. Figure 5.2 gives a complete overview about data types and correlation coefficients that were calculated in the present analysis.

Significance of correlation In the present analysis we encounter both of the statistical problems we have described in Section 2.4.7:

on the one hand, EEG time courses, tension ratings and extracted audio features have different degrees of auto-correlation. In order to gain more insight into the advantages/-drawbacks of the two methods for correction of auto-correlation that we have described in Chapter 2.4.7, we assessed the significance of CACor at single subject and single presentation level using both methods. For Pyper et al.'s method the maximal time lag of autocorrelation that was taken into account was 2 s.

Furthermore, the set of features extracted from the audio waveforms is correlated to different degrees. For the present selection of stimuli the correlation coefficients ranged between r=-0.74 and r=0.93. To account for this correlation, the relation between CACor time courses/tension ratings and music features was determined using the partial correlation coefficient (see Section 3.2). For assessing the significance of partial correlation coefficients, however, the (most likely more sensitive) permutation testing approach is not optimal, since it is not possible to generate surrogate data with the same intrinsic relations as the extracted music features. Therefore, Pyper et al.'s method (maximal lag 10 s) is applied to correct for autocorrelation in order to estimate the significance of partial correlation coefficients. The resulting p-values were corrected for multiple comparisons (false discovery rate (FDR), $q < 0.05$).

Multiway CCA After generic pre-processing (see Section 5.1.2) we performed Multiway CCA (for details see Chapter 2, Section 2.4.2) in order to find EEG components that are common to all subjects that listen to the same piece of music. In our main analysis (see above) we performed Ridge Regression with a stimulus-derived target function and, therefore had to take into account a delay of the brain response with respect

to the stimulus. Now, we maximize the correlation between EEG components of nine subjects, and, assuming similar processing times for all subjects, do not need a temporal embedding step. We train a regularized CCA model on the EEG recordings of all nine participants such that the sum over all pairwise correlations (for all pairs of subjects) between the extracted EEG projections (called canonical components) is maximized. In general, CCA produces as many components as dimensions in the data (here: as EEG channels). Here, we focused on the first component that is associated with the highest correlation. As before, we trained and validated the CCA models in a leave-one-presentation-out cross-validation procedure. Consequently, for each presentation we obtained one EEG projection (called canonical component) that was derived by a CCA filter that was trained on a separate portion of data, the filter itself and a corresponding pattern. From the canonical components we calculated the mean Canonical Inter Subject Correlation (CISC), the average of all 36 pairwise CISCs for each stimulus presentation as a measure of similarity within the group of subjects. For details of the calculation see Algorithm 3, Figure 5.8. We tested the significance of these mean CISCs with a permutation testing approach with 500 iterations where surrogate versions of all nine EEG time courses were generated (see Section 2.4.7) and a mean CISC was calculated. The resulting p-values were corrected with a Bonferroni correction for N=3 repetitions for each stimulus.

The main goal of this additional step of analysis was to learn whether there is a relation between the most inter-individually consistent feature of the listeners' EEG and the reflection of tone onsets as detected in the previous analysis. Therefore, we calculated CCA-CACor coefficients between the CCA-derived EEG projection and the audio power slope. In order to do this, we first examined the cross-correlation sequences between Grand Average Canonical Components for each stimulus. This resulted in an optimal time lag of 210 ms. We, then, calculated the CCA-CACor coefficient between the Grand Average Canonical Components (across all 27 presentations of a stimulus) and the audio power slope (delayed by 210 ms) in a permutation testing approach with 500 iterations using surrogate versions of the audio power slope.

Algorithm 3 Pseudocode for deriving a mean CISC coefficient from a given set of EEG training and test data of s subjects.

Require: EEG data: s data sets $X_{11}, \ldots X_{1s} \in N \times T_1$ (training data), s data sets $X_{21}, \ldots X_{2s} \in N \times T_2$ (test data), after generic preprocessing as described in Section 4.2.1

3. **Train CCA model on training data**

$W_1, \ldots, W_s \leftarrow \text{CCA}\ (X_{11}, \ldots, X_{1s})$

4. **Apply CCA model to test data**

for $i = 1$ **to** s **do**

  $comp(:, i) = W_i^\top X_{2i}$

end for

4. **Calculate pairwise inter-subject-correlations**

$cr \leftarrow corr(comp)$

5. **Calculate mean inter-subject-correlations**

$pairwise\_corr \leftarrow zeros(1, \frac{s(s-1)}{2})$

$count \leftarrow 0$

for $u = 1$ **to** s **do**

  for $v = u + 1$ **to** s **do**

    $count \leftarrow count + 1$

    $pairwise\_corr(count) = cr(u, v)$

  end for

end for

$mCISC \leftarrow \frac{sum(pairwise\_corr)}{\frac{s(s-1)}{2}}$

FIGURE 5.8: Algorithm 3: Pseudocode for calculating mean Canonical Inter Subject Correlation (mCISc).

5.2.3.3 Results

Tension ratings Figure 5.9 gives an example of the continuous tension ratings obtained in the behavioral experiment. The top panel shows the audio waveform (blue) of the entire stimulus (*Rachmaninov*) and the sound intensity (red). In the middle panel the individual tension ratings are plotted along with the Grand Average tension ratings. The bottom panel shows the activity indices (see 5.2.3.2) that indicate how consistently tension ratings rise/fall within the group of subjects at a given time point. While the individual tension ratings vary considerable between subjects, the Grand Average shows clearly that the three-part macro-structure contained in time course of the sound intensity is represented in the tension ratings.

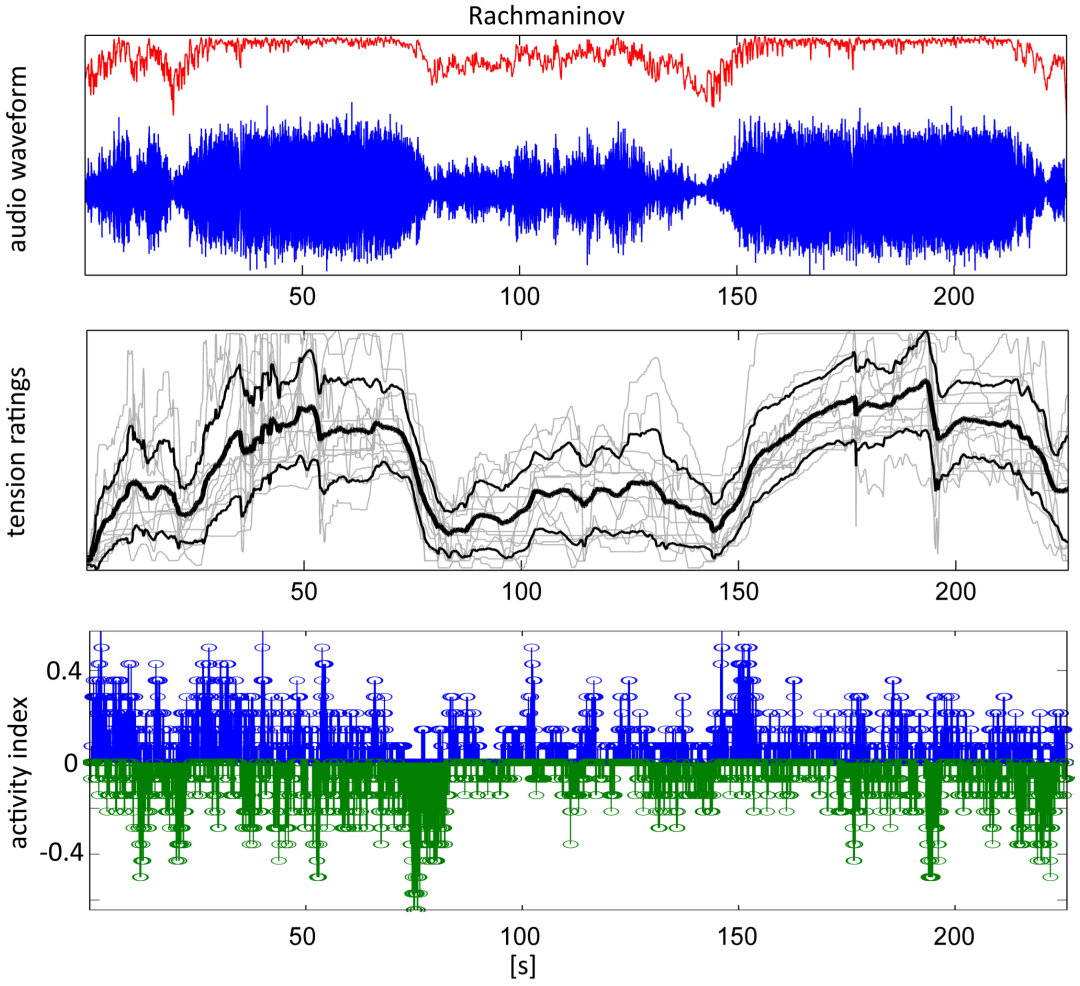

FIGURE 5.9: Example of stimulus waveform, tension ratings and results of the activity analysis. Top: Audio signal (blue) and sound intensity (red) for the *Rachmaninov* (entire stimulus). Middle: Tension ratings of single subjects (grey, N=14), Grand Average (black, thick line) and standard deviation (black, thin line). Bottom: Activity index indicating the percentage of rising(blue) or falling(green) tension ratings at a given time point.

Cortico-Acoustic Correlation

Cortico-Acoustic Correlation for single presentations Table 5.7 gives CACor coefficients for each single subject and each stimulus presentation. The bottom line contains the CACor coefficients for the Grand Average EEG projection (average of the 27 EEG projections for each stimulus). Note that correlation coefficients cannot be compared between stimuli, since duration and autocorrelation properties differ between stimuli. Shaded cells indicate significant correlation at the level of $\alpha = 0.05$. Significance was determined by applying permutation tests and subsequently correcting for the number of 27 presentations per stimulus (see Section 2.4.7). Stimuli were ordered according to the total number of presentations with significant CACor (called CACor score in the following, see Section 4.2.4). Since these were derived in a cross-validation approach significant correlation coefficients can be regarded as reflecting a genuine influence of the stimulus on brain responses that generalize across several presentations of a stimulus. To give an impression of the extracted EEG projections and their relation to the audio power slope three examples are shown in Figure 5.10. Table 5.8 summarizes the corresponding CACor scores into a CACor score profile for the set of nine stimuli. For comparison, both, the CACor score profile that was derived by applying permutation tests (dark blue bars), and that derived by applying Pyper et al.'s method Pyper and Peterman (1998) to assess the significance of correlation in signals containing serial autocorrelation (light blue bars) are given in Figure 5.11. Although there are differences in the absolute CACor scores for both methods, the ranking of the stimuli does not change. Importantly, the comparison shows that the zero scores for *Orchestra* and *Jungle* (the non-musical stimuli) indicate an absence of significant CACor for these stimuli and are not introduced by the permutation test approach. In the following the profile derived by permutation tests is used for further analysis.

Stimuli differ strongly with respect to the consistent presence of significant CACor for the 27 presentations with a CACor score ranging between 23/27 presentations for *Chord sequence* and 0/27 presentations for *Orchestra* and *Jungle*. In the Grand Average CACor is significant for *Chord sequence* and all music pieces, but not for the non-musical stimuli *Orchestra* and *Jungle*. Subjects differed with respect to the total number of presentations with significant correlation, scoring between 4/27 presentations (S1, S9) and 15/27 presentations (S5). Within the group of stimuli, first, second and third presentation do not differ significantly with respect to CACor scores.

Chord sequence	Pres 1	Pres 2	Pres 3
S1	0.15	0.15	0.21
S2	0.48	0.50	0.47
S3	0.18	0.23	0.25
S4	0.31	0.27	0.31
S5	0.23	0.18	0.22
S6	0.28	0.28	0.21
S7	0.28	0.32	0.34
S8	0.37	0.42	0.40
S9	0.15	0.14	0.14
GA	0.34		

Chopin	Pres 1	Pres 2	Pres 3
S1	0.02	0.01	0.04
S2	0.02	0.04	0.02
S3	0.15	0.13	0.16
S4	0.11	0.05	0.08
S5	0.11	0.16	0.12
S6	0.12	0.08	0.10
S7	0.09	0.11	0.09
S8	0.10	0.13	0.06
S9	0.06	0.00	0.04
GA	0.33		

Rachmaninov	Pres 1	Pres 2	Pres 3
S1	0.00	0.00	0.00
S2	0.06	0.09	0.04
S3	0.03	0.07	0.03
S4	0.05	0.06	0.04
S5	0.02	0.07	0.06
S6	0.07	0.02	0.10
S7	0.05	0.08	0.03
S8	0.01	0.04	0.06
S9	0.02	0.01	0.05
GA	0.16		

Vivaldi, Spring	Pres 1	Pres 2	Pres 3
S1	0.03	0.03	0.04
S2	0.10	0.10	0.09
S3	0.04	0.03	0.07
S4	0.03	0.02	0.01
S5	0.02	0.06	0.10
S6	0.04	0.07	0.02
S7	0.08	0.03	0.07
S8	0.08	0.10	0.02
S9	0.02	0.06	0.04
GA	0.18		

Vivaldi, Summer	Pres 1	Pres 2	Pres 3
S1	0.05	0.04	0.0
S2	0.03	0.06	0.03
S3	0.01	0.03	0.01
S4	0.03	0.01	0.02
S5	0.08	0.05	0.05
S6	0.02	0.01	0.01
S7	0.06	0.04	0.05
S8	0.04	0.08	0.03
S9	0.06	0.05	-0.01
GA	0.23		

Bach	Pres 1	Pres 2	Pres 3
S1	0.01	0.03	0.02
S2	0.4	0.0	0.01
S3	0.02	0.02	0.10
S4	0.08	0.04	0.00
S5	0.03	0.07	0.12
S6	0.03	0.03	0.01
S7	0.06	0.01	0.08
S8	0.02	0.03	0.01
S9	0.03	0.05	0.01
GA	0.12		

Williams	Pres 1	Pres 2	Pres 3
S1	0.0	0.03	0.02
S2	0.02	0.00	0.01
S3	0.03	0.05	0.04
S4	0.0	0.02	0.01
S5	0.0	0.04	0.05
S6	0.01	0.02	0.01
S7	0.03	0.02	0.0
S8	0.0	0.01	0.01
S9	0.0	0.0	0.02
GA	0.05		

Orchestra	Pres 1	Pres 2	Pres 3
S1	-0.05	-0.01	-0.02
S2	0.0	0.06	0.05
S3	0.05	0.01	0.02
S4	0.06	0.05	0.04
S5	0.03	0.06	0.0
S6	0.05	0.01	0.02
S7	0.05	0.03	0.04
S8	0.05	0.0	0.02
S9	0.02	0.0	0.03
GA	0.05		

Jungle	Pres 1	Pres 2	Pres 3
S1	0.0	-0.01	0.02
S2	-0.02	0.0	-0.0
S3	-0.05	0.03	0.0
S4	0.04	0.02	0.04
S5	0.04	0.02	0.0
S6	-0.03	-0.02	0.0
S7	0.04	0.02	-0.02
S8	0.0	-0.03	-0.05
S9	0.02	0.02	-0.01
GA	0.04		

TABLE 5.7: CACor coefficients for each single subject and each stimulus presentation; GA: Grand Average (N=27). Pink shading indicates significant positive correlation between EEG projection and audio power slope. Significance was determined by applying permutation tests and subsequently performing Bonferroni-correction for N= 27 presentations per stimulus.

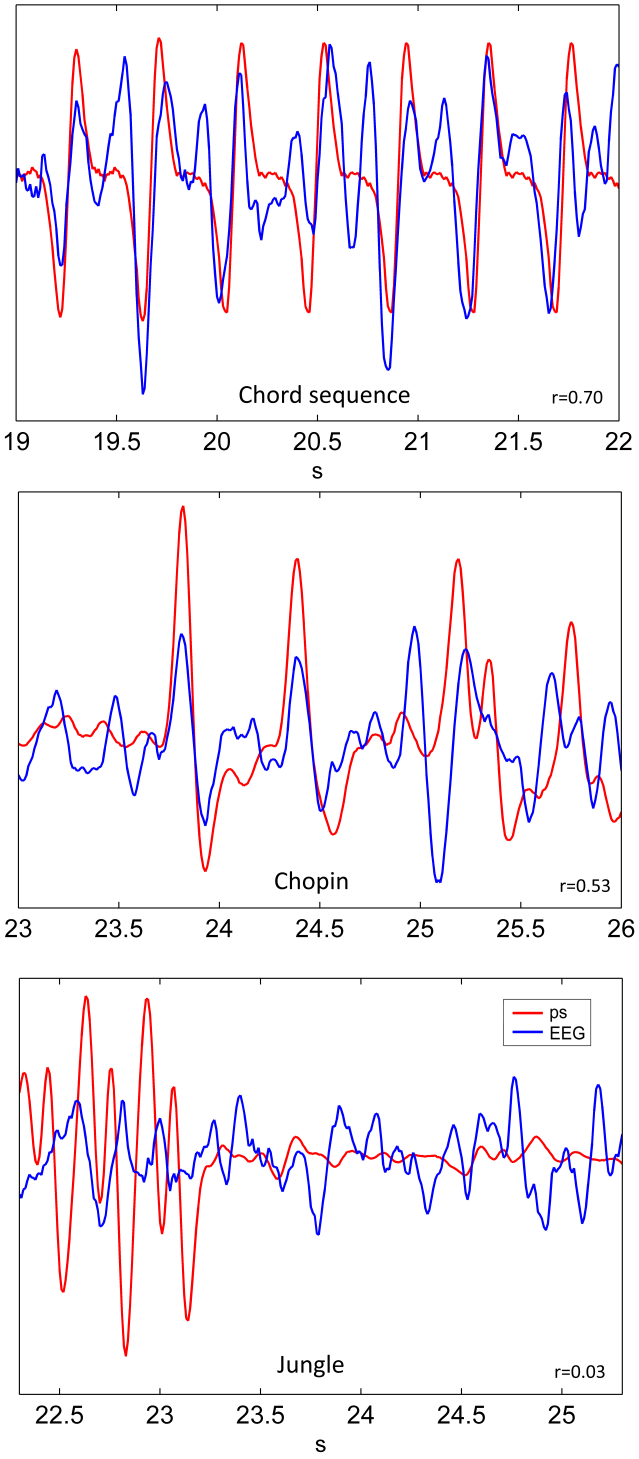

FIGURE 5.10: EEG projections. Three examples (from *Chord sequence*, *Chopin* and *Jungle*) that show 3s-long segments of an extracted EEG projection (blue) for a single stimulus presentation and a single subject and the respective audio power slope (red). Note that in the optimization procedure a time lag between stimulus and brain response is integrated in the spatio-temporal filter, and that, consequently, the EEG projections shown here are not delayed with respect to the audio power slope. The correlation coefficients indicate the magnitude of correlation for the shown segment of 3 s duration.

global music descriptor	CACor score	Coordination score
Sound intensity	0.24	0
Sharpness	0.71	0.57
Spectral centroid	-0.69	-0.57
Spectral entropy	-0.52	0.23
Spectral flux	0.28	0.11
Fluctuation centroid	0.09	-0.14
Fluctuation entropy	-0.65	-0.19
Pulse clarity	-0.32	-0.02
Key clarity	0.58	0.38
Coordination Score	0.90	

TABLE 5.8: Spearman's correlation coefficient (a) between CACor score profile and music feature profiles for nine acoustic/higher-level music features, (b) between Coordination score profile and music feature profiles. Bottom line: Correlation between CACor scores and Coordination scores. Significant positive/negative correlation is indicated by pink/blue shading, respectively.

CACor, tension ratings, and stimulus features Figure 5.11 shows CACor score profiles and the behavioral Coordination score profile for the nine stimuli. Note that the regular stimulus *Chord sequence* was not included in the behavioral experiment. However, in the CACor score profile blank bars for *Orchestra* and *Jungle* denote a score of zero. Table 5.8 gives correlation coefficients (Spearman's rho) that quantify the correlation of (a) CACor score profile and (b) Coordination score profile with music feature profiles that relate to nine stimulus features. Music feature profiles are represented by scores that indicate the average magnitude of a specific acoustic/higher-level musical property for the nine stimuli. Figure 5.13 gives a detailed overview how stimuli differ with respect to each of the nine acoustic/higher-level musical properties. The bottom line of Table 5.8 contains the correlation between CACor score profile and Coordination score profile. The CACor score profile is significantly positively correlated with the Sharpness profile and significantly negatively correlated with the Spectral centroid profile. Furthermore, moderate, but non-significant negative correlation of the CACor score profile with the Fluctuation entropy profile is present. None of the correlation coefficients between Coordination score profile and music feature profiles reached significance. CACor score profiles and Coordination score profiles are significantly correlated with r=0.90, p=0.005. Figure 5.12 illustrates the relationship between CACor scores and Coordination scores.

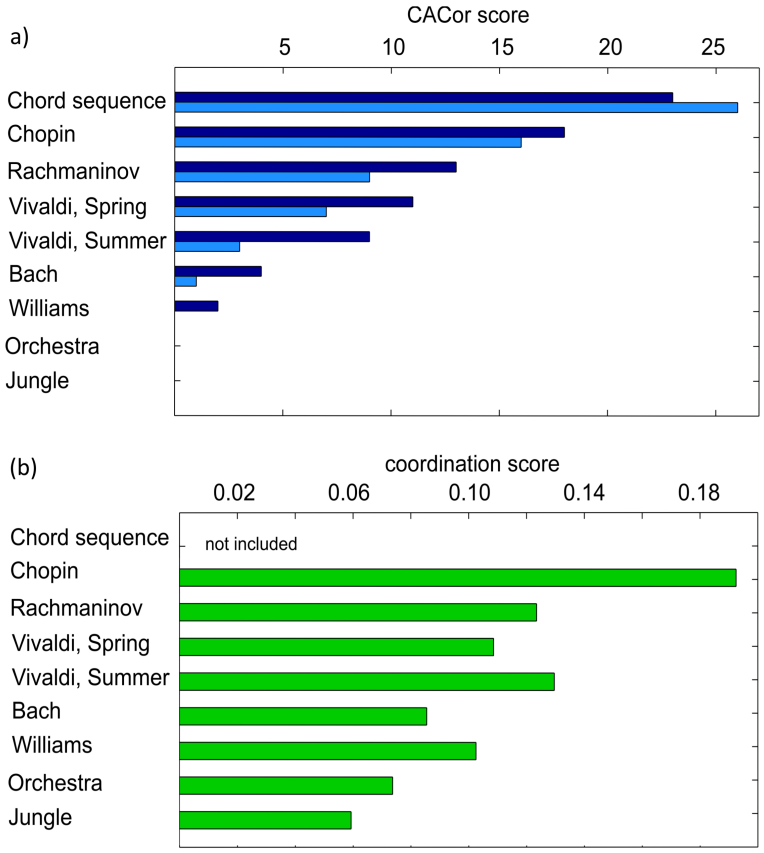

FIGURE 5.11: CACor score profile (blue) and Coordination score profile (green). The CACor score profile for the set of nine stimuli summarizes in how many of the 27 presentations significant Cortico-Acoustic Correlation was detected. Significance of correlation was determined (1) in a permutation-testing approach (darkblue bars) and (2) with Pyper et al.'s method (lightblue bars) to estimate the effective degrees of freedom (for details see 2.4.7). b) Behavioral Coordination score profile. All profiles are sorted according to the descending CACor score.

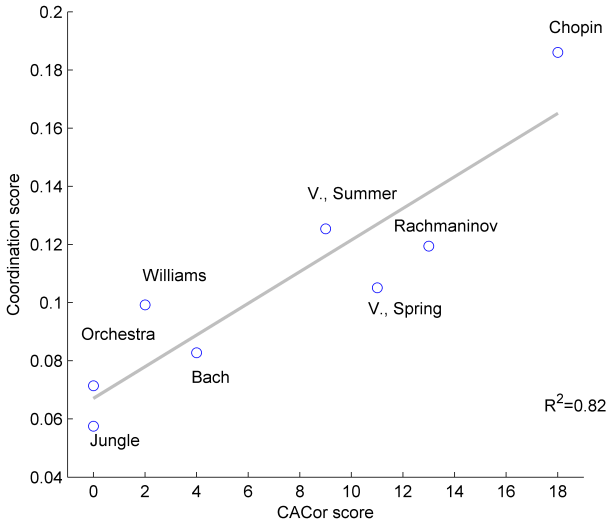

FIGURE 5.12: The scatter plot of CACor scores versus Coordination scores illustrates the relationship between both measures.

FIGURE 5.13: Music feature profiles for nine acoustic/higher-level music features. The bars indicate the average magnitude of a music feature for the set of nine stimuli. This illustrates differences in global stimulus characteristics.

Dynamics of CACor We examined how changes in group-level CACor relate to changes in stimulus features and in mean tension ratings during the course of a stimulus. The full set of partial correlation coefficients (see Section 4.2.4 for details of the calculation) is given in Table 5.9 and Table 5.10. Of these, only a small number of correlation coefficients were found significant. Note that in this part of the analysis correlation coefficients indicate whether CACor and a particular property of the music stimulus co-vary, but do not necessarily inform about significance of global CACor. Five out of the nine music features co-varied significantly at least in one stimulus with time-resolved CACor: Sound intensity, Sharpness, Spectral Centroid, Fluctuation Centroid and Fluctuation Entropy. Of these, only Sharpness (three stimuli, positive correlation) and Spectral Entropy (two stimuli, negative correlation) in more than one stimulus. Sharpness is the music feature with most consistent relation to tension ratings (3/8 stimuli), while for Sound Intensity, Spectral Entropy and Fluctuation Entropy significant correlation is present in one stimulus. Out of the nine (eight for the tension ratings, respectively) stimuli the romantic piano pieces (*Chopin* and *Rachmaninov*) show the clearest relations between CACor and music features, followed by both *Vivaldi* stimuli and, then, by one of the non-music stimuli. A similar trend is visible for the relation between tension ratings and music features.

music feature	Chord sequence	Chopin	Rach.	Vivaldi, Spring	Vivaldi, Summer	Bach	Williams	Orch.	Jungle
Sound intensity	0.0	0.19	0.01	0.06	-0.0	-0.01	-0.06	-0.03	0.0
Sharpness	-0.03	0.27	0.20	0.10	0.16	0.12	0.18	0.37	0.25
Spectral centroid	-0.03	-0.16	0.08	-0.23	-0.07	0.03	0.11	0.09	-0.02
Spectral entropy	-0.01	0.26	-0.02	0.19	0.08	0.02	-0.04	-0.0.7	0.10
Spectral flux	-0.02	0.01	-0.11	-0.02	-0.0	-0.0	-0.02	-0.10	-0.05
Fluctuation centroid	-0.02	0.02	-0.01	-0.18	-0.04	0.20	0.04	0.20	-0.13
Fluctuation entropy	-0.10	-0.20	-0.07	0.0	-0.31	-0.27	0.02	-0.18	-0.14
Pulse clarity	0.05	0.09	-0.03	-0.15	-0.02	-0.01	0.03	-0.40	-0.21
Key clarity	0.07	-0.02	-0.04	-0.23	-0.11	0.06	0.04	0.30	0.19
Tension ratings		-0.26	-0.04	-0.13	-0.13	-0.08	0.04	-0.36	

TABLE 5.9: Partial correlation of time-resolved CACor of Grand Average with nine music features. Pink shading indicates significant positive partial correlation, blue shading indicates significant negative partial correlation (α=0.05, Bonferroni correction).

music feature	Chord sequence	Chopin	Rach.	Vivaldi, Spring	Vivaldi, Summer	Bach	Williams	Orch.	Jungle
Sound intensity	-	0.27	0.22	0.18	0.13	0.05	0.03	0.0	0.14
Sharpness	-	0.16	-0.39	-0.11	-0.40	0.02	-0.04	-0.19	-0.56
Spectral centroid	-	-0.21	-0.02	-0.1	0.17	-0.05	0.18	0.03	0.20
Spectral entropy	-	0.36	0.29	0.12	-0.07	0.04	-0.03	-0.03	-0.24
Spectral flux	-	-0.04	0.01	0.05	0.06	0.05	0.05	-0.03	0.06
Fluctuation centroid	-	-0.12	0.32	0.17	0.37	-0.13	0.03	0.17	-0.08
Fluctuation entropy	-	0.28	-0.03	0.14	-0.27	0.21	0.33	-0.17	0.21
Pulse clarity	-	0.12	0.1	-0.06	-0.20	0.02	-0.19	-0.11	-0.11
Key clarity	-	0.10	-0.06	-0.22	0.23	-0.15	-0.02	0.11	0.00

TABLE 5.10: Partial correlation of mean tension ratings with nine music features. Pink shading indicates significant positive partial correlation, blue shading indicates significant negative partial correlation (α=0.05, Bonferroni correction).

Interpretation of spatio-temporal patterns The upper part of Figure 5.3 (c) shows the scalp topography that was used as a reference to select one MUSIC component per subject and stimulus for comparison (for details see Section 5.2.1). It is represented by the average ERP scalp pattern across subjects for the *Chord sequence* in a time interval of 20 ms around the individually determined point of maximum amplitude within the range of 150 ms to 230 ms after tone onset.

Figure 5.14 shows the spatial and temporal dimensions of the selected MUSIC component for the nine stimuli, averaged across subjects. A complete collection of the individual scalp patterns (one for each stimulus and subject) is contained in Figure 5.15. Note that these patterns were selected for maximal similarity with a reference pattern as described above, and, therefore, necessarily are similar to some extent. The averaged scalp topographies for all nine stimuli show a positive fronto-central pattern. The temporal patterns shown in Figure 5.16 are much more variable.

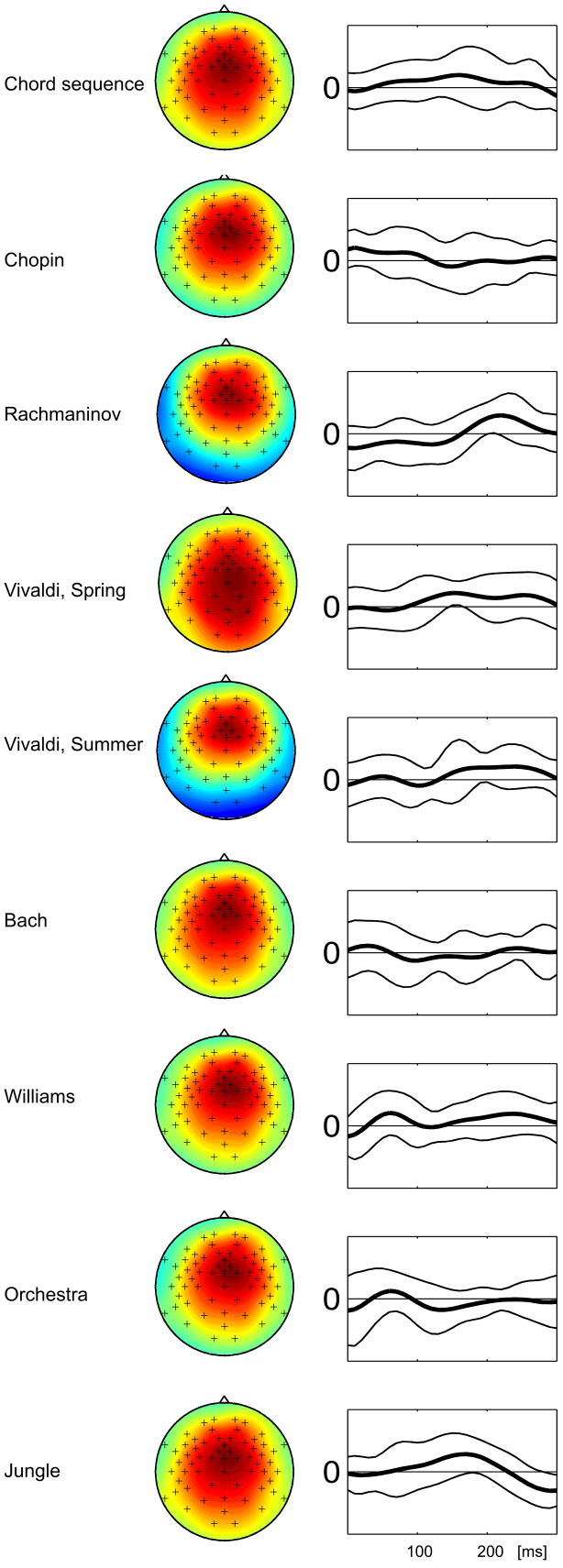

FIGURE 5.14: MUSIC components (Grand Average) for all stimuli. Scalp topographies and time courses (thick line: average of all subjects, thin lines: standard deviation) of the MUSIC component that was extracted most consistently from the decomposed spatio-temporal patterns. Single-subject scalp topographies that are included in these averages are given in Figure 5.15.

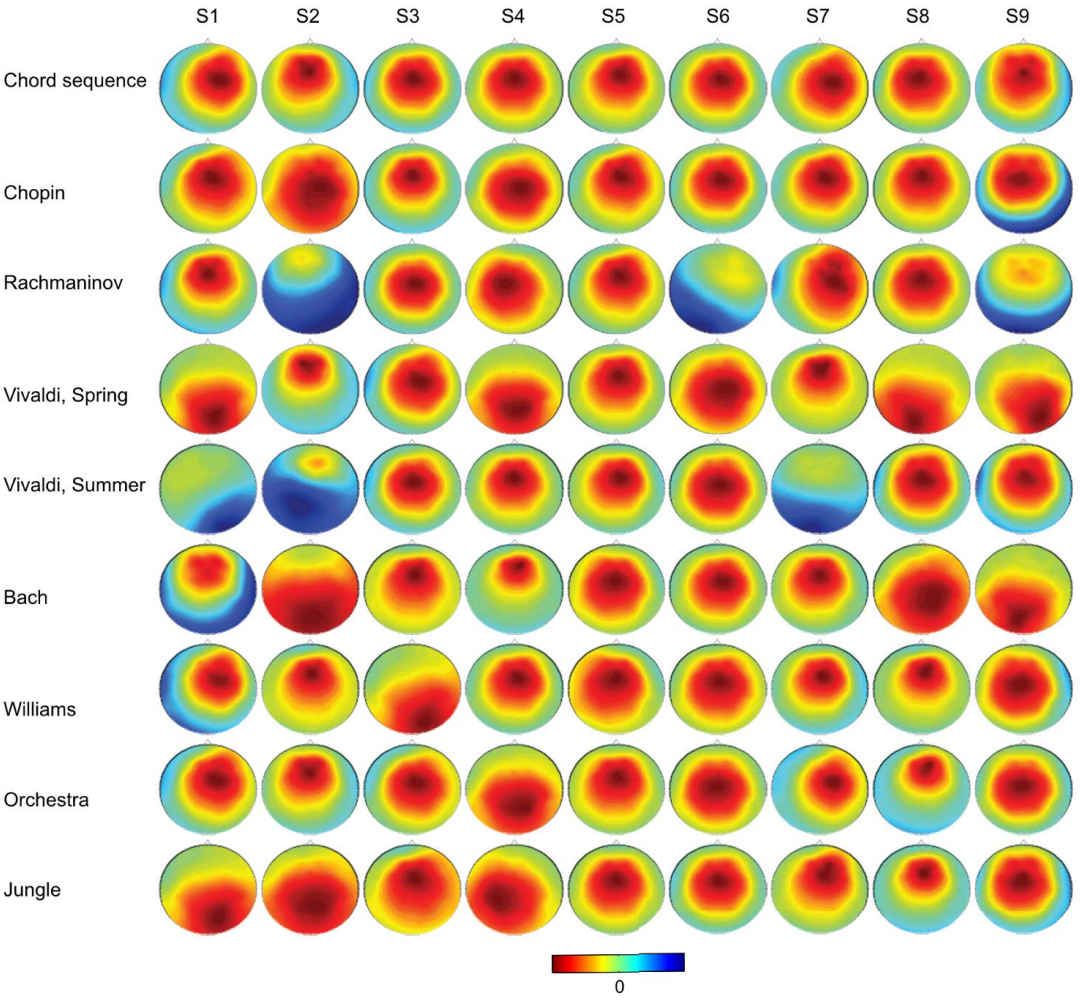

FIGURE 5.15: Scalp topographies of selected MUSIC components for all subjects and stimuli.

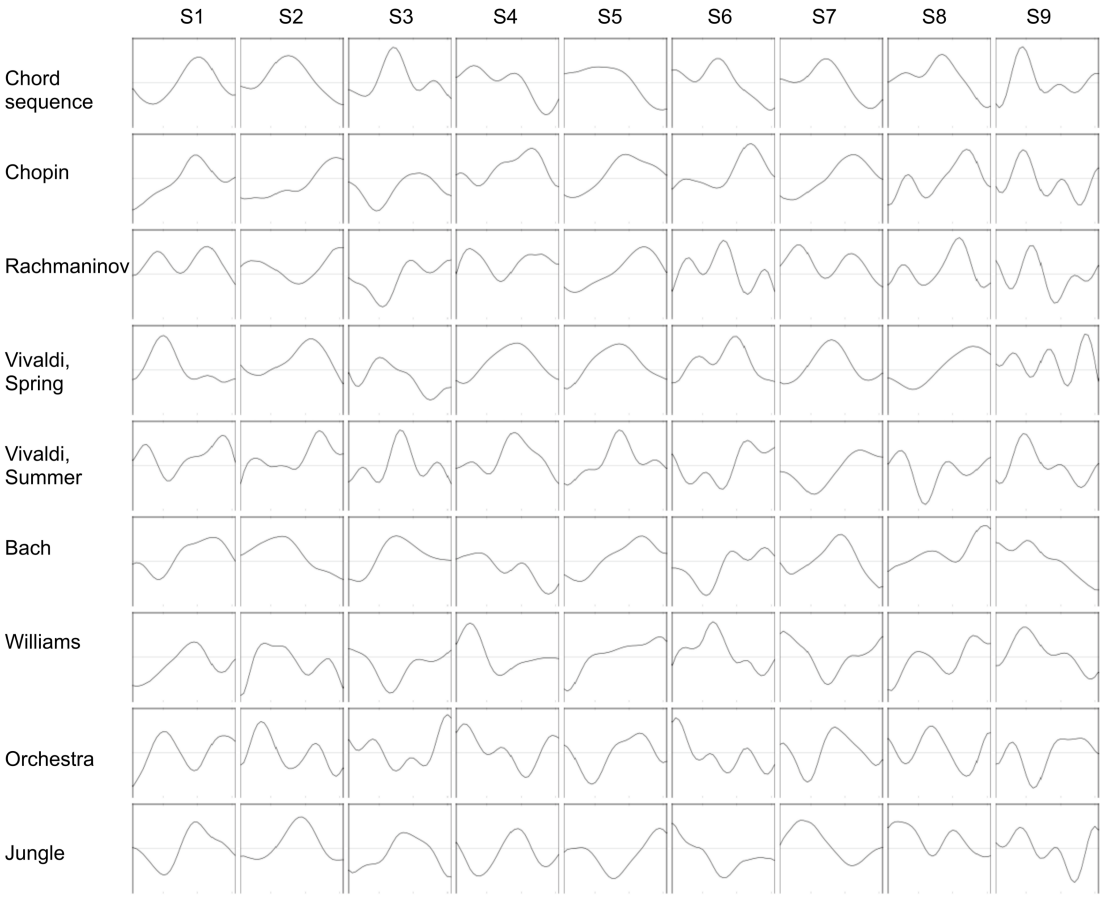

FIGURE 5.16: Time courses of selected MUSIC components for all subjects and stimuli.

Multiway CCA Table 5.11 gives in the first three columns the mean Canonical Inter Subject Correlation Coefficient (CISC) for each presentation and each stimulus. Shaded cells indicate significant correlation at the level of $\alpha = 0.05$. Significance of correlation was determined by permutation tests and subsequently correcting for the number of N=3 presentations of each stimulus (see Section 5.2.3.2). In three of the nine stimuli the mean CISC indicates significant between-subjects-similarity. Significant mean CISCs occur in those stimuli that rank high in the CACor score ranking (see Figure 5.11). In general, the magnitude of CISCs is much lower than that of the Regression-CACor coefficients.

The cross-correlation sequence of the Grand Average of CCA-derived EEG projections with the audio power slope (Figure 5.17) shows a clear peak at 210 ms for the stimuli with the highest correlation coefficients, for *Chord sequence*, *Chopin* and *Rachmaninov*.

The fourth column of Table 5.11 gives the CCA-CACor coefficient between the Grand Average of canonical components and the audio power slope (as used in the previous analysis). This correlation coefficient was calculated taking into account the optimal time delay of 210 ms (see Figure 5.17) of the EEG projection with respect to the stimulus. Note that the calculation of the CCA-CACor is a way of post-hoc relating the results of the CCA analysis (that, in principle, are independent of the audio power slope) to our previous results to probe in how far the reflection of tone onsets in the EEG may result in concomitant similarities in the EEGs of several subjects. The CCA-CACor coefficients in column 4 can be compared directly with those derived in the Regression analysis that are shown in the rightmost column for comparison. The magnitude of the CCA-CACor coefficients is (not surprisingly) smaller than of the Regression-CACor coefficients, but still significant in four stimuli.

Figure 5.18 shows examples of Grand Average EEG projections derived with multiway CCA for the same three excerpts from *Chord sequence*, *Chopin* and *Jungle* that were shown in Figure 5.10. The EEG projections (blue) and the respective audio power slope (red) are considerably less similar than those in Figure 5.10. However, a marked difference in magnitude of correlation is present between the stimuli *Chord sequence* and*Chopin* with respect to the non-musical stimulus *Jungle*. The optimal time lag of 210 ms was taken into account for plotting and for calculating correlation.

Figure 5.19 shows scalp patterns derived from the CCA filters for all subjects and three stimuli. They show a fronto-central scalp pattern that is consistently present in all subjects and all stimuli with exception of subjects S1 and S2 for *Chord sequence*. Individual differences between subjects (e.g., an occipital contribution in subject S4) occur consistently across stimuli.

Stimulus	Mean CISC Pres 1	Pres 2	Pres 3	CCA-CACor GA	Regression-CaCor GA
Chord sequence	0.010	0.021	0.012	0.20	0.34
Chopin	0.013	0.012	0.010	0.13	0.33
Rachmaninov	0.010	0.011	0.010	0.09	0.16
Vivaldi, Spring	0.002	0.006	0.003	0.04	0.18
Vivaldi, Summer	0.005	0.000	0.001	0.02	0.23
Bach	0.001	-0.001	- 0.001	0.01	0.12
Williams	-0.001	0.000	-0.001	0.01	0.05
Orchestra	-0.004	0.006	0.004	0.01	0.05
Jungle	0.001	-0.004	0.004	0.00	0.04

TABLE 5.11: Results of Multiway CCA: Column 1 to 3: Mean Canonical Inter Subject Correlation Coefficients (CISCs) for each single presentation. Mean ICSCs were obtained by averaging all pairwise CISCs. Pink shading indicates significant positive CISCs. Significance was determined by applying permutation tests and subsequently performing Bonferroni correction for N=3 presentations. Column 4: CCA-Cortico-Acoustic Correlation between the Grand Average of canonical components and the audio power slope (as used in the previous analysis). This correlation coefficient was calculated taking into account the optimal time lag of 210 ms. Column 5: Regression-CACor: Cortico-Acoustic Correlation between the Grand Average of regression-derived EEG projections and audio power slope (see also Table 5.7)

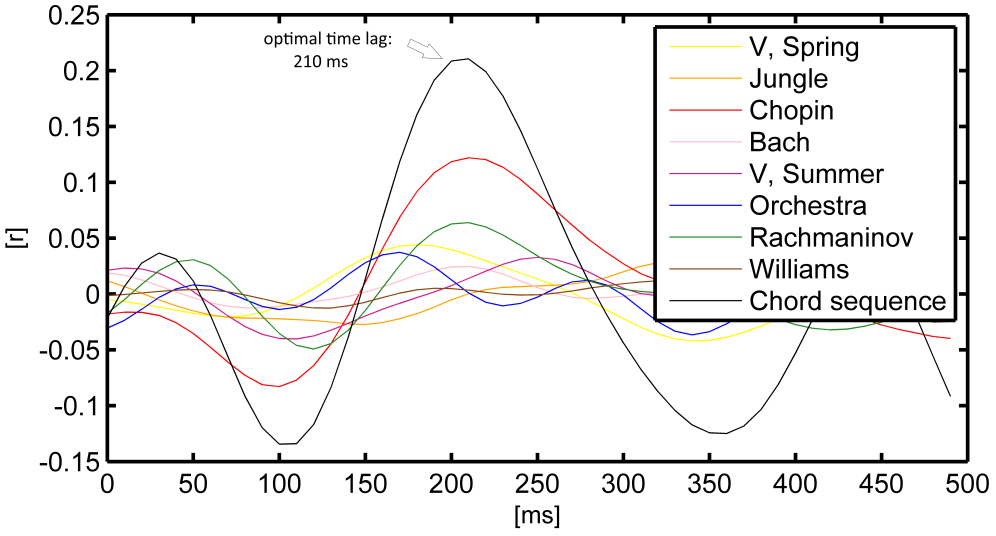

FIGURE 5.17: Cross-correlation sequences for all stimuli for the correlation between the Grand Average of canonical components and the audio power slope.

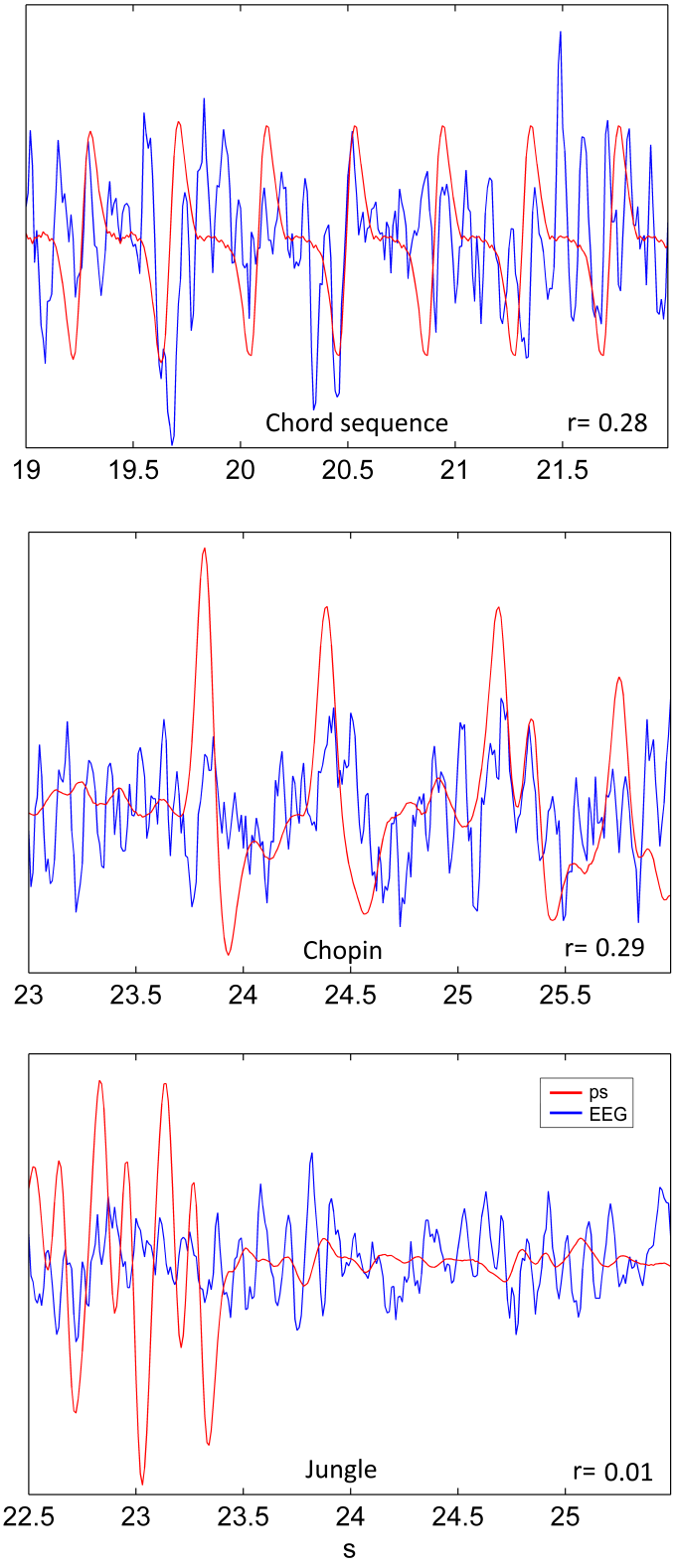

FIGURE 5.18: EEG projections derived with multiway CCA. The same three excerpts from *Chord sequence*, *Chopin* and *Jungle* that were shown in Figure 5.10 showing 3s-long segments of an extracted EEG projection (blue) and the respective audio power slope (red). Here, the EEG projection represents the Grand Average of canonical components that were derived by optimizing Canonical-Inter-Subject-Correlation (CISC). In contrast, the time courses in Figure 5.10 were derived by directly optimizing the correlation between EEG and audio power slope. The correlation coefficients indicate the magnitude of correlation of the extracted EEG projection and the audio power slope for the shown segment of 3 s duration. Correlation was calculated taking into account the optimal time lag of 210 ms.

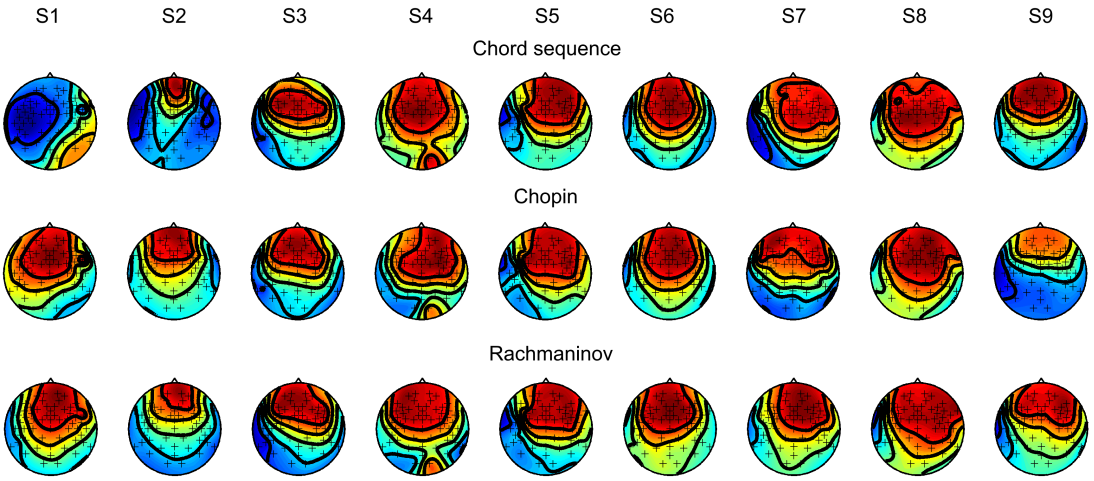

FIGURE 5.19: Scalp patterns of CCA filters for the stimuli with significant mean CISCs (single presentation). Each pattern shows to which extent an electrode reflects the canonical component that maximizes the inter-subject-correlation of EEG.

5.2.3.4 Discussion

In this chapter we have examined the relation between brain signals and music in a setting that (relative to typical highly controlled experimental paradigms) resembles our everyday-life music experiences more closely. Utilizing Cortico-Acoustic Correlation (CACor) we have demonstrated that listeners' EEG signals indeed can synchronize significantly with full-length complex music stimuli. This happened reliably when a rhythmically regular sequence of chords was played, often in romantic piano music and rarely in sound scenes without regular musical structure. Stimuli that consistently led to significant CACor had a high level of Sharpness, were rhythmically simple, and/or were dominated by low-frequency content. The same stimuli produced strongly coordinated ratings in a group of 14 listeners, a finding that provides a tentative link from CACor to conscious experience of music. An additional branch of our analysis showed that even if no stimulus feature is employed to guide the optimization process, but, instead, the criterion of between-subjects similarity, still a distinct reflection of tone onsets can be detected in the EEG.

Methods Our results demonstrate that the proposed regression-based method allows to extract a neural representation of the sequence of note onsets that constitute a complex naturalistic stimulus at the level of single presentations and single subjects. Our stimulus set contained six original pieces of music, a simple chord sequence with repetitive structure, and two non-musical complex stimuli. In all but the non-musical stimuli (*Orchestra* and *Jungle*) the note onset structure of the stimulus was reconstructed from the EEG with significant correlation at least once at single-presentation level and also from the Grand Average EEG features (see Table 5.7). Our results help to confirm onset-related brain responses as one distinct component of the ongoing EEG signal of subjects who listen to complex music. This insight complements previous approaches that aimed at decomposing ongoing EEG with respect to stimulus structure (Cong et al., 2012, Schaefer et al., 2011a, 2009), and, therefore, adds to the growing body of knowledge about how a naturalistic complex music signal is represented in the brain.

The present analysis represents (to our knowledge) the first attempt to apply a sound envelope-related stimulus-reconstruction method in an experimental paradigm where subjects listen without any task to stimulus material that is comparable to music/non-musical sounds that we consume in everyday life.

Neurophysiological interpretation For each subject and stimulus we derived a set of MUSIC components (see 4.2.3 for details) from the spatio-temporal regression patterns with a temporal and spatial dimension each. We found that each of these sets contained a scalp topography that resembles that of the typical N1-P2 complex (as shown in Figure 5.3 for example). We, thus, provide a piece of evidence that the feature that links EEG signal and music stimulus still relies on the N1-P2 complex even in the case of highly variable and complex stimuli and, thus, is in line with previous findings (Kaneshiro et al., 2008, Schaefer et al., 2009). Like in our previous results, the temporal dimension of the selected MUSIC components is highly variable across stimuli, a finding that may relate to the sensitivity of the N1-P2 complex to several auditory features (Meyer et al., 2006, Näätänen and Picton, 1987) and also to attentional state (Choi et al., 2014).

Between-stimulus comparison of CACor The present approach represents an exploratory analysis; therefore, it is not known under which circumstances CACor can be detected. Onset-related brain responses may depend on a multitude of stimulus properties, most probably on the amplitude, tempo, attack, and regularity of the onsets. Technically, their detectability may depend on the number of samples (i.e., stimulus length) and on its stationarity. To gather experience, we applied the proposed method to a range of stimuli from different sound categories.

To assess a stimulus' power to influence the brain signal we calculated CACor scores (see Section 5.2.3.2). The CACor score ranking is led by the semi-musical repetitive *Chord Sequence* (CACor score 22/27), followed by the two romantic piano pieces by *Chopin* and *Rachmaninov*. These were followed by both *Vivaldi* stimuli, then by *Bach* and *Williams*. For the non-music stimuli significant CACor was present in none of the presentations. Remarkably, in the CACor score ranking (Figure 5.11) stimuli from the same category, e.g., both pieces by *Vivaldi* or both romantic piano pieces are adjacent, suggesting that their acoustic/musical structure influences CACor in similar way.

Descriptions of the stimuli with respect to average magnitudes of nine acoustic/musical features revealed that the CACor scores are significantly positively correlated with global measures of Sharpness (see Table 5.8), negatively with Spectral centroid. Sharpness can be described as a measuring how many 'auditory edges' a sound presents (Doelling et al., 2014) and depends on the number and amplitude of peaks in the power slope. Spectral Centroid is related to the perception of brightness. Taken together, our results suggest that significant CACor is detected most reliably in stimuli with high onset density (containing a high number of note onsets) and with onsets that are characterized by a high contrast between baseline and peak sound intensity. Furthermore, lower frequencies (as, e.g., indicated by a relatively low Spectral Centroid) in the stimulus seem to be more effective in 'capturing' the brain. Both, beat density and beat salience have been reported as strongly promoting the experience of 'groove' in songs (Madison et al., 2011). An influential role of energy in low frequency bands in the induction of movement has been reported in Burger et al. (2013). The sharpness of 'auditory edges' has been linked to the strength of stimulus-tracking of cortical oscillatory features and speech intelligibility in Doelling et al. (2014).

Even though the negative correlation of global Fluctuation entropy values and CACor scores are not significant, the relatively strong negative correlation indicates generally lower CACor scores for stimuli with complex rhythmic structure which is in line with results at ERP level reported in Pereira et al. (2014).

As a behavioral counterpart of CACor scores we examined Coordination scores, a measure of how strongly a stimulus effects similar trends in tension ratings in a group of subjects. CACor scores and Coordination Scores were significantly correlated, even though none of the global stimulus descriptions had a significant influence on the Coordination scores. This means that the consistent experience of tension in a group of listeners depends on more complex and variable configurations of musical parameters that, nevertheless, may encompass also those that have been linked to CACor in the present analysis. Thus, stimuli that leave a marked physiological reflection in the EEG, also lead to congruent perception of tension in different subjects. Tentatively, this might provide a link between (low-level) physiological reactions to sounds and the complex concept of musical tension.

Within-stimulus dynamics of CACor, music features and tension ratings Considering the variable surface (Lerdahl and Jackendoff, 1983) of a naturalistic stimulus, it can be assumed that stimuli not only differ with respect to global measures of CACor, but also that the synchronization between brain signal and stimulus varies during the course of the stimulus. The comparison between the dynamics of different acoustic/higher-level music features and time-resolved CACor amounted to a small but distinct set of significant correlation coefficients. In particular, Sharpness, which was one of the main influences on CACor scores found in the between-stimulus comparison of CACor (see Section 5.2.3.3), and Spectral Entropy seem to modulate local within-stimulus CACor. This, again, points to an important role of 'auditory edges' and spectral properties in the cortical processing of note onsets. It demonstrates that even at this fine-grained time scale a direct relation of these properties to brain-stimulus synchronization can be detected. Interestingly, Sharpness was the feature that co-varied with tension ratings most consistently (see Table 5.9), while Sound intensity, Spectral Entropy and Fluctuation Entropy influenced tension ratings only in one stimulus each. This result adds new aspects to previous findings that identified loudness (which is closely related to Sound Intensity) as main factor for the experience of tension several times Farbood and Upham (2013), Lehne et al. (2013b).

A rather general notion of tension has been associated with conflict, instability, or uncertainty (Lehne et al., 2013a). Along these lines, a tentative explanation of the present influence of Sharpness on the experience of musical tension may be formulated: A change from a high level of Sharpness which is characterized by the presence of distinct salient 'auditory edges' to a passage where musical events are less clearly defined can be thought to produce uncertainty in the listener and may lead to an increase in experienced tension. This demonstrates how a stimulus property that can be considered as 'merely' acoustic contributes to an emotion-related aspect of experiencing music.

The set of nine stimuli in this exploratory analysis contained a wide range of different sounds. Of these, two full length romantic piano pieces represented the stimuli for those influences of stimulus structure, both, on brain-stimulus synchronization and tension ratings was detected most consistently. Acoustically, this can be related to the fact that the piano, as a struck string instrument, has a characteristic attack 'thump' (Askenfelt, 1993) and therefore is likely to provide salient onsets that are reflected well in the brain response. Remarkably, the two piano pieces (within our set of stimuli) represent the most complex musical surfaces as (typically for this period) they contain long, lyrical melodies, wide skips, chromaticism, strong contrasts, expressive performance and interpretive freedom (rubato). These highly complex pieces reveal a relation between stimulus structure and physiological measures and behavioral measures most clearly. To a lesser extent, this relation is visible in two movements of a baroque violin concerto, that feature a rhythmic regular pulse and a contrast between 'solo' passages of the solo violin and 'tutti' passages of the orchestra. Taken together, this suggests that structural variance/richness and strong contrasts are aspects of music that give rise to a clear physiological and behavioral reflection of structural elements of music, a finding that represents a strong argument for experimental paradigms using full-length naturalistic music stimuli.

Multiway CCA In general, our results show that CCA retrieves inter-individually consistent EEG components from a group of subjects who listen to natural music, at least for some stimuli and, therefore, contribute to the series of investigations of multi-subject analyses for naturalistic stimuli (Bießmann et al., 2014, Bridwell et al., 2015,

Gaebler et al., 2014, Hasson, 2004, Hasson et al., 2010). Although the significant mean CISCs seem extremely small, our permutation testing procedures have shown that they reflect genuine similarities between subjects' EEG recordings and not spurious correlations. The canonical components seem to be dominated by a strong alpha-like rhythm (see Figure 5.18). Yet, a certain degree of alignment to the stimulus can be observed. Based on these preliminary observations, it would be interesting to examine the role of alpha phase information with respect to synchronization with the stimulus. This might provide complementary aspects to other findings on oscillatory activity in the alpha range during music listening, e.g., on frontal alpha asymmetry distinguishing musical emotions (Schmidt and Trainor, 2001, Trochidis and Bigand, 2012).

In the context of this thesis, the most important result of the CCA analysis is the fact that the canonical components that were obtained using the criterion of inter-subject-correlation and *without imposing any stimulus feature on the optimization process*, clearly seem to be related to tone onsets. This is demonstrated by the significant CACor coefficients of the Grand Average canonical components for the four 'top-ranking' stimuli in our set (see Table 5.11). Since this result relates to the first - and, therefore, predominant - CCA component, our previous suggestions that brain responses to tone onsets are one distinct component of the ongoing EEG signal of subjects who listen to complex music are confirmed strongly.

The scalp topographies that were derived from the CCA filters are consistent across subjects and stimuli, but, with a more frontal distribution differ from the central MUSIC patterns. However, the MUSIC patterns are the results of an additional source reconstruction processing step, a procedure that is not necessary in the case of the present scalp patterns. A relation to the N1-P2 complex is still plausible, in particular, since in several stimuli the optimal time lag for CACor (210 ms) was in the typical range of the P2 component.

In summary, the CCA-based extension of our analysis has led to two important conclusions: first, it has (besides other precautions, such as cross-validation and permutation-testing) proven that the reflection of tone onsets in the listener's EEG is a genuine physiological reaction. Secondly, it has demonstrated (in an 'unsupervised' manner) that the most reliable common feature of the listener's (broadband) EEG is the synchronization to tone onsets.

Limitations The present results are a proof-of-concept that multivariate methods of EEG analysis can achieve considerable advances in extracting onset-related brain responses from the ongoing EEG, enabling more naturalistic experimental paradigms. Notwithstanding, several issues call for further exploration. Our results have demonstrated that complex stimuli with different characteristics vary with respect to the consistency of the occurrence of CACor in a group of subjects. The present analysis has identified acoustic properties that are related to these differences. However, the variance of CACor between presentations within subjects that occurred in artifact-cleaned data has not been explained yet. Onset-related ERPs are known to be influenced by a range of subject variables, some of them being situation-dependent, such as attentional state, vigilance or familiarity (Hillyard et al., 1973, Low et al., 2007). In principle, the observed differences between presentations might reflect changes in such user variables. If so, CACor might even be a means of monitoring these mental states. A systematic investigation is needed to explore this.

The present analysis revealed that EEG signals may significantly reflect the onset structure of music, and, that, if this is the case, also the dynamics of tension ratings are consistent for a group of subjects. At a more detailed level (within stimuli), however, we found only weak relations between CACor and reported tension. This means that the onset structure of music can 'drive' cortical activity, but that it is not clear whether and how this impacts on conscious experience of musical tension.

5.2.3.5 Future Work

Beyond the concept perceived tension, further measures related to experience of music may be required for investigating this aspect in detail. Behaviorally reported qualities describing listening experience, such as immersion or involvement, could be evaluated in future experiments. In addition, it would be interesting to probe whether stimuli effecting significant CACor also synchronize other physiological parameters. In particular, respiration patterns have been found to modulated by musical tempo Bernardi et al. (2006), Gomez and Danuser (2004, 2007). Based on these findings, an examination of between-subject synchronization of respiration or of synchronization of a listener's breathing to the stimulus could represent a valuable extension of the present analysis.

5.2.3.6 Conclusion

The present results demonstrate that, in principle, the sequence of note onsets which constitutes an individual presentation of an original complex piece of music, e.g., a piece of piano music from the romantic repertoire, can be 'reconstructed' from the listener's EEG using spatio-temporal regression-derived filters. The distribution of significant Cortico-Acoustic Correlation coefficients suggests that the presence of a simple salient beat and percussive elements enhances the synchronization of a listener's brain signals to music. The same acoustic properties led to (relatively) strongly coordinated behavioral responses. The proposed approach represents a promising first attempt to connect detailed accounts of the cortical processing of natural music with behavioral reports of experiencing music.

Chapter 6

General Discussion and Conclusion

In this thesis we present methodological approaches for investigating the perception of natural music by combining the analysis of electrophysiological signals recorded from the listening brain with the analysis of music signals and of behavioral reports on the musical experience.

As we have pointed out in Chapter 1, ecological validity is a highly desirable property of music-related experimental paradigms for a number of reasons. Yet, our overview about the state-of-the-art of research shows that the number of studies is small, most likely due to data-analytical and experimental challenges. In particular, the relationship between electrophysiological signals recorded from the listener and the music signal has not received much attention, although these signals clearly share some properties and, thus, offer opportunities for direct comparison.

Therefore, the main goal of this thesis was to explore what insight can be gained from electrophysiological signals recorded while listening to music.

6.1 Investigating music processing with ECoG

In the first study we took advantage of invasive electrophysiological recordings (ECoG) which, by virtue of their properties, are highly favorable for single-trial analysis. We have expanded the degree of specificity of existing approaches towards examining the cortical reflection of multidimensional structure of music in individual participants' brain signals for the single presentations of a stimulus. Directly addressing the problem of the typically non-orthogonal structure of multi-dimensional descriptions of music, we have proposed the partial correlation coefficient as a simple, but effective, method to differentially examine neural correlates of the processing of multiple features of music. The results obtained demonstrate this approach's contribution towards connecting music signal analysis techniques and brain data analysis.

Our very detailed, individual results suggest an influence of context on the representation of the features that characterize a complex auditory stimulus - an insight that could only be obtained using natural music and speech for stimulation.

However, one conceptual problem of this approach has to be acknowledged: the extreme level of specificity of the results obtained in this analysis naturally makes it hard to generalize our findings. Therefore, our results only provide a small glimpse on the processing of one specimen of music and language each. This means that more examples would be needed to draw more general conclusions, and to confirm the advantages of the proposed method.

6.2 Investigating music processing with EEG

From the rare example of a electrocorticographic study we have progressed towards the much more widely applicable recording modality of (scalp) EEG.

6.2.1 Method

Recognizing the susceptibility of EEG to the impulses that carry the temporal structure of music, we have explored how this stimulus-related information can be extracted from the EEG and whether it can be employed to investigate several aspects of music perception. To this end, we have proposed Linear Ridge Regression with the audio power slope as a target function which is a variant of state-of-the-art techniques for stimulus reconstruction from speech-processing research (see Chapter 4.1). From the perspective of the generative model of EEG this method can be described as a way of supervised EEG decomposition that makes use of the stimulus structure and thus fuses EEG analysis and music signal analysis analysis – one main aim of this thesis. With the complementary multiway-CCA analysis we have given an example of a 'less supervised' related technique for multi-subject analysis that confirmed our previous findings.

The proposed analysis pipeline is assembled from well-established methods (Müller et al., 2003) with relatively simple underlying ideas, such as Ridge Regression, Single Value Decomposition (SVD) and Least-Squares Optimization. Our results, however, show that enriching these 'old' mechanisms with specific extensions allows obtaining significant results at single-trial level in a challenging setting. In particular, the combination of multivariate EEG analysis (that in itself is highly effective for enhancing the signal-to-noise-ratio) with temporal embedding adds to the strength of the method as it enables the fine-tuning of spatio-temporal filters to, both, stimulus characteristics and individual differences in latencies of brain responses. Furthermore, relating brain responses to a stimulus' power slope (instead of the sound envelope that numerous approaches focus on) exploits the brain's sensitivity to change.

Here, we derived a *backward* mapping between the brain signal and the sound envelope, in contrast to several *forward* mapping approaches that are described in the literature (see Section 4.1). We would like to point out that in the EEG setting *backward* modeling has one key advantage over *forward* modeling: In *backward* modeling the full set of electrodes is integrated in order to learn the mapping between EEG and target function, a technique that is beneficial to the signal-to-noise ratio and provides us with information about the scalp distribution of relevant information (Parra et al., 2005). In contrast, when modeling in the reverse direction, either a specific electrode has to be selected or the multichannel signal has to be reduced to a single source. This step needs further assumptions, while in the backward modeling it is integrated. Typically, *forward* modeling

derives a convolution model that can be viewed as an estimate of the impulse response of the auditory system. With the proposed extension of the Multiple Signal Classification technique for decomposing the regression patterns we have shown that it is possible to obtain a comparable model in a *backward* modeling approach and, thus, demonstrated that the advantages of *forward* and *backward* modeling can be combined in this case.

Technically, the application of this method is not restricted to a particular type of auditory stimulus, since the power slope can be derived in a simple procedure from any audio waveform. In principle, this technique may be useful in a range of scenarios related to onset-ERPs as it mitigates the demand for a high number of trials that is typical for EEG averaging techniques. It is applicable at single-subject and single-stimulus presentation level and it is appropriate for complex long stimuli. Since in the extracted EEG projections the time-resolution of the EEG is preserved, this method allows for subsequent investigations at several time scales. It may open up a new avenue for transferring findings that have been made with respect to simple stimuli and at the group level to the level of subject-individual analysis and towards more naturalistic stimuli.

6.2.2 Application examples

Progressing from simple tone sequences to natural music we have applied the proposed method in several typical scenarios. To begin with, we have directly compared neural signatures derived with the proposed method to (conventionally) averaged ERPs. This comparison helped to establish a link from the brain response we extracted from the EEG to the N1-P2 ERP component. In the next study we applied our method in a music-related variant of the 'cocktail party problem' where a simplified music-like stimulus was presented either in solo (single voice) versions or in an ensemble version. Our results to some extent demonstrated that a listener's EEG can indeed be decomposed into identifiable sources that correspond to the voices of a multi-voiced piece of music (if the single voices are known). They further showed an effect of frequency masking between the three voices. On the one hand, this shows that non-linear properties of the auditory system that are well known and included in auditory models can be demonstrated in naturalistic listening scenarios. On the other hand, it also proves the existence of imprecisions/shortcomings in the simple linear mapping we assumed.

Using a broad range of musical pieces and other naturalistic sounds we have successfully approached a 'pure' decoding task at the level of single presentations of complex music and non-musical sounds. In addition, we have demonstrated that even without explicit integration of the sound envelope into the optimization process, the most consistently occurring stimulus-related EEG component is related to tone onsets. Using audio descriptors we have explained to some extent how stimuli differ in their likelihood to 'capture' the brain both, at a global and a local level. Finally, we have complemented our findings on the relation between brain signal and music signal with behavioral data. A main insight from this exploratory analysis is that stimuli with certain acoustic properties, such as a high level of Sharpness and a low Spectral Centroid yield not only a tight synchronization in the EEG with the stimulus, but also coordinate behavioral responses in a different group of subjects. This finding supports an relevant link between acoustic properties and, both, behavioral and physiological responses to music.

6.2.3 Lessons learned

Looking back at our initial research questions we conclude that for one generic aspect of music processing, responses to tone onsets, we succeeded to advance the analysis of single-presentation brain signals from isolated sounds towards more ecologically valid, continuous music excerpts. The proposed method, in principle, is a way to forego conventional averaging techniques. However, a few points need consideration:

Firstly, one could argue that the information we extract at single-trial level (e.g., a single CACor coefficient) is of limited use and that the main results of this thesis were derived by looking at distributions of CACor coefficients, such as CACor scores. Thus, in order to utilize the proposed method as a tool to characterize determinants of experiencing music we remain to rely on conventional experimental practices of contrasting controlled listening conditions and populations of subjects. Nevertheless, the advance from simple sounds to continuous music represents a qualitative change. It opens up new avenues for researching aspects of experiencing music that cannot possibly be accessed with short, simplified stimuli, such as affective responses, personal factors, such as familiarity or liking, as well as aspects related to the processing of musical macro-structure.

Furthermore, the EEG feature we have been focusing on is a sensory trace of the stimulus. All inferences about listening to and experiencing music are indirect and operate based on the assumption that this sensory trace is modulated by certain aspects of the listening experience, such as mental states, user variables or depend on the stimulus. The N1-P2 components to tone onsets have been a promising and useful 'workhorse' for our purpose known to be sensitive to a number of aspects of listening (see Chapter 4.1). At the same time, this poses a conceptual problem, as effects on N1-P2 responses can be hard to pinpoint to their specific origin. Rigid experimental control of the listening situation is required to draw specific conclusions from here. The unexplained variance that we observed in CACor coefficients in our least controlled listening paradigm in experiment 2, may relate to this.

In summary, our approach enables to approximate ecological validity in terms of stimulus material. The additional, simultaneous approximation of an ecologically valid listening situation (unconstrained music listening like in everyday life) adds considerably to the recognized conceptual problem, as the interpretability of results strongly depends on clearly defined experimental conditions.

6.2.4 Future Work

An important part of our work was to identify technical problems, e.g., how to deal with auto-correlation, propose solutions and discuss their specific advantages and drawbacks. As a result we have assembled a framework of rather robust analysis steps of analysis that allows examining the relation between music signals, EEG signals and behavioral data at several time scales. Obviously, there exist a number of variants, extensions and refinements that could enrich this framework. For instance, for modeling the mapping between EEG data and music stimulus other (non-linear) methods could be employed, e.g. including auditory modeling in the extraction of the sound envelope as proposed in (Aiken and Picton, 2008). Furthermore, there are numerous ways of varying the music signal analysis. Also, other physiological parameters, such as information on heart rate or respiration, could be added in order to obtain a comprehensive overview on physiological

and behavioral responses to music. Finally, tension ratings, as described in Chapter 2.2.3, are only one way of collecting behavioral responses to music. In summary, this thesis presents first examples how different perspectives on listening to music can be combined to learn about music perception in a comprehensive way. Most importantly, with the Cortico-Acoustic Correlation analysis we have proposed a tool that allows combining indices of stimulus-brain response strength with other measures at several time scales.

All scalp EEG analyses presented in this thesis were restricted to the broadband EEG signal and did not consider oscillatory brain activity, e.g., bandpower modulations. One reason for this is that in the preliminary stages of analysis power modulations of specific frequency bands (e.g., theta, alpha, beta band) did not reveal any consistent relation to the music stimulus that encouraged further investigations, a finding that recently has been confirmed in Jäncke et al. (2015). However, this thesis has revealed one feature of music that is consistently reflected in the EEG and it has helped to identify a set of stimuli that were very effective in stimulating the brain and coordinating behavior. Therefore, from today's perspective and based on the outcome of this work, an investigation with respect to oscillatory features could be a tangible topic for the future, in particular, since highly effective multi-variate methods for the extraction of oscillatory neural activity have been refined in the meantime (Dähne et al., 2014a). Instead of trying to relate neural activity directly to the stimulus, it may be here equally insightful to access inter-individually consistent phenomena in the listening process through multi-subject analyses.

6.2.5 Scenarios for practical application

The effect oft directed attention on neural responses to (concurrent) auditory stimuli has been put to practical use in a great number of Brain-Computer-Interfacing (BCI) applications (Choi et al., 2013, Hill et al., 2012, Lopez-Gordo et al., 2012, Treder et al., 2014). Our method (and applications thereof) can be regarded as step towards a more 'musical' BCI that, using the representation of the sound envelope in the EEG, allows for far more natural stimulus material than that of typical P3-based paradigms or those based on the Auditory Steady State Response (ASSR). Recently, similar approaches built on sound envelope-responses have produced promising results in the speech domain (Kerlin et al., 2010, Lopez-Gordo et al., 2012, O'Sullivan et al., 2014) and also in the music domain (Choi et al., 2013).

In a similar way, to the ability to read the user's target of attention, EEG reflections of the sound envelope could be used in hearing aids to provide selective enhancement of attended sounds in multi-speaker or multi-stream environments. Moreover, the representation of the sound envelope in the EEG could be a tool to guide hearing aid fitting. In particular, such an application might be useful for adjusting hearing aid function for listening to music or even for particular requirements, as, e.g., those of musicians with hearing loss. In general, also in clinical settings diagnostics have moved towards more complex sound paradigms, e.g., to evaluate cochlea implant users (Koelsch et al., 2004, Timm et al., 2014). In the future, brain responses to complex music might also be of interest in this field.

Finally, the proposed method might open up novel possibilities for EEG-based assessment of listening experience, such as listener engagement, with potential commercial applications, e.g., in the field of marketing.

6.3 Conclusion

Electrophysiological measurements enable to investigate the processing of complex music with considerable sensitivity and a fine level of detail. In combination with music analysis and behavioral techniques they represent a promising tool for bridging the gap between brain studies that often are concerned with basic aspects of music processing and behavioral studies that present complex music, but rely on the participants' (often retrospective) reports of their listening experience. In this thesis we have provided a number of methodological advances that may help to utilize the relation between brain signal and music signal for further understanding of perceptional processes. We have given a number of examples of how these advances can be employed to study the perception of natural music and have identified relevant conceptual problems that require consideration. While our findings still remain to be confirmed by a more comprehensive selection of music stimuli, we are confident that they can contribute to a better understanding of the process leading from soundwave to personal experience.

Bibliography

Abdi, H. (2007). Part (semi partial) and partial regression coefficients. In Salkind, J., editor, *Encyclopedia of Measurements and Statistics*. SAGE Publications.

Abecasis, D., Brochard, R., Granot, R., and Drake, C. (2005). Differential brain response to metrical accents in isochronous auditory sequences. *Music Perception*, 22(3):549–562.

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., and Menon, V. (2011). Decoding Temporal Structure in Music and Speech Relies on Shared Brain Resources but Elicits Different Fine-Scale Spatial Patterns. *Cerebral Cortex*, 21(7):1507–1518.

Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., and Menon, V. (2013). Inter-subject synchronization of brain responses during natural music listening. *European Journal of Neuroscience*, 37(9):1458–1469.

Aiken, S. J. and Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiology and Neurotology*, 11(4):213–232.

Aiken, S. J. and Picton, T. W. (2008). Human Cortical Responses to the Speech Envelope. *Ear and Hearing*, 29(2):139–157.

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., and Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, 59(4):3677–3689.

Angluin, D. and Smith, C. (1983). Inductive Inference: Theory and Methods. *ACM Computing Surveys*, 15(3):237–269.

Askenfelt, A. (1993). Observations on the transient components of the piano tone. *STL-QPSR*, 34(4):15–22.

Bartz, D. and Müller, K. R. (2013). Generalizing analytic shrinkage for arbitrary covariance structures. In *Advances in neural information processing*, pages 1869–1877.

Baumann, S., Meyer, M., and Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *Journal of Cognitive Neuroscience*, 20(12):2238–2249.

Bernardi, L., Porta, C., and Sleight, P. (2006). Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. *Heart*, 92(4):445–452.

Bhattacharya, J., Petsche, H., and Pereda, E. (2001a). Interdependencies in the spontaneous EEG while listening to music. *International Journal of Psychophysiology*, 42(3):287–301.

Bhattacharya, J., Petsche, H., and Pereda, E. (2001b). Long-range synchrony in the gamma band: role in music perception. *The Journal of Neuroscience*, 21(16):6329.

Biesmans, W., Vanthornhout, J., Wouters, J., Moonen, M. F. T., and Bertrand, A. (2015). Comparison of speech envelope extraction methods for EEG-based auditory attention detection in a cocktail party scenario. *Internal Report KU Leuven STADIUS)*.

Bießmann, F. (2011). *Data-driven analysis for multimodal neuroimaging*. PhD thesis, Technische Universität Berlin, Berlin and Germany.

Bießmann, F., Gaebler, M., Lamke, J., Ju, S., Wallraven, C., and Müller, K. R. (2014). Data-driven multisubject neuroimaging analyses for naturalistic stimuli. In *2014 International Workshop on Pattern Recognition in Neuroimaging*, pages 1–4. IEEE.

Bießmann, F., Meinecke, F. C., Gretton, A., Rauch, A., Rainer, G., Logothetis, N., and Müller, K. R. (2010). Temporal kernel CCA and its application in multimodal neuronal data analysis. *Machine Learning*, 79(1-2):5–27.

Bießmann, F., Plis, S., Meinecke, F. C., Eichele, T., and Müller, K. (2011). Analysis of multimodal neuroimaging data. *Biomedical Engineering, IEEE Reviews in*, 4:26–58.

Billings, C. J., Tremblay, K. L., and Miller, C. W. (2011). Aided cortical auditory evoked potentials in response to changes in hearing aid gain. *International Journal of Audiology*, 50(7):459–467.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K. R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage*, 56(2):814–825.

Blood, A. and Zatorre, R. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, 98(20):11818.

Borga, M. (1998). *Learning multidimensional signal processing*. PhD thesis, Universitet Linköping, Department of Electrical Engineering, Linköping and Sweden.

Borga, M., Landelius, T., and Knutsson, H. (1997). A unified approach to PCA, PLS, MLR and CCA, Technical Report. Linköping.

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., and Tervaniemi, M. (2011). A functional MRI study of happy and sad emotions in music with and without lyrics. *Frontiers in Psychology*, 2(308).

Brattico, E., Bogert, B., and Jacobsen, T. (2013). Toward a neural chronometry for the aesthetic experience of music. *Frontiers in Psychology*, 4(206).

Brattico, E., Jacobsen, T., Baene, W. d., Glerean, E., and Tervaniemi, M. (2010). Cognitive vs. affective listening modes and judgments of music-An ERP study. *Biological Psychology*.

Brattico, E., Tervaniemi, M., Naatanen, R., and Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Research*, 1117(1):162–174.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound.* MIT press.

Bridwell, D. A., Roth, C., Gupta, C. N., and Calhoun, V. D. (2015). Cortical Response Similarities Predict which Audiovisual Clips Individuals Viewed, but Are Unrelated to Clip Preference. *PloS One*, 10(6):e0128833.

Brown, E. C., Muzik, O., Rothermel, R., Juhász, C., Shah, A. K., Fuerst, D., Mittal, S., Sood, S., and Asano, E. (2013). Evaluating Signal-Correlated Noise as a Control Task with Language-Related Gamma Activity on Electrocorticography. *Clinical Neurophysiology*, 1257:1312–1323.

Brown, S., Martinez, M., and Parsons, L. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport*, 15(13):2033.

Brunet, N., Vinck, M., Bosman, C. A., Singer, W., and Fries, P. (2014). Gamma or no gamma, that is the question. *Trends in cognitive sciences*, 18(10):507–509.

Brunner, P., Ritaccio, A. L., Emrich, J. F., Bischof, H., and Schalk, G. (2011). Rapid communication with a "P300" matrix speller using electrocorticographic signals (ECoG). *Frontiers in Neuroscience*, 5(5).

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., and Toiviainen, P. (2013). Influences of rhythm-and timbre-related musical features on characteristics of music-induced movement. *Frontiers in Psychology*, 4:183.

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M.-H., and McAdams, S. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, 18(12):1959–1972.

Caclin, A., Giard, M. H., Smith, B. K., and McAdams, S. (2007). Interactive processing of timbre dimensions: A Garner interference study. *Brain Research*, 1138:159–170.

Campbell, J. D., Cardon, G., and Sharma, A. (2011). Clinical application of the P1 cortical auditory evoked potential biomarker in children with sensorineural hearing loss and auditory neuropathy spectrum disorder. In *Seminars in hearing : NIH Public Access*, volume 32, page 147.

Carrus, E., Koelsch, S., and Bhattacharya, J. (2011). Shadows of music–language interaction on low frequency brain oscillatory patterns. *Brain and language*, 119(1):50–57.

Chakalov, I., Draganova, R., Wollbrink, A., Preissl, H., and Pantev, C. (2013). Perceptual organization of auditory streaming-task relies on neural entrainment of the stimulus-presentation rate: MEG evidence. *BMC Neuroscience*, 14(1):120.

Chapin, H., Jantzen, K., Kelso, J., Steinberg, F., and Large, E. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PloS One*, 5(12):e13812.

Chew, E. (2000). *Towards a mathematical model of tonality.* PhD thesis, Massachusetts Institute of Technology, Boston and USA.

Choi, I., Le Wang, L., and Shinn-Cunningham, B. (2014). Individual differences in attentional modulation of cortical responses correlate with selective attention performance. *Hearing Research*, 314:10–19.

Choi, I., Rajaram, S., Varghese, L. A., and Shinn-Cunningham, B. (2013). Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography. *Frontiers in Human Neuroscience*, 7(115).

Cong, F., Phan, A. H., Zhao, Q., Nandi, A. K., Alluri, V., Toiviainen, P., Poikonen, H., Huotilainen, M., Cichocki, A., and Ristaniemi, T. (2012). Analysis of ongoing EEG elicited by natural music stimuli using nonnegative tensor factorization. In *Signal Processing Conference (EUSIPCO)*, pages 494–498.

Correa, D., Saito, J., and Costa, L. (2010a). Musical genres: beating to the rhythms of different drums. *New Journal of Physics*, 12:053030.

Correa, N. M., Eichele, T., Adalı, T., Li, Y., and Calhoun, V. D. (2010b). Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI. *NeuroImage*, 50(4):1438–1445.

Coutinho, E. and Cangelosi, A. (2011). Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion*, 11(4):921–937.

Crawley, E. J., Acker-Mills, B. E., Pastore, R. E., and Weil, S. (2002). Change detection in multi-voice music: the role of musical structure, musical training, and task demands. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2):367.

Crone, N. E., Korzeniewska, A., Supratim, R., and Franaszczuk, P. J. (2009). Cortical function mapping with intracranial EEG. In *Quantitative EEG analysis methods*, pages 369–399.

Crone, N. E., Sinai, A., and Korzeniewska, A. (2006). High-frequency gamma oscillations and human brain mapping with electrocorticography. *Progress in brain research*, 159:275–295.

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

Dähne, S. (2015). *Decomposition Methods for the Fusion of Multimodal Functional Neuroimaging Data: PhD Thesis*. PhD thesis, Technische Universität Berlin, Berlin.

Dähne, S., Bießmann, F., Meinecke, F. C., Mehnert, J., Fazli, S., and Muller, K. (2013). Integration of multivariate data streams with bandpower signals. *Multimedia, IEEE Transactions on*, 15(5):1001–1013.

Dähne, S., Bießmann, F., Samek, W., Haufe, S., Goltz, D., Gundlach, C., Villringer, A., Fazli, S., and Muller, K. (2015). Multivariate Machine Learning Methods for Fusing Multimodal Functional Neuroimaging Data. *Proceedings of the IEEE, accepted.*

Dähne, S., Meinecke, F. C., Haufe, S., Höhne, J., Tangermann, M., Müller, K., and Nikulin, V. V. (2014a). SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *NeuroImage*, 86:111–122.

Dähne, S., Nikulin, V. V., Ramírez, D., Schreier, P. J., Müller, K., and Haufe, S. (2014b). Finding brain oscillations with power dependencies in neuroimaging data. *NeuroImage*, 96:334–348.

Daikoku, T., Ogura, H., and Watanabe, M. (2012). The variation of hemodynamics relative to listening to consonance or dissonance during chord progression. *Neurological research*, 34(6):557–563.

Darvas, F., Pantazis, D., Kucukaltun-Yildirim, E., and Leahy, R. (2004). Mapping human brain function with MEG and EEG: methods and validation. *NeuroImage*, 23:289–299.

Deike, S., Gaschler-Markefski, B., Brechmann, A., and Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport*, 15(9):1511–1514.

del Bimbo, A., Chang, S. F., Smeulders, A., Cannam, C., Landone, C., and Sandler, M. (2010). Sonic Visualiser. In *Proceedings of the international conference on Multimedia - MM '10*, page 1467. ACM Press.

Derix, J., Iljina, O., Weiske, J., Schulze-Bonhage A., Aertsen, A., and Ball, T. (2014). From speech to thought: the neuronal basis of cognitive units in non-experimental, real-life communication investigated using ECoG. *Frontiers in Human Neuroscience*, 8(383).

Deutsch, D. (2013). *Psychology of music*. Elsevier.

Dimitrijevic, A., Michalewski, H. J., Zeng, F. G., Pratt, H., and Starr, A. (2008). Frequency changes in a continuous tone: auditory cortical potentials. *Clinical Neurophysiology*, 119(9):2111–2124.

Ding, N. and Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, 109(29):11854–11859.

Ding, N. and Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, 107(1):78–89.

Ding, N. and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Frontiers in Human Neuroscience*, 8(311).

Dmochowski, J. P., Sajda, P., Dias, J., and Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention–a possible marker of engagement? *Frontiers in Human Neuroscience*, 6(112).

Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, 85:761–768.

Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, 95:2670.

Eerola, T., Lartillot, O., and Toiviainen, P. (2009). Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models. In *ISMIR*, pages 621–626.

Einhäuser, W. and König, P. (2010). Getting real—sensory processing of natural stimuli. *Current Opinion in Neurobiology*, 20(3):389–395.

Farbood, M. (2006). *A quantitative, parametric model of musical tension.* PhD thesis, Massachusetts Institute of Technology.

Farbood, M. (2012). A quantitative, parametric model of musical tension. *Music Perception*, 29(4):387–428.

Farbood, M. and Schoner, B. (2009). Determining Feature Relevance in Subject Responses to Musical Stimuli. *Mathematics and Computation in Music*, pages 115–129.

Farbood, M. and Upham, F. (2013). Interpreting expressive performance through listener judgments of musical tension. *Frontiers in Psychology*, 4(998).

Fredrickson, B. L. and Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of personality and social psychology*, 65(1):45.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., and Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, 129(10):2593–2608.

Fujioka, T., Trainor, L., Ross, B., Kakigi, R., and Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, 16(6):1010–1021.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of Cognitive Neuroscience*, 17(10):1578–1592.

Gabrielsson, A. and Lindström, E. (2001). The influence of musical structure on emotional expression. In Sloboda, J., editor, *Music and emotion: Theory and research. Series in affective science*, page 487 pp. Oxford University Press, New York and NY and US.

Gabrielsson, A. and Lindström E. (2010). The role of structure in the musical expression of emotions. In Juslin, P. N. and Sloboda, J. A., editors, *Handbook of music and emotion: Theory, research, applications*, pages 367–400. Oxford University Press.

Gaebler, M., Bießmann, F., Lamke, J. P., Müller, K. R., Walter, H., and Hetzer, S. (2014). Stereoscopic depth increases intersubject correlations of brain networks. *NeuroImage*, 100:427–434.

Geisser, S. (1993). *Predictive inference: An introduction*, volume 55 of *Monographs on statistics and applied probability*. Chapman & Hall, New York.

Goebl, W., Dixon, S., and Schubert, E. (2014). Quantitative Methods: Motion Analysis, Audio Analysis, and Continuous Response Techniques. In *Expressiveness in music performance: Empirical approaches across styles and cultures*, pages 221–238. Oxford University Press, Oxford and UK.

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., and Simon, J. Z. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, 77(5):980–991.

Gomez, P. and Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, 53(2):91–103.

Gomez, P. and Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, 7(2):377.

Goodglass, H., Kaplan, E., and Barresi, B. (1983). *BDAE: The Boston Diagnostic Aphasia Examination*. Lea and Febiger, Philadelphia.

Gordon, J. W. (1987). The perceptual attack time of musical tones. *The Journal of the Acoustical Society of America*, 82(1):88.

Goydke, K., Altenmüller, E., Möller, J., and Münte, T. (2004). Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Cognitive Brain Research*, 21(3):351–359.

Grahn, J. A. and Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *The Journal of Neuroscience*, 29(23):7540–7548.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *The Journal of Neuroscience*, 25(22):5382–5388.

Halpern, A., Martin, J., and Reed, T. (2008). An ERP study of major-minor classification in melodies. *Music Perception*, 25(3):181–191.

Hamamé, C. M., Vidal, J. R., Perrone-Bertolotti, M., Ossandón, T., Jerbi, K., Kahane, P., Bertrand, O., and Lachaux, J. P. (2014). Functional selectivity in the human occipitotemporal cortex during natural vision: Evidence from combined intracranial EEG and eye-tracking. *NeuroImage*, 95:276–286.

Harte, C., Gasser, M., Sandler, and M. (2006). Detecting harmonic change in musical audio. In *AMCMM '06 Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, pages 21–26.

Hasson, U. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science*, 303(5664):1634–1640.

Hasson, U. and Honey, C. J. (2012). Future trends in Neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, 62(2):1272–1278.

Hasson, U., Malach, R., and Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences*, 14(1):40–48.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., and Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96–110.

Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., and Ackermann, H. (2012). Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. *Psychophysiology*, 49(3):322–334.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *The American Journal of Psychology*, pages 246–268.

Higuchi, M. K., Kodama, K., Fornari, J., Del Ben, C. M., Graeff, F. G., and Leite, J. P. (2011). Reciprocal modulation of cognitive and emotional aspects in pianistic performances. *PloS One*, 6(9):e24437.

Hill, K. T., Bishop, C. W., and Miller, L. M. (2012). Auditory grouping mechanisms reflect a sound's relative position in a sequence. *Frontiers in Human Neuroscience*, 6(158).

Hill, K. T., Lal, T. N. B. K., Birbaumer, N., and Schölkopf, B. (2004). An auditory paradigm for brain-computer interfaces. *Advances in neural information processing systems*, (17):569–576.

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

Höhne, J., Schreuder, M., Blankertz, B., and Tangermann, M. (2011). A novel 9-class auditory ERP paradigm driving a predictive text entry system. *Frontiers in Neuroscience*, 5(99).

Höller, Y., Thomschewski, A., Schmid, E. V., Höller, P., Crone, J. S., and Trinka, E. (2012). Individual brain-frequency responses to self-selected music. *International Journal of Psychophysiology*, 86(3):206–213.

Horton, C., D'Zmura, M., and Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *Journal of Neurophysiology*, 109(12):3082–3093.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, (28):321–377.

Hyde, K. L., Peretz, I., and Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, 46(2):632–639.

Istók, E., Brattico, E., Jacobsen, T., Ritter, A., and Tervaniemi, M. (2013). 'I love Rock 'n'Roll'—Music genre preference modulates brain responses to music. *Biological Psychology*, 92(2):142–151.

Jackendoff, R. and Lerdahl F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, 100(1):33–72.

Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*, 7(2):153–164.

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, 19(11):2579.

Janata, P., Tillmann, B., and Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective, & Behavioral Neuroscience*, 2(2):121–140.

Jäncke, L., Kühnis, J., Rogenmoser, L., and Elmer, S. (2015). Time course of EEG oscillations during repeated listening of a well-known aria. *Frontiers in Human Neuroscience*, 9(401).

Jasper, H. and W., P. (1949). Electrocorticograms in man: effect of voluntary movement upon the electrical activity of the precentral gyrus. *Archiv für Psychiatrie und Nervenkrankheiten*, 183(1-2):163–174.

Jensen, O. and Colgin, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends in cognitive sciences*, 11(7):267–269.

Jentschke, S., Friederici, A. D., and Koelsch, S. (2014). Neural correlates of music-syntactic processing in two-year old children. *Developmental Cognitive Neuroscience*, 9:200–208.

Jongsma, M., Desain, P., and Honing, H. (2004). Rhythmic context influences the auditory evoked potentials of musicians and nonmusicians. *Biological Psychology*, 66(2):129–152.

Juslin, P. N. and Sloboda, J. A., editors (2010). *Handbook of music and emotion: Theory, research, applications*. Oxford University Press.

Kaneshiro, B. B., Dmochowski, J. P., Norcia, A. M., and Berger, J. (2008). Toward an objective measure of listener engagement with natural music using inter-subject EEG correlation. *Jodi*, 4:24.

Kanoh, S., Miyamoto, K. I., and Yoshinobu, T. (2008). A brain-computer interface (BCI) system based on auditory stream segregation. In *Engineering in Medicine and Biology Society, 2008. 30th Annual International Conference of the IEEE*, pages 642–645.

Kendall, M. G., Stuart, A., and Ord, J. K. (1973). *Inference and Relationship*, volume 2 of *The Advanced Theory of Statistics*. Griffin, London, 3 edition.

Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional Gain Control of Ongoing Cortical Speech Representations in a 'Cocktail Party'. *Journal of Neuroscience*, 30(2):620–628.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451.

Khalfa, S., Schon, D., Anton, J., and Liégeois-Chauvel, C. (2005). Brain regions involved in the recognition of happiness and sadness in music. *Neuroreport*, 16(18):1981.

Kim, C., Lee, S., Kim, J., Seol, J., Yi, S., and Chung, C. (2014). Melody effects on ERANm elicited by harmonic irregularity in musical syntax. *Brain Research*, 1560:36–45.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, 7(3):302–307.

Koelsch, S. and Mulder, J. (2002). Electric brain responses to inappropriate harmonies during listening to expressive music. *Clinical Neurophysiology*, 113(6):862–869.

Kong, Y., Mullangi, A., and Ding, N. (2014). Differential modulation of auditory responses to attended and unattended speech in different listening conditions. *Hearing Research*, 316:73–81.

Koskinen, M., Viinikanoja, J., Kurimo, M., Klami, A., Kaski, S., and Hari, R. (2013). Identifying fragments of natural speech from the listener's MEG signals. *Human Brain Mapping*, 34(6):1477–1489.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception*, pages 401–432.

Krumhansl, C. L. and Toiviainen, P. (2001). Tonal Cognition. *Annals of the New York Academy of Sciences*, 930(1):77–91.

Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., Schalk, G., and Rodriguez-Fornells, A. (2013). The Tracking of Speech Envelope in the Human Cortex. *PloS One*, 8(1):e53398.

Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., Patterson, R. D., Howard, M. A., Friston, K. J., and Griffiths, T. D. (2011). Predictive coding and pitch processing in the auditory cortex. *Journal of Cognitive Neuroscience*, 23(10):3084–3094.

Kuwada, S., Batra, R., and Maher, V. L. (1986). Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones. *Hearing Research*, 21(2):179–192.

Lalor, E. C. and Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European Journal of Neuroscience*, 31(1):189–193.

Lalor, E. C., Power, A., Reilly, R. B., and Foxe, J. J. (2009). Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. *Journal of Neural Physiology*, 102(1):349–359.

Lane, R. D., Reiman, E. M., Ahern, G. L., Schwartz, G. E., and Davidson, R. J. (1997). Neuroanatomical correlates of happiness, sadness, and disgust. *American Journal of Psychiatry*, 154(7):926–933.

Large, E. W. (2001). Periodicity, pattern formation, and metric structure. *Journal of New Music Research*, 30(2):173–185.

Lartillot, O., Eerola, T., Toiviainen, P., and Fornari, J. (2008a). Multi-Feature Modeling of Pulse Clarity: Design, Validation and Optimization. In *ISMIR 2008*, pages 521–526.

Lartillot, O., Toiviainen, P., and Eerola, T. (2008b). A Matlab Toolbox for Music Information Retrieval. In Preisach, C., Burkhardt, H., Schmidt-Thieme, L., and Decker, R., editors, *Studies in Classification, Data Analysis, and Knowledge Organization*, pages 261–268. Springer Berlin Heidelberg.

Leaver, A. M., van Lare, J., Zielinski, B., Halpern, A. R., and Rauschecker, J. P. (2009). Brain Activation during Anticipation of Sound Sequences. *Journal of Neuroscience*, 29(8):2477–2485.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411.

Lehne, M. and Koelsch, S. (2015). Towards a general psychological model of tension and suspense. *Frontiers in Psychology*, 6(79).

Lehne, M., Rohrmeier, M., Gollmann, D., and Koelsch, S. (2013a). The influence of different structural features on felt musical tension in two piano pieces by Mozart and Mendelssohn. *Music Perception: An Interdisciplinary Journal*, 31(2):171–185.

Lehne, M., Rohrmeier, M., and Koelsch, S. (2013b). Tension-related activity in the orbitofrontal cortex and amygdala: an fMRI study with music. *Social cognitive and affective neuroscience*, pages 1–9.

Leonard, M. K. and Chang, E. F. (2014). Dynamic speech representations in the human temporal lobe. *Trends in cognitive sciences*, 18(9):472–479.

Lerdahl, F. and Jackendoff, R. (1983). An Overview of Hierarchical Structure in Music. *Music Perception: An Interdisciplinary Journal*, 1(2):229–252.

Lerdahl, F. and Krumhansl, C. (2007). Modeling tonal tension. *Music Perception*, 24(4):329–366.

Leuthardt, E. C., Gaona, C., Sharma, M., Szrama, N., Roland, J., Freudenberg, Z., and Schalk, G. (2011). Using the electrocorticographic speech network to control a brain–computer interface in humans. *Journal of Neural Engineering*, 8(3):036004.

Leuthardt, E. C., Miller, K. J., Schalk, G., Rao, R. P., and Ojemann, J. G. (2006). Electrocorticography-based brain computer interface-the Seattle experience. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 14(2):194–198.

Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., and Moran D. W. (2004). A brain–computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering*, 1(2):63.

Liebenthal, E., Ellingson, M. L., Spanaki, M. V., Prieto, T. E., Ropella, K. M., and Binder, J. R. (2003). Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *NeuroImage*, 19(4):1395–1404.

Lin, Y., Duann, J., Feng, W., Chen, J., and Jung, T. (2014). Revealing spatio-spectral electroencephalographic dynamics of musical mode and tempo perception by independent component analysis. *Journal of neuroengineering and rehabilitation*, 11(1):18.

London, J. (2012). *Hearing in Time*. Oxford University Press.

Lopez-Gordo, M. A., Fernandez, E., Romero, S., Pelayo, F., and Prieto, A. (2012). An auditory brain–computer interface evoked by natural speech. *Journal of Neural Engineering*, 9(3):036013.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, 103(49):18866–18869.

Loui, P., Grent, T., Torpey, D., and Woldorff, M. (2005). Effects of attention on the neural processing of harmonic syntax in Western music. *Cognitive Brain Research*, 25(3):678–687.

Low, Y., Corona-Strauss, F. I., Adam, P., and Strauss, D. (2007). Extraction of auditory attention correlates in single sweeps of cortical potentials by maximum entropy paradigms and its application. In *Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on : Neural Engineering IEEE/EMBS*, pages 469–472.

Madison, G., Gouyon, F., Ullén, F., and Hörnström, K. (2011). Modeling the tendency for music to induce movement in humans: First correlations with low-level audio descriptors across music genres. *Journal of Experimental Psychology: Human Perception and Performance*, 37(5):1578.

Madsen, C. K. and Fredrickson, W. E. (1993). The experience of musical tension: A replication of Nielsen's research using the continuous response digital interface. *Journal of Music Therapy*, 30(1):46–63.

Maiste, A. and Picton, T. (1989). Human auditory evoked potentials to frequency-modulated tones. *Ear and Hearing*, 10(3):153–160.

Majima, K., Matsuo, T., Kawasaki, K., Kawai, K., Saito, N., Hasegawa, I., and Kamitani, Y. (2014). Decoding visual object categories from temporal correlations of ECoG signals. *NeuroImage*, 90:74–83.

Marie, C. and Trainor, L. J. (2012). Development of simultaneous pitch encoding: infants show a high voice superiority effect. *Cerebral Cortex*, (bhs050):1–10.

Marie, C. and Trainor, L. J. (2014). Early development of polyphonic sound encoding and the high voice superiority effect. *Neuropsychologia*, 57:50–58.

Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J., and Benali, H. (2006). Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage*, 32(1):228–237.

Martin, B. A., Tremblay, K. L., and Korczak, P. (2008). Speech evoked potentials: from the laboratory to the clinic. *Ear and Hearing*, 29(3):285–313.

Martin, B. A., Tremblay, K. L., and Stapells, G. (2007). Principles and applications of cortical auditory evoked potentials. *Auditory evoked potentials: basic principles and clinical application*, pages 482–507.

Martin, S., Brunner, P., Holdgraf, C., Heinze, H. J., Crone, N. E., Rieger, J., Schalk, G., Knight, R. T., and Pasley, B. N. (2014). Decoding spectrotemporal features of overt and covert speech from the human cortex. *Frontiers in Neuroengineering*, 7:14.

Mauch, M. and Dixon, S. (2010). Approximate Note Transcription for the Improved Identification of Difficult Chords. In *ISMIR*, pages 135–140.

McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. d., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58(3):177–192.

McMenamin, B. W., Shackman, A. J., Greischar, L. L., and Davidson, R. J. (2011). Electromyogenic artifacts and electroencephalographic inferences revisited. *NeuroImage*, 54(1):4–9.

McMenamin, B. W., Shackman, A. J., Maxwell, J. S., Bachhuber, D. R. W., Koppenhaver, A. M., Greischar, L. L., and Davidson, R. J. (2010). Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. *NeuroImage*, 49(3):2416–2432.

Mechler, F., Victor, J. D., Purpura, K. P., and Shapley R. (1998). Robust temporal coding of contrast by V1 neurons for transient but not for steady-state stimuli. *The Journal of Neuroscience*, 18(16):6583–6598.

Menon, V. and Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *NeuroImage*, 28(1):175–184.

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., and Friederici, A. D. (2012). Perception of Words and Pitch Patterns in Song and Speech. *Frontiers in Psychology*, 3:76.

Mesgarani, N. and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397):233–236.

Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Journal of Neurophysiology*, 102(6):3329–3339.

Meyer, L. (1961). *Emotion and meaning in music*. University of Chicago Press.

Meyer, M., Baumann, S., and Jäncke, L. (2006). Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. *NeuroImage*, 32(4):1510–1523.

Mikutta, C. A., Altorfer, A., Strik, W., and Koenig, T. (2012). Emotions, arousal, and frontal alpha rhythm asymmetry during Beethoven's 5th symphony. *Brain topography*, 25(4):423–430.

Mikutta, C. A., Schwab, S., Niederhauser, S., Wuermle, O., Strik, W., and Altorfer, A. (2013). Music, perceived arousal, and intensity: Psychophysiological reactions to Chopin's 'Tristesse'. *Psychophysiology*, 50(9):909–919.

Miller, K. J., Schalk, G., Fetz, E. E., den Nijs, M., Ojemann, J. G., and Rao, R. P. (2010). Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proceedings of the National Academy of Sciences*, 107(9):4430–4435.

Mitrović, D., Zeppelzauer, M., and Breiteneder, C. (2010). Features for content-based audio retrieval. *Advances in Computers*, 78:71–150.

Mitterschiffthaler, M., Fu, C., Dalton, J., Andrew, C., and Williams, S. (2007). A functional MRI study of happy and sad affective states induced by classical music. *Human Brain Mapping*, 28(11):1150–1162.

Moore, B. C. J. and Gockel, H. E. (2002). Factors influencing sequential stream segregation. *Acta Acustica*, 88(3):320–333.

Mosher, J., Lewis, P., and Leahy, R. (1992). Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Transactions on Biomedical Engineering*, 39(6):541–557.

Mosher, J. C. and Leahy, R. M. (1998). Recursive MUSIC: a framework for EEG and MEG source localization. *Biomedical Engineering, IEEE Transactions on*, 45(11):1342–1354.

Mulert, C., Jäger, L., Propp, S., Karch, S., Störmann, S., Pogarell, O., Möller, H. J., Juckel, G., and Hegerl, U. (2005). Sound level dependence of the primary auditory cortex: Simultaneous measurement with 61-channel EEG and fMRI. *NeuroImage*, 28(1):49–58.

Mullen, T., Acar, Z. A., Worrell G., and Makeig S. (2011). Modeling cortical source dynamics and interactions during seizure. *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*.

Müller, K. R., Anderson, C. W., and Birch, G. E. (2003). Linear and nonlinear methods for brain-computer interfaces. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):165–169.

Müller, K. R., Mika, S., Rätsch, G., Tsuda, K., and Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. *Neural Networks, IEEE Transactions on*, 12(2):181–201.

Näätänen, R. and Picton, T. W. (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.

Nagel, F., Kopiez, R., Grewe, O., and Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods*, 39(2):283–290.

Nan, Y. and Friederici, A. D. (2013). Differential roles of right temporal cortex and Broca's area in pitch processing: evidence from music and Mandarin. *Human Brain Mapping*, 34(9):2045–2054.

Nan, Y., Knösche, T. R., and Friederici, A. D. (2006). The perception of musical phrase structure: a cross-cultural ERP study. *Brain Research*, 1094(1):179–191.

Nielsen, F. V. (1983). *Oplevelse af musikalsk spænding (The experience of musical tension)*. Bilag. Akademisk Forlag.

Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience*, 31(28):10234–10240.

Nozaradan, S., Peretz, I., and Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(49):17572–17581.

Nozaradan, S., Zerouali, Y., Peretz, I., and Mouraux A. (2013). Capturing with EEG the neural entrainment and coupling underlying sensorimotor synchronization to the beat. *Cerebral Cortex*, 25(3):736–747.

Ostroff, J. M., Martin, B. A., and Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear and Hearing*, 19(4):290–297.

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., and Lalor, E. C. (2014). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, pages 1–10.

Palmer, C. and Holleran S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Attention, Perception, & Psychophysics*, 56(3):301–312.

Pampalk, E., Rauber, A., and Merkl, D. (2002). Content-based organization and visualization of music archives. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 570–579.

Panksepp, J. (1995). The emotional sources of chills induced by music. *Music Perception*, pages 171–207.

Parra, L., Spence, C., Gerson, A., and Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage*, 28(2):326–341.

Pascual-Marqui, R. D. (1999). Review of methods for solving the EEG inverse problem. *International journal of bioelectromagnetism*, 1(1):75–86.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., and Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1):e1001251.

Patel, A. D. (2007). *Music, language, and the brain*. Oxford University Press.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4):767–776.

Perani, D., Saccuman, M., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., and Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences*, 107(10):4758.

Pereira, C., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., and Brattico (2011). Music and emotions in the brain: familiarity matters. *PloS One*, 6(11):e27241.

Pereira, D. R., Cardoso, S., Ferreira-Santos, F., Fernandes, C., Cunha-Reis, C., Almeida, P. R., Silveira, C., Barbosa, F., and Marques-Teixeira, J. (2014). Effects of inter-stimulus interval (ISI) duration on the N1 and P2 components of the auditory event-related potential. *International Journal of Psychophysiology*, 94(3):311–318.

Peretz, I. and Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review Psychology*, 56:89–114.

Peter, V. M. G. T. W. F. (2012). Discrimination of stress in speech and music: a mismatch negativity (MMN) study. *Psychophysiology*, 49(12):1590–1600.

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

Pisarenko, V. F. (1973). The Retrieval of Harmonics from a Covariance Function. *Geophysical Journal International*, 33(3):347–366.

Plack, C. J., Barker, D., and Hall, D. A. (2014). Pitch coding and pitch processing in the human brain. *Hearing Research*, 307:53–64.

Plotnikov, A., Stakheika, N., Gloria, A., Schatten, C., Bellotti F., Berta, R., and Ansovini, F. d. (2012). Exploiting real-time EEG analysis for assessing flow in games. In *Advanced Learning Technologies (ICALT), IEEE 12th International Conference on*, pages 688–689. IEEE.

Potes, C., Brunner, P., Gunduz, A., Knight, R. T., and Schalk, G. (2014). Spatial and temporal relationships of electrocorticographic alpha and gamma activity during auditory processing. *NeuroImage*, 97:188–195.

Potes, C., Gunduz, A., Brunner, P., and Schalk, G. (2012). Dynamics of electrocorticographic (ECoG) activity in human temporal and frontal cortical areas during music listening. *NeuroImage*, 61(4):841–848.

Power, A. J., Foxe, J. J., Forde, E. J., Reilly, R. B., and Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. *European Journal of Neuroscience*, 35(9):1497–1503.

Power, A. J., Lalor, E. C., and Reilly, R. B. (2011). Endogenous Auditory Spatial Attention Modulates Obligatory Sensory Activity in Auditory Cortex. *Cerebral Cortex*, 21(6):1223–1230.

Pratt, H., Starr, A., Michalewski, H. J., Dimitrijevic, A., Bleich, N., and Mittelman, N. (2009). Cortical evoked potentials to an auditory illusion: Binaural beats. *Clinical Neurophysiology*, 120(8):1514–1524.

Pressnitzer, D., McAdams, S., Winsberg, S., and Fineberg, J. (2000). Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Attention, Perception, & Psychophysics*, 62(1):66–80.

Purcell, D. W., John, S. M., Schneider, B. A., and Picton, T. W. (2004). Human temporal auditory acuity as assessed by envelope following responses. *The Journal of the Acoustical Society of America*, 116(6):3581–3593.

Pyper, B. J. and Peterman, R. M. (1998). Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Canadian Journal of Fisheries and Aquatic Sciences*, 55(9):2127–2140.

Regnault, P., Bigand, E., and Besson, M. (2001). Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: Evidence from auditory event-related brain potentials. *Journal of Cognitive Neuroscience*, 13(2):241–255.

Rieke, F., Bodnar, D. A., and Bialek, W. (1995). Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proceedings of the Royal Society of London B: Biological Sciences*, 262(1365):259–265.

Roebuck, K. (2012). *Tangible User Interfaces: High-impact Emerging Technology-What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors.* Emereo Publishing.

Rohrmeier, M. A. and Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology*, 83(2):164–175.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 336(1278):367–373.

Ruiz, M., Koelsch, S., and Bhattacharya, J. (2009). Decrease in early right alpha band phase synchronization and late gamma band oscillations in processing syntax in music. *Human Brain Mapping*, 30(4):1207–1225.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

Sabisch, B., Hahne, A., Glass, E., Suchodoletz, W., and Friederici A. von (2006). Auditory language comprehension in children with developmental dyslexia: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 18(10):1676–1695.

Salimpoor, V., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, (14(2)):257–262.

Sammler, D., Baird, A., Valabrègue, R., Clément, S., Dupont, S., Belin, P., and Samson, S. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *The Journal of Neuroscience*, 30(10):3572–3578.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H. J., Knösche, T. R., Wellmer, J., Widman, G., Elger, C. E., Friederici, A. D., and Schulze-Bonhage, A. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *NeuroImage*, 64:134–146.

Sammler, D., Koelsch, S., and Friederici, A. D. (2011). Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex*, 47(6):659–673.

Santos, A., Joly-Pottuz, B., Moreno, S., Habib, M., and Besson, M. (2007). Behavioural and event-related potentials evidence for pitch discrimination deficits in dyslexic children: Improvement after intensive phonic intervention. *Neuropsychologia*, 45(5):10.

Schaefer, R., Vlek, R., and Desain, P. (2011a). Decomposing rhythm processing: electroencephalography of perceived and self-imposed rhythmic patterns. *Psychological Research*, 75(2):95–106.

Schaefer, R. S., Desain, P., and Suppes, P. (2009). Structural decomposition of EEG signatures of melodic processing. *Biological Psychology*, 82(3):253–259.

Schaefer, R. S., Farquhar, J., Blokland, Y., Sadakata, M., and Desain, P. (2011b). Name that tune: decoding music from the listening brain. *NeuroImage*, 56(2):843–849.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1):32.

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). BCI2000: a general-purpose brain-computer interface (BCI) system. *Biomedical Engineering, IEEE Transactions on*, 51(6):1034–1043.

Schalk, G. and Mellinger, J. (2010). *Human-Computer Interaction: Practical Guide to Brain-Computer Interfacing with BCI2000: General-Purpose Software for Brain-Computer Interface Research, Data Acquisition, Stimulus Presentation, and Brain Monitoring*. Springer.

Schalk, G., Miller, K. J., Anderson, N. R., Wilson, J. A., Smyth, M. D., Ojemann J. G., and Leuthardt, E. C. (2008). Two-dimensional movement control using electrocorticographic signals in humans. *Journal of Neural Engineering*, 5(1):75.

Scherer, K. R. (2000). Psychological models of emotion. *The Neuropsychology of Emotion*, 137(3):137–162.

Schmidt, L. A. and Trainor, L. J. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition & Emotion*, 15(4):487–500.

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280.

Schölkopf, B., Smola, A., and Müller, K. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319.

Schreuder, M., Blankertz, B., and Tangermann, M. (2010). A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. *PloS One*, 5(4):e9813.

Schreuder, M., Rost, T., and Tangermann, M. (2011). Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. *Frontiers in Neuroscience*, 5(112).

Schubert, E. (2010). Continuous self-report methods. In Juslin, P. N. and Sloboda, J. A., editors, *Handbook of music and emotion: Theory, research, applications*, pages 223–253. Oxford University Press.

Schubert, E., Wolfe, J., and Tarnopolsky, A. (2004). Spectral centroid and timbre in complex, multiple instrumental textures. In *Proceedings of the international conference on music perception and cognition, North Western University, Illinois 2004*, pages 112–116.

Shahbazi, A. F., Ewald, A., and Nolte, G. (2012). Localizing True Brain Interactions from EEG and MEG Data with Subspace Methods and Modified Beamformers. *Computational and Mathematical Methods in Medicine*, 2012(1):1–11.

Shahin, A. J., Bosnyak, D., Trainor, L., and Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *The Journal of Neuroscience*, 23(13):5545–5552.

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., and Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *NeuroImage*, 41(1):113–122.

Shahin, A. J., Roberts, L. E., Pantev, C., Trainor, L. J., and Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport*, 16(16):1781–1785.

Shahin, A. J., Trainor, L., Roberts, L., Backer, K., and Miller L.M (2010). Development of auditory phase-locked activity for music sounds. *Journal of Neurophysiology*, 103(1):218.

Skouras, S., Gray, M., Critchley, H., and . Koelsch, S. (2013). fMRI scanner noise interaction with affective neural processes. *PloS One*, 8(11):e80564.

Sloboda, J. A., O'Neill, S. A., and Ivaldi, A. (2001). Functions of music in everyday life: An exploratory study using the experience sampling method. *Musicae Scientiae*, 5(1):9–32.

Snyder, J. S. and Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin*, 133(5):780.

Snyder, J. S., Alain, C., and Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, 18(1):1–13.

Snyder, J. S. and Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research*, 24(1):117–126.

Spüler, M., Walter, A., Ramos-Murguialday, A., Naros, G., Birbaumer, N. a. A. R. W., and M., B. (2014). Decoding of motor intentions from epidural ECoG recordings in severely paralyzed chronic stroke patients. *Journal of Neural Engineering*, 11(6):1–9.

Stanley, G. B., Li, F. F., and Dan, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *The Journal of Neuroscience*, 19(18):8036–8042.

Sturm, I., Blankertz, B., Potes, C., Schalk, G., and Curio, G. (2014). ECoG high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song. *Frontiers in Human Neuroscience*, 8(798).

Sundberg, J., Nord, L., and Carlson, R. (1991). Music, language, speech and brain.

Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *The Journal of the Acoustical Society of America*, 117(3):1285–1298.

Sussman, E. S., Horváth, J., Winkler, I., and Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics*, 69(1):136–152.

Thaut, M. (2005). The future of music in therapy and medicine. *Annals of the New York Academy of Sciences*, 1060(1):303–308.

Theiler, J., Eubank, S., Longtin, A., G., B., D., and Farmer, J. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1):77–94.

Thompson, J. (2013). *Neural decoding of subjective music listening experiences - unpublished Master's Thesis*. PhD thesis, Dartmouth College Hanover, New Hampshire, Hanover.

Tikhonov, A. N. and Arsenin, V. I. (1977). *Solutions of ill-posed problems*. Scripta series in mathematics. Winston and Halsted Press, Washington and New York.

Timm, L., Vuust, P., Brattico, E., Agrawal, D., Debener, S., Büchner, A., Dengler, R., and Wittfoth, M. (2014). Residual neural processing of musical sound features in adult cochlear implant users. *Frontiers in Human Neuroscience*, 8(181).

Toiviainen, P., Alluri, V., Brattico, E., Wallentin, M., and Vuust, P. (2014). Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data. *NeuroImage*, 88:170–180.

Tomioka, R. and Müller, K. (2010). A regularized discriminative framework for EEG analysis with application to brain–computer interface. *NeuroImage*, 49(1):415–432.

Trainor, L., McDonald, K., and Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience*, 14(3):430–442.

Trainor, L. J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1664):20140089.

Treder, M. S. and Blankertz, B. (2010). Research (C) overt attention and visual speller design in an ERP-based brain-computer interface. *Behav. Brain Funct*, 6:1–13.

Treder, M. S., Purwins, H., Miklody, D., Sturm, I., and Blankertz, B. (2014). Decoding auditory attention to instruments in polyphonic music using single-trial EEG classification. *Journal of Neural Engineering*, 11(2):026009.

Tremblay, K. L. and Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *Journal of Speech, Language, and Hearing Research*, 45(3):564–572.

Tremblay, K. L., Kraus, N., McGee, T., Ponton, C., and Otis, B. (2001). Central Auditory Plasticity: Changes in the N1-P2 Complex after Speech-Sound Training. *Ear & Hearing*, 22(2):79–90.

Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., and Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Frontiers in systems neuroscience*, 8(28).

Trochidis, K. and Bigand, E. (2012). EEG-based emotion perception during music listening. *Proceedings of the ICMPC 2012*, pages 1018–1021.

Turner, R. I. A. (2009). Brain, music and musicality: inferences from neuroimaging. In Malloch, S. and Trevarthen C., editors, *Communicative musicality: Exploring the basis of human companionship*, pages 148–182. Oxford University Press, USA.

Tuuri, K. and Eerola, T. (2012). Formulating a revised taxonomy for modes of listening. *Journal of New Music Research*, 41(2):137–152.

Upham, F., Cambouropoulos, E., Tsougras, C., Mavromatis, P., and Pastiades, K. (2012). Limits on the application of statistical correlations to continuous response data. In E. Cambouropoulos, e. a., editor, *Proceedings of the 12th International Conference on Music Perception and Cognition*, pages 1037–1041.

van Noorden, L. H. (1975). *Temporal coherence in the perception of tone sequences*. PhD thesis, Netherlands.

van Vansteensel, M. J., Hermes, D., Aarnoutse, E. J., Bleichner, M. G., Schalk G., Rijen, P. C., and Ramsey, N. F. (2010). Brain–computer interfacing based on cognitive control. *Annals of neurology*, 67(6):809–816.

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Statistics for Engineering and Information Science. Springer New York and Imprint: Springer, New York and NY, second edition. edition.

Varèse, E. and Chou, W. C. (1966). The Liberation of Sound. *Perspectives of New Music*, 5(1):11–19.

Whiting, K. A., Martin, B. A., and Stapells, D. R. (1998). The effects of broadband noise masking on cortical event-related potentials to speech sounds/ba/and/da. *Ear and Hearing*, 19(3):218–231.

Whittingstall, K., Bartels, A., Singh, V., Kwon, S., and Logothetis, N. K. (2010). Integration of EEG source imaging and fMRI during continuous viewing of natural movies. *Magnetic resonance imaging*, 28(8):1135–1142.

Wilkins, R. W., Hodges, D. A., Laurienti, P. J., Steen, M. R., and Burdette, J. H. (2012). Network science: a new method for investigating the complexity of musical experiences in the brain. *Leonardo*, 45(3):282–283.

Wilson, J. A., Felton, E. A., Garell, P. C., Schalk, G., and Williams, J. C. (2006). ECoG factors underlying multimodal control of a brain-computer interface. In *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 14.2*, pages 246–250.

Winkler, I., Denham, S. L., and Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in cognitive sciences*, 13(12):532–540.

Winkler, I., Takegata, R., and Sussman, E. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Cognitive Brain Research*, 25(1):291–299.

Woods, D. L. and Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 65(6):447–459.

Wu, J., Z., J., Ding, X., Li, R., and Zhou, C. (2013). The effects of music on brain functional networks: a network analysis. *Neuroscience*, 250:49–59.

Wundt, W. M. (1913). *Grundriss der Psychologie*. A. Kröner.

Yabe, H., Winkler, I., Czigler, I., Koyama, S., Kakigi, R., Sutoh, T., Hiruma, T., and Kaneko, S. (2001). Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. *Brain Research*, 897(1):222–227.

Yao, H., Shi L., Han F., Gao H., and Dan Y. (2007). Rapid learning in cortical coding of visual scenes. *Nature Neuroscience*, 10(6):772–778.

Yoncheva, Y., Maurer U., Zevin J. D., and McCandliss, B. D. (2014). Selective attention to phonology dynamically modulates initial encoding of auditory words within the left hemisphere. *NeuroImage*, 97:262–270.

Zatorre, R. and Krumhansl, C. (2002). Mental models and musical minds. *Science*, 298(5601):2138.

Zeng, F. G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (1999). Temporal and speech processing deficits in auditory neuropathy. *Neuroreport*, 10(16):3429–3435.

Zentner, M. (2010). Homer's Prophecy: an Essay on Music's Primary Emotions. *Music Analysis*, 29(1-3):102–125.

Ziehe, A., Kawanabe, M., Harmeling, S., and Müller, K. (2003). Blind separation of post-nonlinear mixtures using linearizing transformations and temporal decorrelation. *The Journal of Machine Learning Research*, 4:1319–1338.

Ziehe, A., Laskov, P., Nolte, G., and Müller, K. R. (2004). A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *The Journal of Machine Learning Research*, 5:777–800.

Ziehe, A. and Müller, K. (1998). TDSEP—an efficient algorithm for blind separation using time structure. In *ICANN 98*, pages 675–680. Springer.

Ziehe, A., Müller, K., Nolte, G., Mackert, B., and Curio, G. (2000). Artifact reduction in magnetoneurography based on time-delayed second-order correlations. *Biomedical Engineering, IEEE Transactions on*, 47(1):75–87.

Zwicker, E. and Scharf, B. (1965). A model of loudness summation. *Psychological Review*, 72(1):3.