TOWARD AN OBJECTIVE NEUROPHYSIOLOGICAL MEASURE OF MUSICAL ENGAGEMENT

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MUSIC
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Blair Bohannan Kaneshiro
July 2016

© 2016 by Blair Bohannan Kaneshiro. All Rights Reserved.
Re-distributed by Stanford University under license with the author.

This work is licensed under a Creative Commons Attribution-3.0 United States License.
http://creativecommons.org/licenses/by/3.0/us/

This dissertation is online at: http://purl.stanford.edu/xk371tf6758

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Jonathan Berger, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Anthony Norcia, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Julius Smith, III

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

Engaging listeners is an inherent goal of music. The concept of 'musical engagement', however, carries multiple connotations and remains difficult to quantify or even define. In particular, an objective measure of musical engagement is lacking.

Over past decades, cortical responses have been used to investigate processing of music. While these responses are objective and can be recorded in real time, they suffer from a low signal-to-noise ratio and reflect, at best, an abstraction of the corresponding stimuli. As a result, approaches to this research have historically focused primarily on controlled stimuli with limited ecological validity, and event-related averaging of responses, which requires short stimulus epochs and numerous stimulus presentations. Responses to real-world stimuli have proven challenging to analyze and interpret. How can we move beyond these limitations to derive a measure of engagement with 'real' music (i.e., naturalistic and complete musical works) from the brain response?

In this thesis, we address these limitations by introducing a novel analysis framework for interpreting listeners' responses to music, with the ultimate goal of developing a meaningful, quantitative, and dynamically changing index of musical engagement. We draw from recent approaches in neuroscience and physiology that use synchrony of audience responses to study engagement in other domains. Specifically, we examine time-resolved inter-subject correlations (ISCs) of cortical, physiological, and behavioral responses to musical pieces heard in their entirety. The current approach is facilitated by a recently developed method that efficiently extracts relevant, stimulus-related activity from a complex, noisy response. This method allows for full-length, ecologically valid stimuli to be presented in a single-listen experimental paradigm.

The proposed methodologies are tested and evaluated in two experiments. First, we validate the approach by deriving cortical components from scalp-recorded electroencephalographic (EEG) responses to intact and scrambled songs and computing their ISCs. In a

second experiment, we broaden the context of the approach by comparing EEG-ISCs to the activity and synchrony of physiological and continuous behavioral responses.

This work makes several novel contributions to the field of music cognition. First, we show that the presence of temporally relevant musical features produces a consistent component topography in the brain response. Furthermore, the ISCs computed from this component are higher when such musical features are retained. We additionally employ a novel approach to experimental design, choosing highly engaging stimuli that were unknown to our participants, and introducing computational procedures for manipulating the stimuli. Finally, we demonstrate that brain responses to full-length musical works from various genres and styles can be successfully analyzed in a single-listen paradigm.

Acknowledgments

There are many people who have made the completion of this thesis, and my PhD, possible. First and foremost, my adviser Jonathan Berger has provided invaluable support and guidance over the years. Thank you for believing in me, offering advice when I needed it, and giving me the flexibility to explore new approaches to music research in a variety of domains. To my co-adviser Anthony Norcia, thank you for your tremendous mentorship and generosity. I have learned so much from our ongoing discussions about representation, methodology, vision, and music, and look forward to continuing the conversation.

I thank Julius Smith for being an ever-positive presence throughout my graduate career. Despite a lack of background when I started, you encouraged me to pursue an engineering degree, the results of which continue to show through my work. Thank you to Ge Wang for setting an example of fearlessness in pursuing new avenues of research, and for your sincere feedback on my work over the years. I would also like to thank Trevor Hastie for serving as the Chair for my defense. Your suggestions are already leading to new research ideas! Finally, I must thank the late Patrick Suppes, my first academic mentor and the person who set me on the path to graduate school. Pat saw something in me worth developing, and I would not be where I am now without his support and intellectual influence.

I send a warm and heartfelt thank you to Duc Nguyen. Duc, our work together over the past year and a half has made this intense period in my life not only bearable but enjoyable. It's been a privilege to be your colleague and your friend. I am also indebted to Daniel Abrams, Jacek Dmochowski, and Marcos Perreau Guimaraes for their technical mentorship, career advice, and friendship over the years. Each of you has shared your expertise and your time with me, and I am truly appreciative. Several other mentors have helped me along the way as well, including Jonathan Abel, Fred Gibbons, Malcolm Slaney, Jason Titus, and Avery Wang. And a special thank you to Nola Nahulu, Marcia Stratman, and especially the late Janet Stotts for shaping my musical identity early in life.

To Steinunn Arnardóttir, Jorge Herrera, Hyung-Suk Kim, and Jieun Oh: Thank you for being awesome friends, classmates, and collaborators. I feel very lucky to have been a student at the same time as you! I'd also like to express my appreciation to my colleagues, past and present, in what is now the Music Engagement Research Initiative: Tysen Dauer, Nick Gang, Evan Gitterman, Kristin Kueter, Sophia Laurenzi, Steven Losorelli, Megha Makam, Jeff Rector, Anna Cecilia Rosenkranz, Karanvir Singh, and Jeff Smith. Finally, thank you to Tom Collins, Rebecca Schaefer, and Sebastian Stober, who inspire me to keep learning; as well as my former labmates from the Suppes Brain Lab and researchers from the Stanford Vision and Neuro-Development Lab.

Thank you to Jay Kadis and Vladimir Vildavski for helping me with my hardware and software configurations; Fernando Lopez-Lezcano, Colin Sullivan, and Carr Wilkerson for computing help; and John Granzow, Romain Michon, and Kurt Werner for sawing and splicing various items on my behalf. Many thanks also to Debbie Barney, Charlotte Cattivera, Amita Kumar, Michelle Lodwick, and Nette Worthey for considerable administrative support over the years.

Thank you to everyone at CCRMA, the Stanford Department of Music, Shazam, and the ICMPC and ISMIR communities. I've been extremely fortunate to work among you in the pursuit of understanding the human experience of music.

Finally, thank you to my family. My parents and brothers have shown me love and support for my entire life, and in recent years have been extremely patient and understanding regarding my distance and lack of communication. And most of all, I would like to thank my wonderful husband Lewis. You have provided unwavering support through all of the successes and obstacles I have encountered during this experience, and I could not have done it without you. I'm very excited for us as we embark on our next adventures!

Contents

Abstract **iv**

Acknowledgments **vi**

1 Introduction **1**

1.1 Musical Engagement 1
1.2 Overview 2
1.3 Main Contributions 3
1.4 Outline 4
1.5 Co-Authored Publications 5

2 Background **6**

2.1 Investigating Music Processing Using EEG 6
2.1.1 Averaging-Based Approaches 6
2.1.2 Multivariate Approaches 8
2.1.3 Criteria for Present Research 11
2.2 Reliable Components Analysis 11
2.2.1 RCA Derivation 12
2.2.2 Composition of Data Matrices 15
2.3 Inter-Subject Correlations 17
2.3.1 EEG-ISCs 17
2.3.2 fMRI-ISC Studies 17
2.4 Implications for Musical Engagement 18

3 Experiment 1 **20**

3.1 Introduction 20

3.2 Methods . . . 21
3.2.1 Ethics Statement . . . 21
3.2.2 Stimuli . . . 21
3.2.3 Participants . . . 25
3.2.4 Experimental Paradigm and Data Acquisition . . . 26
3.2.5 EEG Preprocessing and Analysis . . . 28
3.2.6 Extraction of Stimulus Features . . . 32
3.2.7 Statistical Analyses . . . 32
3.3 Results . . . 33
3.3.1 Behavioral Ratings . . . 33
3.3.2 EEG Results . . . 34
3.4 Discussion . . . 46

4 Physiological and Behavioral Measures **48**
4.1 Physiological Responses . . . 48
4.1.1 Experimental Approaches . . . 50
4.1.2 Analysis Approaches . . . 53
4.1.3 Summary of Current Findings . . . 54
4.1.4 Reliability of Physiological Responses . . . 56
4.2 Continuous Behavioral Responses . . . 56
4.2.1 Response Collection Interfaces . . . 57
4.2.2 Dimensions of Self-Report . . . 57
4.2.3 Reliability of Continuous Behavioral Responses . . . 59
4.2.4 Experimental and Analytical Approaches . . . 59
4.3 Discussion . . . 60
4.3.1 Considerations . . . 60

5 Experiment 2 **63**
5.1 Introduction . . . 63
5.2 Methods . . . 64
5.2.1 Ethics Statement . . . 64
5.2.2 Stimuli . . . 64
5.2.3 Participants . . . 65
5.2.4 Experimental Paradigm and Data Acquisition . . . 67

5.2.5 Data Analysis . . . 70
5.3 Results . . . 75
5.3.1 Behavioral Ratings . . . 75
5.3.2 EEG Responses . . . 76
5.3.3 Continuous Behavioral Responses . . . 77
5.3.4 ECG Responses . . . 78
5.3.5 Respiratory Responses . . . 78
5.4 Discussion . . . 81
5.4.1 Main Findings . . . 81
5.4.2 Considerations . . . 85

6 Conclusion **89**
6.1 A Narrative Framework for Musical Engagement . . . 89
6.2 Future Work . . . 90
6.3 Closing Remarks . . . 92

A Experiment 1 Supplement **93**
A.1 Stimulus Figures, Songs 2–4 . . . 93
A.2 Inter-Subject Correlations . . . 96
A.2.1 RC1 and RC2 ISCs, Songs 2–4 . . . 96
A.2.2 RC1 and RC2 ISCs for Manipulated Stimuli . . . 98
A.2.3 First- and Second-Listen RC1 ISCs . . . 101
A.2.4 ISC-Amplitude Envelope Plots . . . 104

List of Tables

3.1 Hindi stimulus information 22
3.2 Wilcoxon test results comparing listen-1 and listen-2 RC1 ISCs 40
3.3 ISC-amplitude envelope correlation information 44
3.4 Correlation of original and flipped reversed ISCs 45

List of Figures

3.1 Waveforms, spectrograms, and magnitude spectra of Song 1 stimuli 25
3.2 Behavioral ratings of Hindi stimuli 34
3.3 RC1–RC3 topographies 35
3.4 RC1 topographies by stimulus condition and listen 36
3.5 Time-resolved RC1 and RC2 ISCs for Song 1 37
3.6 Time-resolved RC1 and RC2 ISCs for all original songs 38
3.7 Proportion of significant RC1 and RC2 ISCs, first listen 39
3.8 Proportion of significant RC1 ISCs, first versus second listen 40
3.9 RC1 ISCs of original stimuli, first versus second listen 41
3.10 First-listen RC1 ISCs of original stimuli, plotted over song parts 42
3.11 RC1 ISCs of Song 4 responses, plotted with stimulus amplitude envelopes . 43
3.12 RC1 ISCs for original stimuli plotted with flipped ISCs for reversed stimuli 45

5.1 Elgar stimulus waveform and spectrogram 66
5.2 Physiological sensor configuration 69
5.3 Behavioral ratings of Elgar stimuli 75
5.4 RC1 topographies for responses to Elgar stimuli 76
5.5 Elgar EEG-ISCs 77
5.6 Continuous behavioral responses 79
5.7 HR activity and synchrony over time 80
5.8 Respiratory amplitude over time 82
5.9 Respiratory rate over time 83
5.10 Aggregate responses, original stimulus 84
5.11 Aggregate responses, reversed stimulus 85
5.12 Summary of proportion of significant results 86

A.1 Waveforms, spectrograms, and magnitude spectra of Song 2 stimuli 93
A.2 Waveforms, spectrograms, and magnitude spectra of Song 3 stimuli 94
A.3 Waveforms, spectrograms, and magnitude spectra of Song 4 stimuli 95
A.4 Time-resolved RC1 and RC2 ISCs for Song 2 96
A.5 Time-resolved RC1 and RC2 ISCs for Song 3 97
A.6 Time-resolved RC1 and RC2 ISCs for Song 4 97
A.7 Time-resolved RC1 and RC2 ISCs for all reversed songs 98
A.8 Time-resolved RC1 and RC2 ISCs for all measure-shuffled songs 99
A.9 Time-resolved RC1 and RC2 ISCs for all phase-scrambled songs 100
A.10 RC1 ISCs of reversed stimuli, first versus second listen 101
A.11 RC1 ISCs of measure-shuffled stimuli, first versus second listen 102
A.12 RC1 ISCs of phase-scrambled stimuli, first versus second listen 103
A.13 RC1 ISCs of Song 1 responses, plotted with stimulus amplitude envelopes . 104
A.14 RC1 ISCs of Song 2 responses, plotted with stimulus amplitude envelopes . 105
A.15 RC1 ISCs of Song 3 responses, plotted with stimulus amplitude envelopes . 106

Chapter 1

Introduction

1.1 Musical Engagement

The enjoyment of music is ubiquitous among humans. The experience of enjoyment varies in quality and intensity of *engagement*, from passive and barely aware, to deeply engrossed and attentive. Engagement with music can, and typically does, vary over the course of listening to a single musical work. Research in numerous domains, including the theory, psychology, perception, and cognition of music, is concerned with understanding how, when, and why a listener's engagement with music varies, even while the very notion of musical engagement eludes definition and quantification.

The challenge of assigning a precise definition to musical engagement stems, in part, from the many functions that music can serve. Deeply engaged listening that might occur in the context of dedicated listening sessions, whether in a concert hall or at home, implicates music as the target of focused attention. However, humans engage with music in numerous other ways. Listeners can *engage* music to serve as background for other activities such as socialization, driving, or work (Sloboda et al., 2001; Lonsdale and North, 2011; Schäfer et al., 2013). Engaging with music can also denote active undertakings such as learning an instrument, performing music, or attending an event; music can additionally provide a framework to guide movement and action, as in a dance or exercise setting (Chin and Rickard, 2012). The utility of music can extend even beyond the musical content itself. Live musical events, such as concerts and performance ensembles, facilitate social interactions by providing opportunities for collective participation and listening (Lonsdale and North, 2011; Chin and Rickard, 2012). Musical tastes can serve as a vehicle for communicating

one's identity, conveying beliefs, expressing emotion, and relating to others (Schäfer et al., 2013; Laplante and Downie, 2011; Rentfrow, 2012).

Evidence of engagement similarly takes many forms. Engagement can be manifested through a listener's brain response, physiological activity, or expressed musical preferences. A person's musical practices and activities may communicate engagement. We may assess musical engagement by examining how a person consumes, learns, discovers, and shares music, or how music is employed to serve his emotional, practical, and social needs. As a result, the term *engagement*, as it relates to music, carries a variety of connotations, and consequently lacks a universal measure.

For the scope of this thesis, we are interested specifically in quantifying the state of *focused* engagement with music—that is, a state of 'being compelled, drawn in, connected to what is happening, and interested in what will happen next' (Schubert et al., 2013). Additionally, we make the assumption that a composer intends to 'reliably create experiences in audience members' (Davies, 2014). Under this definition and assumption, therefore, content that is engaging will drive the subjective experiences of audience members in a similar fashion; put another way, an engaged audience will experience and process the content in a similar fashion. Consequently, if we may assume that mental states are reflected in brain states, and that brain states are measurable by means of brain activity (Hasson et al., 2008b), then it is reasonable to conclude that brain activity—specifically, the *synchrony* of brain activity across audience members, which can be quantified by computing inter-subject correlations (ISCs) of responses—may constitute a modality for indexing this type of focused engagement.

1.2 Overview

We are interested here in validating a metric for quantifying a state of engagement while listening to music. The focus will be primarily on brain responses, but will also consider continuous physiological and behavioral responses, as well as behavioral ratings of musical excerpts. By applying a modern analysis method to brain responses to music and interpreting our experimental results in relation to findings from cognitive neuroscience, music perception and cognition, and narrative transportation theory, we make significant contributions to the study of musical engagement and—more broadly—to the fields of music cognition and music neuroscience.

In this thesis we introduce a state-of-the-art methodology as a foundational tool for studying musical engagement. We apply this analysis approach, recently developed for electroencephalographic (EEG) responses and shown to index engagement with audiovisual film excerpts, for the first time to music.

We present two experiments advancing research in musical engagement using EEG. In Experiment 1, we validate a recently developed spatial filtering method for EEG for music research. This method, termed Reliable Components Analysis (RCA) (Dmochowski et al., 2012, 2015), is a spatial filtering technique that derives maximally correlated components across a collection of EEG records. ISCs computed in the subspace of single components serve as a measure of temporal reliability of the responses and have been shown to index audience engagement with audiovisual film excerpts (Dmochowski et al., 2012, 2014). In this first experiment we apply the technique for the first time to responses to music, and show that stimuli retaining musically relevant features such as beat, meter, and melody yield more plausible component topographies, as well as higher ISCs, than phase-scrambled controls. We also show a correspondence between the temporal reliability of the neural response and participant ratings of the stimuli.

In Experiment 2, we further validate the use of RCA and EEG-ISCs while taking a first step toward examining the relationship between engagement and arousal. We analyze EEG responses in conjunction with continuous physiological and behavioral measures in response to a musical excerpt characterized by large fluctuations of arousal. We assess the temporal reliability of these responses, as well as the physiological and behavioral activations themselves, to identify periods of statistically significant responses and their relation to a predefined set of musically salient events.

1.3 Main Contributions

This thesis presents several novel contributions to the fields of music perception, cognition, and neuroscience. The EEG analysis approach used here is still a novel neuroscientific method, and this is its first application to music. Next, as we will discuss in future chapters, a neurophysiological measure of engagement is considered an *objective* measure—one that does not divert audience attention or suffer from self-report bias. Finally, on a methodological level, this thesis constitutes a substantial contribution to the field of music neuroscience. A multivariate approach such as RCA utilizes the full brain response, not just preselected

electrodes and time points. This provides a data-driven approach toward identifying temporal and spatial components of interest in the brain response, and facilitates the use of real-world, naturalistic music excerpts—that is, musical excerpts that were created with the intention of being consumed in real life. Importantly, while most music EEG studies to date use averaging approaches, requiring hundreds or thousands of stimulus presentations, the present method allows for a single-listen experimental paradigm. As a result, we may use longer stimuli (on the order of minutes rather than seconds), as well as ecologically valid stimuli, which are more difficult to analyze in an averaging paradigm but may present a more 'real' musical experience than more traditional, controlled stimuli (Madsen and Geringer, 1990). A broader implication of the single-listen paradigm is that it facilitates the study of a concept such as engagement. Hearing a song that was created to engage the listener, hearing it in its entirety, and not needing to hear it several times in a row more closely resembles the experience of music listening in real life. This is a critical component in eliciting measurable states of engagement in our listeners.

1.4 Outline

The rest of the thesis is structured as follows.

In Chapter 2, we review background literature that led to the overall formulation of the thesis. We begin by reviewing approaches to investigating musical processing using EEG, with a focus on multivariate approaches and the incorporation of naturalistic stimuli. From there we introduce RCA, its derivation, and its application to the study of engagement using EEG responses to audiovisual film excerpts. We conclude by contextualizing this approach within a broader field of research using inter-subject correlations of cortical responses to derive data-driven insights into neural processing of naturalistic stimuli from various modalities.

Chapter 3 describes Experiment 1, which is our initial validation of the analysis approach. We describe the design of the experiment, our custom hardware configuration for stimulus delivery, and custom software implementations for preprocessing and analyzing the data. We present results and discuss their significance in relation to the seminal audiovisual study in which the analysis approach was introduced (Dmochowski et al., 2012).

Next, Chapter 4 presents a review of foundational literature on physiological and continuous behavioral responses to music. We trace the use of these responses back to their early

applications, review data-collection apparatuses and experimental paradigms, and summarize the general consensus (or lack thereof) of the findings to date. We draw connections between the approaches of these studies and approaches used in the EEG-ISC paradigm, which motivate Experiment 2.

Experiment 2 is presented in Chapter 5. The methodological focus here is on the acquisition and analysis of additional physiological and continuous behavioral responses. We interpret the various results within a framework guided by musical events that we consider to be especially salient due to thematic elements, building or resolution of musical tension, and extremes of musical texture and dynamics.

Chapter 6 concludes the thesis. We discuss the relation of the present work to the transportation/cognitive elaboration framework used to characterize states of narrative engagement. We highlight potential avenues for future research in musical engagement, both as a continuation of the approach used here, as well as in other experimental settings and with other forms of response data.

1.5 Co-Authored Publications

Much of the literature reviewed in §2.1 (in particular §2.1.2) was reviewed previously in the published paper by Kaneshiro and Dmochowski (2015). Experiment 1 (reported in Chapter 3) is a collaborative work with supporting authors Jacek P. Dmochowski, Duc T. Nguyen, Anthony M. Norcia, and Jonathan Berger (in preparation). The data from this experiment are published in Kaneshiro et al. (2016a). An earlier version of this experiment is published in Kaneshiro et al. (2014). Selected results from Experiment 2 (reported in Chapter 5) are published in Kaneshiro et al. (2016b).

Chapter 2

Background

2.1 Investigating Music Processing Using EEG

EEG is the measure of the electrical activity of the brain. When a sufficient number of neurons, ranging from hundreds to tens of thousands, fire synchronously, the resulting electrical field is strong enough in aggregate to be measured noninvasively using electrodes placed on the surface of the scalp (Cohen, 2014).

Relative to functional magnetic resonance imaging (fMRI), EEG provides superior temporal resolution—on the order of milliseconds—deeming it particularly useful for studying time-based processes such as music. However, scalp-recorded EEG offers relatively diminished spatial resolution due to signal propagation through the skull and scalp; the localization of underlying cortical sources from the scalp recording thus remains an open field of research. In addition, the signal-to-noise ratio (SNR) of EEG is low—estimated to be on the order of -20 dB (Kaneshiro and Dmochowski, 2015); thus, recovering relevant stimulus- or task-related components from the response can prove challenging, especially in the analysis of single trials. Even so, the low expense, noninvasiveness, and high temporal resolution of EEG have led to its wide adoption in neuroscience and cognitive psychology research.

2.1.1 Averaging-Based Approaches

A widely used approach to analyzing EEG-recorded responses to sensory stimuli involves averaging of time-locked Event-Related Potentials (ERPs). In this paradigm, responses to a given stimulus condition are aggregated and averaged across trials. This approach generally employs univariate analysis techniques, by which data from one or a few electrodes

are averaged, and amplitudes and latencies of preselected peaks—or components—of the averaged waveforms are compared across stimulus conditions. Response latencies of interest for event-related analyses typically range from less than 10 msec after stimulus onset for auditory brainstem responses to 50–500 msec for cortical responses. Averaging-based ERP analyses of cortical responses typically require on the order of tens or hundreds of stimulus presentations; for subcoritcal responses, such as those generated by the auditory brainstem, thousands of stimulus repetitions are required (Skoe and Kraus, 2010).

Music ERP studies often focus on various aspects of fulfillment or violation of musical expectations with regard to such attributes as pitch or tonal organization, beat, and timbre. Typically, short stimulus epochs (e.g., single chords) drawn from longer stimuli (e.g., chord progressions) are analyzed. Different components have been found to reflect processing of different musical attributes. For example, the P300 is a positive deflection that reflects processing of an 'oddball'—that is, improbable or unexpected—stimulus event. Occurring approximately 300 msec after the onset of the unexpected stimulus, this component has been found to generally require active attention to the stimulus, and its amplitude is proportional to the degree of unexpectedness of the oddball event (Picton, 1992). Janata (1995) used this component to study the tonal hierarchy of expectations in a chord-progression paradigm, finding that unexpected chord events in the place of an expected cadence affects both the amplitude and latency of P300 sub-components. The P300 has also been used to study rhythmic expectancy, for example to demonstrate different processing strategies employed by rhythmic experts versus nonmusicians (Jongsma et al., 2004).

The mismatch negativity (MMN), a negative deflection occurring between 90–150 msec after a stimulus, and the early right anterior negativity (ERAN), occurring slightly later with a latency of around 150–200 msec, are other examples of ERP components implicated in music processing. Both components occur in response to deviant musical events, but have been shown to reflect distinct dimensions of musical processing. The MMN is considered an automatic response to deviant events, whether physical or abstract in nature, while the ERAN is thought to reflect processing of musical syntax, drawing from longer-term models of musical expectancy (Zanto et al., 2006; Leino et al., 2007; Koelsch, 2009). Studies have shown that these two components can be differentiated on the basis of which stimulus dimension is disrupted. For example, the MMN has been linked to acoustical deviance (such as mistuning), while the ERAN is evoked by syntactic deviance (e.g., out-of-key chords, especially once a strong tonal expectation has been established) (Leino et al., 2007;

Koelsch et al., 2007). A comprehensive comparative analysis of these two components is given by Koelsch (2009).

2.1.2 Multivariate Approaches

While the univariate, averaging-based approaches described above are still widely used, especially in the field of music neuroscience, EEG researchers have in recent years begun to adopt multivariate approaches to data analysis. Multivariate analyses not only facilitate utilization of the full response—combining data across electrodes and time samples—but also facilitate data-driven approaches toward identifying temporal and spatial features of interest in the brain response, rather than selecting them in advance.

Single-Trial Classification

One multivariate approach to analyzing EEG data is single-trial classification. In this setting, a statistical model is built from a set of training trials and used to predict the label (descriptor of the stimulus) of unlabeled test trials. A useful introduction and tutorial specific to EEG is provided by Blankertz et al. (2011). Feature-selection procedures inherent to developing classification models, as well as analysis of classifier performance over subsets of the brain response, can serve to reveal spatiotemporal components of the response that successfully discriminate between stimuli or stimulus categories.

The first single-trial EEG classification study focusing on musical stimuli was presented by Schaefer et al. (2011). Here, the authors were able to classify EEG responses to seven short excerpts of naturalistic music from a variety of genres, significantly above chance. A subsequent study by Kaneshiro et al. (2012) appropriated the tonal-expectation paradigm often used in ERP studies—that is, using short, composed chord progressions with expected and unexpected cadential events—and, rather than averaging responses at single electrodes, classified multi-electrode responses to the cadential events. The classifier was able to discriminate tonal functions of cadential events significantly above chance, even when responses were grouped across musical keys. More recently, Stober et al. (2014) classified EEG responses from East African listeners who heard 12 Western and 12 East African rhythms. Here the authors used deep-learning techniques to predict both the rhythm family of the stimulus (2-class problem) as well as individual rhythms (24-class problem).

One application of single-trial EEG classification is the brain-computer interface (BCI) (Blankertz et al., 2002). A successful BCI enables a user who cannot communicate through

conventional means, such as speech or movement, to do so mentally by imagining a cue that would then be detected in the brain response and translated to an action or message. In a musical context, cueing would involve selective attention to—or interaction with—an ongoing musical stimulus. For example, metrical accents mentally imposed over an ongoing beat sequence have been successfully detected in the EEG response (Vlek et al., 2011a), and a subjective-accenting classification model has been successfully trained from responses to acoustical (presented) metrical accents (Vlek et al., 2011b). In a more recent BCI-motivated EEG classification study, Treder et al. (2014) played polyphonic stimuli with intermittent oddball events to listeners, who focused on the activity of just one instrument. The authors then leveraged the aforementioned P300 component, namely that this component is evoked by *attention* to oddball stimuli, and classified responses to just the oddball events from all instruments in order to identify which was being attended to.

The above studies not only contribute advances in analysis methodologies for music EEG research; they also constitute a move toward more naturalistic stimuli. While some of the studies described above use short, parametrically controlled stimuli (Vlek et al., 2011a,b; Kaneshiro et al., 2012; Stober et al., 2014), Schaefer et al. (2011) used naturalistic musical excerpts. However, these stimuli, though drawn from ecologically valid musical works, were fairly short (all < 5 sec) and were epoched to 3.26 sec, the length of the shortest stimulus. Treder et al. (2014) also shortened their 40-sec musical excerpts to 1.4-sec oddball epochs for classification. Thus, the analysis paradigm for single-trial classification may still be considered primarily event related.

Ongoing Responses

Multivariate approaches to analyzing brain data also facilitate the study of *ongoing* responses. Here the focus moves beyond local, event-related processing of discrete musical events toward global processing over longer time epochs. Levitin and Menon (2003) argue that this approach moves the music and speech cognition fields, historically focused heavily on anomaly processing, toward more general analyses of processing meaning. In an EEG setting, Cong et al. (2013) note that the ongoing-response paradigm combines the longer epochs and resulting temporal continuity of ongoing EEG (typically recorded while the participant is in a resting state) with event-related (often shorter) analyses. In early studies assessing ongoing responses to music using fMRI, nonmusician participants were played intact classical music excerpts as well as control versions that were scrambled in 250- to

350-ms fragments. Differential activations between stimulus conditions were analyzed to identify brain regions responding preferentially to intact stimuli (Levitin and Menon, 2003; Menon and Levitin, 2005).

A number of studies focused on ongoing responses have drawn explicitly from music information retrieval techniques, utilizing acoustical features developed specifically for music analysis (Tzanetakis and Cook, 2002). These studies use short-term (e.g., spectral flux, spectral centroid) and long-term (e.g., musical mode, pulse clarity) acoustical features, computationally extracted from musical stimuli, as a basis for quantitatively comparing stimuli with responses. A behavioral study by Alluri and Toiviainen (2010) set the foundation for this approach. Here the authors formulated perceptual scales suitable for assessing timbre of naturalistic music, and then linked human ratings of short musical excerpts to the excerpts' constituent short-term acoustical features. Subsequent fMRI studies used a refined set of short-term features, as well as long-term features, to characterize their musical stimuli. Alluri et al. (2012) identified brain regions whose fMRI time series correlated with those of the acoustical features of a tango piece, and later predicted brain activations from the features of a variety of musical excerpts (Alluri et al., 2013). Toiviainen et al. (2014) have taken the inverse approach, predicting acoustical features from fMRI-recorded responses to Beatles songs.

Acoustical feature representation has also been studied in ongoing EEG and electrocorticography (ECoG). Cong et al. (2013) used the same stimulus and long-term acoustical features as Alluri et al. (2012) in an ongoing-EEG paradigm, decomposing the EEG response into temporally independent sources using Independent Component Analysis (ICA), and then identifying sources whose frequency content corresponded to the time courses of the acoustical features. Lin et al. (2014) also used EEG-ICA sources to link ongoing-EEG responses to musical mode and tempo in shorter musical excerpts. Most recently, Sturm et al. (2015) used a regression approach to extract note onsets from EEG responses to classical music excerpts. ECoG, which offers higher SNR and localization due to the placement of electrodes directly on the surface of the cortex, has been analyzed in an ongoing paradigm to study encoding of sound intensity (Potes et al., 2012) as well as short- and long-term acoustical features (Sturm et al., 2014).

2.1.3 Criteria for Present Research

We seek an analysis technique that can operate upon responses to complete and self-contained excerpts (e.g., an entire 'movement' or 'song') of naturalistic musical works presented in a single-listen paradigm. The conventional ERP approach used most often in music EEG research, requiring short, often parametrically manipulated stimuli and hundreds or thousands of stimulus presentations, is not conducive to ecologically valid listening settings. The single-trial classification approach is amenable to naturalistic stimuli but still requires tens or hundreds of stimulus presentations in order to build the classification model; for this reason, classification studies typically involve stimuli that are substantially shorter than our musical works of interest. The ongoing-EEG paradigm, which allows for single-listen presentations, seems most promising.

2.2 Reliable Components Analysis

The previously mentioned low SNR of EEG hinders the use of single-listen experimental paradigms. One approach toward mitigating this problem is to *spatially filter* the data—that is, to derive linear weightings of electrodes subject to the optimization of some criterion. One example of spatial filtering is Principal Components Analysis (PCA), which returns orthogonal components ordered by descending variance explained. Schaefer et al. (2011) used PCA to decompose single-trial EEG responses prior to classification, and in a subsequent meta-analysis (Schaefer et al., 2013). ICA, the method used in the ongoing-EEG studies by Cong et al. (2013) and Lin et al. (2014), derives temporally independent components by maximizing joint entropy (Bell and Sejnowski, 1995; Jung et al., 1998). Other spatial filtering techniques for EEG include Common Spatial Pattern (CSP), which minimizes variance for one stimulus condition while maximizing variance for the other (Koles, 1991; Blankertz et al., 2008) and Spatio-Spectral Decomposition (SSD), which computes components based upon oscillations-related variance explained (Haufe et al., 2014).

RCA is a recently developed spatial filtering technique that maximizes mutual correlation (specifically, the Pearson Product Moment Correlation Coefficient) among data records. The method was first introduced as 'correlated components analysis' by Dmochowski et al. (2012). To date it has been successfully applied to time-domain (Dmochowski et al., 2012, 2014) and frequency-domain (Dmochowski et al., 2015) representations of scalp-recorded EEG responses.

2.2.1 RCA Derivation

Objective

Given two data matrices

$$X_1 \in \mathbb{R}^{M \times N} \text{ and } X_2 \in \mathbb{R}^{M \times N}, \tag{2.1}$$

RCA will derive a weight vector $w \in \mathbb{R}^N$ such that the projections

$$y_1 = X_1 w \text{ and } y_2 = X_2 w$$

are maximally correlated in $\mathbb{R}^M$.

In the case of EEG data, the X_i matrices comprise M time samples and N electrodes of data. As a spatial filter, RCA thus computes a linear weighting over the *electrodes* (columns) such that the resulting projected data are maximally correlated in *time* (rows). We note here that the matrix dimensions are transposed from those used conventionally in EEG research (where rows represent electrodes and columns represent time) so that the spatial filter is computed across columns of data.

Definitions

We use the definitions given by Dmochowski et al. (2012), with slight modifications to account for matrix transpositions.

First, the sample covariance matrices may be expressed as in Eq. 2.2:

$$\begin{aligned} R_{ij} &= \frac{1}{M} X_i^T X_j \\ M R_{ij} &= X_i^T X_j \end{aligned} \tag{2.2}$$

Next, we define scalar weighted power terms $\sigma_{ij} = w^T R_{ij} w$.

Finally, assuming that two EEG recordings have similar power levels, we may say that $\sigma_{11} \approx \sigma_{22}$.

Derivation

Optimization problem:

$$\begin{aligned}\hat{w} &= \arg\max_w \frac{y_1^T y_2}{\|y_1\| \, \|y_2\|} \\ &= \arg\max_w \frac{w^T X_1^T X_2 w}{\sqrt{w^T X_1^T X_1 w}\sqrt{w^T X_2^T X_2 w}} \\ &= \arg\max_w \frac{M w^T R_{12} w}{\sqrt{M w^T R_{11} w}\sqrt{M w^T R_{12} w}} \\ &= \arg\max_w \frac{w^T R_{12} w}{\sqrt{w^T R_{11} w}\sqrt{w^T R_{22} w}}\end{aligned}$$

We now take derivatives with respect to w.

For the numerator, by definition,

$$\frac{d}{dw}(w^T A w) = (A + A^T)w$$

Also,

$$\begin{aligned}R_{12}^T &= \frac{1}{M}\left(X_1^T X_2\right)^T \\ &= \frac{1}{M} X_2^T X_1 \\ &= R_{21}\end{aligned}$$

Thus,

$$\begin{aligned}\frac{d}{dw}\left(w^T R_{12} w\right) &= \left(R_{12} + R_{12}^T\right) w \\ &= \left(R_{12} + R_{21}\right) w \end{aligned} \tag{2.3}$$

For the first term of denominator, by definition,

$$\frac{d}{dw}\sqrt{u} = \frac{1}{2\sqrt{u}} u'$$

Thus,

$$\frac{d}{dw}\left(\sqrt{w^T R_{11} w}\right) = \frac{1}{2\sqrt{w^T R_{11} w}} 2R_{11} w$$
$$= \frac{R_{11} w}{\sqrt{\sigma_{11}}} \tag{2.4}$$

Similarly, for the second term of the denominator,

$$\frac{d}{dw}\left(\sqrt{w^T R_{22} w}\right) = \frac{1}{2\sqrt{w^T R_{22} w}} 2R_{22} w$$
$$= \frac{R_{22} w}{\sqrt{\sigma_{22}}} \tag{2.5}$$

We compute the product of denominator terms (Eq. 2.4 and Eq. 2.5) using the product rule:

$$\frac{d}{dw}(uv) = vu' + uv'$$

where

$$u = \sqrt{w^T R_{11} w}$$
$$= \sqrt{\sigma_{11}}$$
$$u' = \frac{R_{11} w}{\sqrt{\sigma_{11}}}$$
$$v = \sqrt{w^T R_{22} w}$$
$$= \sqrt{\sigma_{22}}$$
$$v' = \frac{R_{22} w}{\sqrt{\sigma_{22}}}$$

Thus, we have

$$\frac{d}{dw}(uv) = vu' + uv'$$
$$= \sqrt{\sigma_{22}} \frac{R_{11} w}{\sqrt{\sigma_{11}}} + \sqrt{\sigma_{11}} \frac{R_{22} w}{\sqrt{\sigma_{22}}}$$
$$= (R_{11} + R_{22}) w \tag{2.6}$$

Finally we compute the quotient of numerator (Eq. 2.3) and denominator (Eq. 2.6) using

the quotient rule:

$$\frac{d}{dw}\left(\frac{u}{v}\right) = \frac{vu' - uv'}{v^2}$$

where

$$\begin{aligned} u &= w^T R_{12} w \\ &= \sigma_{12} \\ u' &= (R_{12} + R_{21})\, w \\ v &= \sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ &= \sigma_{11} \\ v' &= (R_{11} + R_{22})\, w \end{aligned}$$

We now set the derivative to zero and solve:

$$\begin{aligned} \frac{d}{dw}\left(\frac{u}{v}\right) &= \frac{vu' - uv'}{v^2} \\ 0 &= \frac{\sigma_{11}\,(R_{12} + R_{21})\, w - \sigma_{12}\,(R_{11} + R_{22})\, w}{\sigma_{11}^2} \\ &= \sigma_{11}(R_{12} + R_{21})w - \sigma_{12}(R_{11} + R_{22})w \\ \sigma_{11}\,(R_{12} + R_{21})\, w &= \sigma_{12}\,(R_{11} + R_{22})\, w \\ (R_{11} + R_{22})^{-1}\,(R_{12} + R_{21})\, w &= \frac{\sigma_{12}}{\sigma_{11}} w, \end{aligned}$$

which can be expressed as the eigenvalue equation

$$(R_{11} + R_{22})^{-1}\,(R_{12} + R_{21})\, w = \lambda w \tag{2.7}$$

where $\lambda = \sigma_{12}/\sigma_{11}$.

As such, RCA in fact computes multiple weight vectors w_i, which correspond to the eigenvectors of Eq. 2.7. These Reliable Components (RCs) are returned in descending order of reliability explained.

2.2.2 Composition of Data Matrices

In the case where RCA is computed across K stimuli, the input matrices X_1 and X_2 (defined in Eq. 2.1) are composed of concatenated paired matrices shown in Eq. 2.8,

$$X_1 = \begin{bmatrix} S_1^1 \\ S_2^1 \\ \vdots \\ S_K^1 \end{bmatrix} \quad X_2 = \begin{bmatrix} S_1^2 \\ S_2^2 \\ \vdots \\ S_K^2 \end{bmatrix} \tag{2.8}$$

where S_i^1 and S_i^2 themselves represent concatenations of the time-by-electrodes data matrices for each stimulus $i = 1 : K$ such that all subject pairs appear across corresponding rows of these S_i matrices. For example, if our dataset comprised only three records (trials or participants) per stimulus (i.e., A_1, A_2, and A_3), the data matrices for stimulus i would be structured as follows:

$$S_i^1 = \begin{bmatrix} A_1^i \\ A_1^i \\ A_2^i \end{bmatrix} \quad S_i^2 = \begin{bmatrix} A_2^i \\ A_3^i \\ A_3^i \end{bmatrix} \tag{2.9}$$

such that all three possible pairings of the three A^i datasets occur across corresponding rows of S_i^1 and S_i^2 as notated in Eq. 2.9. It should be noted, then, that for N data records (e.g., participants or trials), a total of $\binom{N}{2}$ pairwise comparisons exist for a given stimulus, and this may be considered the effective sample size.

For K stimuli, RCA outputs a cell array of length K. Each element of this cell array contains a 3D time-by-RC-by-participants data matrix for the given stimulus, with the RC dimension having replaced the electrodes dimension. RCA also returns the electrodes-by-RC matrix W, which provides the linear weightings over the electrodes for the specified number of computed RCs, as well as the forward-model projection matrix A, used for plotting topographies. A is the same size as the weighting matrix W and is derived from W, and data covariance matrix R, as follows: $A = RW(W^T RW)^{-1}$ (Parra et al., 2005; Dmochowski et al., 2012).

2.3 Inter-Subject Correlations

2.3.1 EEG-ISCs

In the study introducing the RCA method (Dmochowski et al., 2012), the authors analyzed EEG responses from 20 participants who viewed three 6-minute film excerpts (two from famous films; one a control with footage from everyday life). Once the data were projected into one-dimensional subspaces of single RCs, the authors computed inter-subject correlations of the responses over time, across all participant pairs. Here they found that ISCs were higher for the narrative film excerpts than for the control, and higher for intact excerpts than for a time-scrambled version that was shown to a separate set of participants. Importantly, they also found that periods of heightened ISCs corresponded to moments of tension and suspense in the film excerpts.

In a subsequent EEG study, Dmochowski et al. (2014) presented participants with a 90-minute television episode with commercials (N=16), as well as SuperBowl television advertisements from 2012 and 2013 (N=12). After performing RCA and computing ISCs on these responses, the authors found, for the television show, that ISCs over time correlated significantly with both scene-related Tweet volume and with Nielsen ratings. They also found that a neural reliability score computed across RC1–RC3 correlated with Facebook-USA Today ratings of the SuperBowl ads with $\rho = 0.81$ (compared with a correlation of only $\rho = 0.51$ between neural reliability and ratings of the experimental participants). The findings from these two studies suggest not only that ISCs may reflect engagement with audiovisual stimuli, but also that ISCs computed from brain responses of a small experimental sample may generalize to large-scale population measures.

2.3.2 fMRI-ISC Studies

The use of cortical ISCs as a measure of engagement, in fact, has a longer history in the fMRI literature. This approach was first introduced by Hasson et al. (2004), who advocate the data-driven approach of ISCs for analyzing cortical responses to complex, naturalistic stimuli—a setting in which conventional, hypothesis-driven approaches would not be feasible (Hasson and Honey, 2012; Ben-Yakov et al., 2012). In this initial study, the authors analyzed responses to a 30-minute film excerpt (one used subsequently by Dmochowski et al. (2012)), demonstrating that the resulting ISCs highlight both the brain regions that 'tick collectively' during natural vision, as well as emotional and surprising moments in the film stimulus.

Subsequent fMRI-ISC studies using narrative stimuli have uncovered relationships between ISCs and successful episodic encoding (Hasson et al., 2008a), salience of different size temporal windows and directionality in time (Hasson et al., 2008c; Regev et al., 2013), shared responses across languages (Honey et al., 2012), effect of predictability in speaker-listener utterances (Dikker et al., 2014), and rhetorical quality (Schmälzle et al., 2015).

2.4 Implications for Musical Engagement

A number of studies have uncovered connections between cortical ISCs and engagement with narrative works. Notable examples using fMRI include connections to arousing film scenes (Hasson et al., 2004) and rhetorically powerful, as opposed to rhetorically weak, speeches (Schmälzle et al., 2015). The development of RCA has made ISC paradigms possible with EEG; here, ISCs have implicated suspense and tension in film scenes (Dmochowski et al., 2012) and also correlated highly with large-scale measures of viewer engagement (Dmochowski et al., 2014). Hasson et al. (2008b) propose a new interdisciplinary field of 'neurocinematics' to study the neuroscience of film. Does the promise of neural synchrony extend beyond engagement with literal narrative works such as films and speeches, into the realm of music?

It should be noted that cortical ISCs have been employed in music studies. Researchers have used fMRI-ISCs to identify brain regions that track acoustical features (Alluri et al., 2012; Trost et al., 2015), musical emotion (Trost et al., 2015), and hierarchical structural segmentation (Farbood et al., 2015), as well as those that respond preferentially to temporal and spectral structure of music (Abrams et al., 2013). ISCs have also been used with ECoG to investigate processing of sound intensity (Potes et al., 2014). Complementary to music, fMRI-ISCs have been used to assess audience responses to unedited and edited dance performances (Jola et al., 2013; Herbec et al., 2015).

The neuroscientific study of musical engagement would be facilitated by a recording modality that operates on the same time scale as music; an analysis technique that allows for both naturalistic, ecologically valid stimuli; and an experimental paradigm that operates on truly single-listen (single-trial) brain responses. The EEG-ISC approach appears to hold much promise toward meeting these criteria. EEG provides the necessary temporal resolution, while ISCs provide a means of studying responses to naturalistic, ecologically

valid stimuli. Finally, RCA is an efficient spatial filtering technique that allows for single-listen experimental paradigms.

However, no study to date has used cortical ISCs to study *engagement* with music. Furthermore, no study to date has used cortical ISCs to study EEG-recorded responses specifically to music. We will begin to address both of these matters in the next chapter, through the study of EEG-recorded responses to intact and scrambled music.

Chapter 3

Experiment 1

3.1 Introduction

In this first experiment, we seek to validate the use of EEG-ISCs—to date, applied to the analysis of cortical responses to visual and audiovisual stimuli—to study responses to ongoing naturalistic music. As this is the first application of RCA-ISC methodologies to a new stimulus modality, we closely model our experimental and analysis approaches after those taken in the first EEG-ISC study, which involved ongoing audiovisual stimuli (Dmochowski et al., 2012). In the present experiment, we are particularly interested in assessing the impact of temporal organization of the stimuli on the temporal reliability of the brain response. Here we utilize four stimulus conditions that maintain aggregate spectral characteristics, while manipulating the temporal organization of acoustical events. We collect behavioral ratings of the stimuli as an additional channel of data. Based on the findings of Dmochowski et al. (2012), our expected results are as follows: First, the topography of the most reliable component (RC1) will broadly agree with spatially filtered components derived in previous naturalistic-music EEG studies. Second, the proportion of statistically significant ISCs will be higher when musically relevant temporal structure is preserved. We may also expect to see periods of heightened ISCs during particularly salient musical events. Finally, while Dmochowski et al. (2012) observed a decrease in significant ISCs during a second viewing of film excerpts, we speculate that this effect will be mitigated for musical stimuli.

A pilot version of this experiment, using only two stimulus conditions, is presented in Kaneshiro et al. (2014). From that study, we have expanded the stimulus set, revised the

experimental design, collected new data, revised the preprocessing and analysis pipeline, and report expanded results.

3.2 Methods

3.2.1 Ethics Statement

This experiment was approved by Stanford University's Institutional Review Board as part of the study IRB-28863: Studies of Musical Learning and Expectation Using Behavioral, Physiological, and Scalp-Recorded EEG Responses. All participants delivered written informed consent prior to their participation in the experiment.

3.2.2 Stimuli

Songs

In selecting a set of songs for this experiment, we imposed the following criteria. First, we sought songs that would satisfy a 'popular yet novel' stimulus paradigm. The reasoning here was that we wanted to use songs that have been proven to engage a large, general audience (and as a result are likely easy to apprehend and enjoy on first listen) yet would be novel to our population of experimental participants (and would therefore have no confounds of familiarity or established preference, and furthermore would not complicate the experience of hearing manipulated versions of the songs). Next, so that the listener experience would be driven by musical content only, and so that our stimulus manipulations would incur no loss or deformation of meaning derived from lyrics, we sought songs containing minimal English lyrics. Finally, our planned stimulus manipulations also required that the songs have a steady beat and unchanging tempo throughout.

To satisfy these requirements, we selected recent successful Hindi pop songs. These songs have proven to effectively engage a massive audience, but would be easy to verify as unfamiliar to our participant pool. These songs were composed broadly in the pop idiom, having verses and choruses, a clear vocal line, easily discernible phrase structure, a steady beat and tempo, and sufficiently Western instrumental, melodic, and timbral palettes; such songs are created with the intention of being easy to grasp and enjoy. Our selected songs were sung in Hindi and Hindi dialects, and contained no more than sporadic, single-word occurrences of English lyrics. All songs were approximately four and a half minutes long

	Song 1	Song 2	Song 3	Song 4
Title	'Ainvayi Ainvayi'	'Daaru Desi'	'Haule Haule'	'Malang'
Movie	*Band Baaja Baaraat*	*Cocktail*	*Rab Ne Bana Di Jodi*	*Dhoom 3*
Year	2010	2012	2008	2013
Length	4:27	4:30	4:24	4:33
Tempo	156.25	93.75	90.36	86.21
iTunes id	797516730	537651546	673596457	775836478

Table 3.1: Hindi stimulus information, including song title, movie title, year of movie release, length (min:sec), tempo (BPM), and iTunes album id.

and were all in duple meter. All songs but one (sung by a male soloist) used a duet format common to Hindi pop songs—that is, a female and male singer performing in alternation. Information and metadata for the selected songs are summarized in Table 3.1.

Stimulus Manipulations

We devised a set of stimulus manipulations that would disrupt the temporal structure of the stimuli at various levels, while leaving the aggregate spectral content unchanged. First, temporally reversed control of conditions of stimuli have been employed in recent fMRI-ISC studies, in both the visual domain with silent-film stimuli (Hasson et al., 2008c), and in the auditory domain with a spoken narrative (Regev et al., 2013). This manipulation is thought to maintain sensory processing of 'instantaneous' events in the stimulus, while preventing the audience from accumulating information over time. In the case of silent films, this manipulation kept objects and characters intact, but hindered comprehension of the plot. In the case of speech, this rendered the narrative unintelligible, though the acoustical content was unchanged.

We wished also to manipulate the musical content at a shorter temporal interval. Levitin and Menon (2003) and Menon and Levitin (2005) produced a temporally scrambled control condition of classical music excerpts by shuffling the stimuli in 250- to 350-ms fragments. While we sought a similar 'scramble' paradigm, we felt that this specific procedure, which partitioned the music according to a predetermined time window and not at musically informed segmentation boundaries, might produce discontinuities of a kind that would not only disrupt a beat-based and metrical framework of the music, but might also distract the

listener through abrupt changes in amplitude. In a more recent variant on the scramble condition, Farbood et al. (2015) segmented and scrambled their musical excerpts at measure, phrase, and section boundaries. We adopt this measure-level partitioning and shuffling for the present experiment.

In their fMRI-ISC study, Abrams et al. (2013) created separate control conditions to disrupt either temporal or spectral structure of their stimuli while keeping the other attribute intact. Temporal structure was disrupted here by phase scrambling the stimuli—that is, transforming the stimulus to the frequency domain, adding a random offset to the phase of each frequency component, and converting the resulting signal back to the time domain. We adopt this approach as our final control condition, which is the most disruptive to temporal coherence.

In sum, we used the following stimulus manipulations in the current study. First, we disrupted the order in which acoustical events unfolded across time by reversing the stimulus. In this condition, musical features such as beat, meter, phrase, song part, and melody are intact, though the trajectory of musical events within a tonal framework, and perhaps the ability to form musical expectations around such trajectories, are disrupted. Timbral characteristics of the stimuli may also be affected by this manipulation, for example, for acoustical events originally characterized by sharp attacks followed by a longer decay. Next, we shuffled each stimulus at the measure level. This procedure preserves beat, meter, and short-term melodic and tonal structure, but disrupts musical trajectory and continuity at higher structural levels such as phrases and song parts. Finally, the most extreme stimulus manipulation was phase scrambling; here, all temporal structure is removed, and the effect (for the present stimuli) is a continuous texture in which the general pitch profile of the original song can be detected, but which lacks melodic and harmonic variation over time. We consider the phase-scrambled stimulus to be the true control condition, as all musically relevant structural elements are absent.

Stimulus Generation

We purchased original versions of the songs in digital format from iTunes (see Table 3.1 for album ids) and used Audacity recording and editing software, version 2.0.3,[1] to convert each .mp4 file to .wav format. Subsequent stimulus manipulations were performed in Matlab.

Our first step was to derive the base stimuli from the original versions, from which the

[1] Audacity® is copyright ©1999–2016 Audacity Team, `http://audacity.sourceforge.net/`

other three conditions would be created. For each song, we first converted the stereo .wav file to mono by taking the average of the channels. Next, we used publicly available beat-tracking software (Ellis, 2007) to extract beat onsets, which were then corrected manually to account for tempo octave errors (Levy, 2011). The resulting tempos ranged from 86.2–156.25 beats per minute (BPM), a common range for music (Moelants and McKinney, 2004). Using measure onsets derived from the beat onsets, we trimmed excess silence from the beginning and end of each audio file so that each song comprised an integer number of measures. We henceforth refer to these versions of the audio as the 'original' versions of the songs.

We derived the other three versions of the stimuli from the original versions. The reversed condition was created by reversing the audio signals. For the measure-shuffled conditions, we applied the beat-tracking procedure once more (since the trimming process affected beat onset times) and re-derived measure boundaries. We then shuffled each song's audio waveform at the measure level. Finally, for phase scrambling, we used the following procedure: For each song, we used the FFT to transform the audio signal from the time domain to the frequency domain. We then independently randomized the phase value of each positive frequency (0 to $fs/2$) to a value between 0 and 2π, and assigned conjugate-symmetric values of the positive-frequency phases to the negative frequencies to preserve phase antisymmetry of real signals (Smith, 2011). Finally, the phase-scrambled frequency representation was transformed back to the time domain using the IFFT. This procedure is similar to that described in Prichard and Theiler (1994) and used by Abrams et al. (2013); those studies introduced a random phase shift over the $(0, 2\pi)$ interval at each frequency bin rather than replacing the values with uniformly sampled random variables outright.

We thus created 16 stimuli total: Four songs, and four versions of each song. Audio waveforms, spectrograms, and magnitude spectra of the four versions of Song 1 are shown in Figure 3.1. Inspection of the original and reversed waveforms shows some evidence of distinct, repeated song parts; the measure-shuffled stimulus lacks the song-part segmentation of the previous versions and contains more fluctuations in spectral and temporal activity over short time periods. In contrast, the spectral content of the phase-scrambled stimulus is smeared across time, and the audio waveform has a fairly static amplitude envelope. However, the magnitude spectra are consistent across stimulus conditions. Stimulus figures for the other three songs can be found in §A.1.

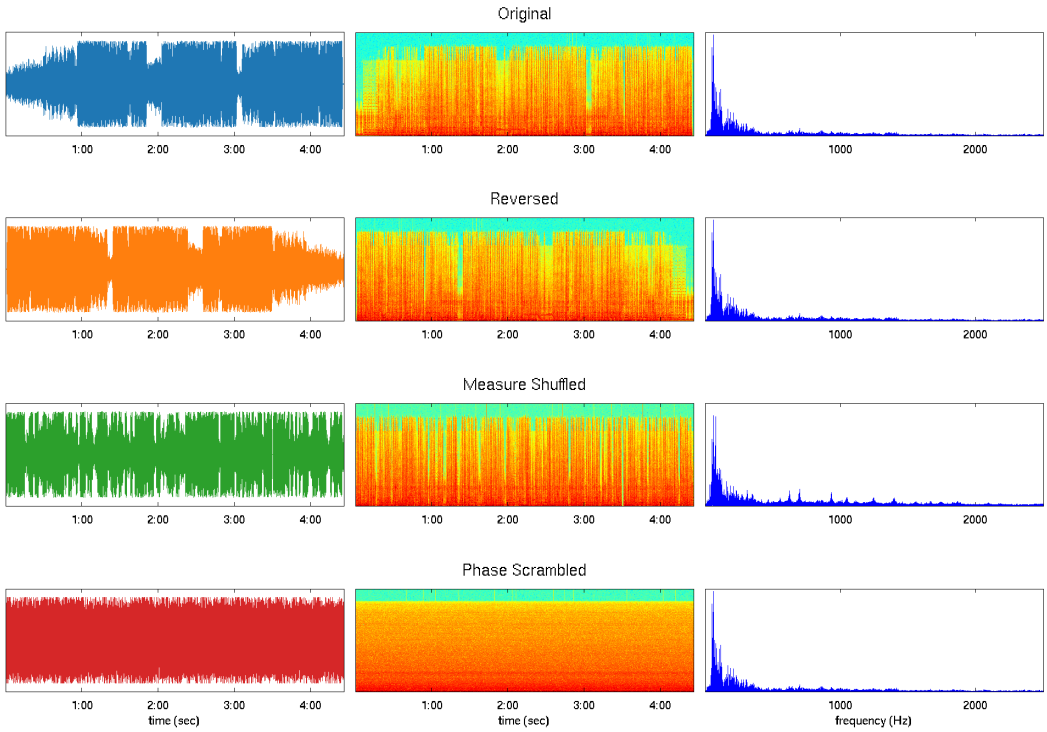

Figure 3.1: Waveforms, spectrograms, and magnitude spectra (up to 2,500 Hz) of Song 1 stimuli. The stimulus waveforms (time domain) and spectrograms (time-frequency domain) vary by stimulus condition. However, the aggregate spectral content across each excerpt is unchanged across original, reversed, and phase-scrambled conditions. The magnitude spectrum of the measure-shuffled stimulus is slightly altered by discontinuities introduced by the shuffling procedure.

3.2.3 Participants

For this experiment we recruited right-handed participants with normal hearing, between 18–35 years of age, who were fluent in English and had no cognitive or decisional impairments. To ensure that the songs would be novel and that no meaning would be imparted by their lyrics, participants were required to have no experience with Hindi language, films, or music. To maximize the potential for engagement with our song set, we recruited participants who reportedly enjoyed listening to music from a variety of genres including pop,

rock, and classical, and who listened to music for at least three hours per week. Finally, as absolute pitch is known to impact cortical organization for music processing (Loui et al., 2011), we required that participants not have absolute pitch. While cortical responses have been shown to be enhanced by formal musical training (Pantev et al., 1998), for this first attempt we seek results that generalize across the general population, and therefore had no requirements related to formal musical training.

Forty-eight participants produced usable datasets (see §3.2.3).[2] All participants met the eligibility requirements described above. Participants ranged from 18–34 years of age (mean = 24.58 years); 25 were male and 23 were female. Twenty participants reported being involved in musical activities at the time of their experimental sessions. Thirty-two reported having received formal musical training; of these, the total duration of training ranged from 3 months to 22 years (mean = 7.57 years). Music listening ranged from 3 to 52.5 hours per week (mean = 15.03 hours).

3.2.4 Experimental Paradigm and Data Acquisition

We assigned each participant four of the 16 stimuli using a Latin square design, wherein each participant was assigned each of the four songs once, in a different stimulus condition for every song. Therefore, each participant was exposed to each song and each stimulus condition exactly once. This procedure not only ensured independent samples when comparing responses to four conditions of a given song or four songs within a given condition, but also meant that participants only ever heard one version of each song. Thus, as all songs were unfamiliar, participants had no basis for comparison regarding the coherence, or 'rightness', of a given song. Under this assignment procedure, 4! = 24 stimulus assignments were possible, and each possible assignment was used twice across the pool of 48 participants, resulting in a total of 12 participants assigned to each of the 16 stimuli.

After delivering written informed consent, each participant began the experimental session by filling out a demographic and musical experience questionnaire. Following this, the experimenter familiarized the participant with the experimental booth and guided the participant through a training run-through of the experiment using 15-second stimuli not used in the actual experiment. Once the participant was familiar with the layout of the booth, the task, and the keyboard interface for answering questions, he was fitted with the electrode net and began the experimental blocks.

[2] In total, 58 participants took part in the experiment.

Each participant completed two experimental blocks. In each block, the four assigned stimuli were presented once in random order. A separate recording was taken for each block, before which electrode impedances were checked and brought under threshold. Participants were instructed to sit still (avoiding head and body movements), focus their eyes on a fixation point presented on a monitor located 57 cm in front of them, and listen attentively while the stimuli played; no other task was performed while the stimuli were playing. At the conclusion of every stimulus, the following questions were presented on-screen one at a time, and the participant delivered responses using the computer keyboard in front of him:

1. How pleasant was this excerpt, on a scale of 1 (not pleasant at all) to 9 (very pleasant)?
2. How musical was this excerpt, on a scale of 1 (not musical at all) to 9 (very musical)?
3. How well ordered was this excerpt, on a scale of 1 (not ordered at all) to 9 (very well ordered)?
4. How much of the excerpt was interesting, on a scale of 1 (none of it) to 9 (all of it)?

In sum, we collected from each participant two listens to his assigned four stimuli, for a total of eight trials.

The participant was seated in a darkened, acoustically and electrically shielded booth (ETS-Lindgren) during the experimental blocks. EEG responses were recorded using the EGI GES 300 system (Tucker, 1993). Data were acquired at a sampling rate of 1 kHz with vertex reference, using unshielded 128-channel HCGSN 110 and 130 nets which connected to a Net Amps 300 amplifier. Amplified EEG signals were recorded using Net Station software (version 4.5.7) on a Power Mac G5 desktop computer running the OSX operating system, version 10.6.8.

The experiment was programmed using Neurobehavioral Systems Presentation software (version 16.5, build 09.17.13),[3] on a Dell Inspiron 3521 laptop running the Windows 7 Professional operating system. The laptop was synced to a keyboard, mouse, and Samsung SyncMaster PX2370 LED monitor in the experiment booth. Triggers for stimulus labels sent by the Presentation software, as well as keyboard responses delivered by the participant in the booth, were output from the stimulus computer via USB to a National Instruments USB-2008 D-to-A converter emulating a printer port. From that device, all but one pin of a modified DE-9P (DB9) cable delivered numbered trigger labels to the DIN_1 input of the EGI amplifier.

Stereo audio signals were output from the stimulus laptop via USB to an external sound

[3] `https://www.neurobs.com/`

card (Native Instruments Komplete Audio 6). From there, the audio channel was split: The channel containing the auditory stimulus was routed to a Behringer Xenyx 502 mixer, which split the mono signal and delivered it to two magnetically shielded Genelec 1030A speakers located 120 cm from the participant in the booth. The second audio channel, containing intermittent clicks for precise time-locking of the stimulus to the EEG recording, was output to the remaining pin of the DB9 cable and delivered to the DIN_1 input of the EGI amplifier along with the other numbered trigger labels.

3.2.5 EEG Preprocessing and Analysis

EEG Preprocessing

Prior to data export, we used Net Station's Waveform Tools software to filter and downsample the EEG recordings. We applied a zero-phase bandpass filter (0.3–50 Hz) to each recording's data frame, and then downsampled the resulting data by a factor of 8 to a sampling rate of 125 Hz. Following this, the data were exported to .mat file format. All subsequent analyses were performed in Matlab.

The present experiment contains a low number of trials that are relatively long in duration. As a result, it is costly to discard a trial. With the present experimental design, we in fact discarded all data from a participant if any one trial was deemed unusable. This approach is in contrast to more traditional experimental paradigms for EEG—using short, repeated trials—in which enough trials of a given stimulus are collected that unusable trials may often be simply excluded from analysis. In the present scenario, therefore, it was critical to clean the data as well as possible of malfunctioning or noisy electrodes, ocular artifacts, movement-related artifacts, and other noisy transients, in order to retain a participant's data. On the other hand, one benefit of the RCA algorithm is its ability to handle missing data. Therefore, we had the option to replace data from bad electrodes and noisy transients with missing values (NaNs), and furthermore did not need to impute missing data by taking, for example, a spatial average of neighboring electrodes.

With these constraints in mind, we devised and coded a custom software pipeline in Matlab to preprocess the data. Each EEG recording, corresponding to one experimental block from one participant, underwent the following procedure. A package of helper functions used in this procedure have been made available for public download.[4]

[4] `https://github.com/blairkan/MatEEGPreproc`

First, after loading in a given recording .mat file, initial preprocessing steps included annotating the file with net ID, name of experimenter, and name of analyzer. Following this, we extracted the trigger labels and timestamps from the DIN_1 variable output by Net Station. We used the click onsets from the secondary audio channel to correct the timestamps sent by Presentation specifying the start of a trial. We also computed and saved information pertaining to onset timing errors and playback rate error across each trial. Next, we extracted the behavioral ratings delivered by the participant at the end of each trial.

The next stage involves the EEG data. Recall that EEG responses were recorded by 128 monopolar electrodes plus a vertex reference. The resulting data frame output by Net Station for a given recording was a 129-by-time matrix. For every trial in a given recording, we epoched the relevant data, removed the linear trend of each electrode across the trial, and performed a median-based DC-offset correction of each electrode[5]. The four trial epochs for a given recording were then concatenated into a single electrodes-by-time matrix. Next, we retained electrodes 1 through 124 for further analysis (excluding the electrodes on the face). We then identified bad electrodes based upon impedances and experimenter notes from experimental sessions, a percent-over-voltage threshold used in a later stage of the preprocessing procedure, and manual inspection of the data. Rows of data corresponding to bad electrodes were removed from the data matrix at this time—that is, the matrix became smaller along the row (electrode) dimension. Finally, we computed horizontal and vertical electrooculogram (HEOG and VEOG) channels, to be used later for removing ocular artifacts from the data.

Ocular activity, such as eye blinks and eye movements, introduces high-amplitude artifacts into EEG data. These artifacts may be addressed in a number of ways, including exclusion of contaminated trials, or regression-based approaches to recover the underlying brain signal. Here, we removed ocular artifacts from the data using a validated approach for EEG involving ICA (Bell and Sejnowski, 1995; Jung et al., 1998). We performed ICA over each recording's concatenated-epoch data frame using the Matlab EEGLAB toolbox implementation of the extended Infomax ICA algorithm (Delorme and Makeig, 2004). Once the unmixing matrix W was computed, we used it to convert the recording's data from electrode

[5]Linear trend and DC offset of electrode data affect the performance of subsequent stages of preprocessing and analysis, including ICA (Groppe et al., 2009), detection of transients, and RCA. Trend and DC offset should have been removed by the highpass component of the EGI bandpass filter. However, as we found these artifacts still present in the exported data frames, we repeated these steps here.

space to component space, $X_{ICA} = WX_{Raw}$ (i.e., rows now represent activations of independent components). We then correlated the time course of every component with the time courses of the HEOG and VOEG channels. Any component whose magnitude correlation with either EOG channel met or exceeded a fixed threshold ($|\rho| \geq 0.3$) was automatically flagged as an EOG component. Additional components for which $0.2 \leq |\rho| < 0.3$ were selectively flagged as EOG components on the basis of manual inspection of their forward-model projection topographies (Parra et al., 2005) and characteristics of their temporal activations. Following this, the temporal activations of all identified EOG components were replaced with rows of zeros in the component-space matrix, and the data were converted back to 'clean' electrode space using the inverse of the unmixing matrix, $X_{Clean} = W^{-1}X_{ICA}$.

Removal of high-amplitude eye artifacts facilitates identification of corrupted electrodes in the data. At this stage, we adopted a percent-over-threshold procedure similar to that used in Dmochowski et al. (2015). For the present study, any electrode for which at least 10% of voltage magnitudes exceeded 50 μV across the recording was marked as a recording-wide bad electrode, and the entire preprocessing procedure was re-started with that electrode removed.[6] Once no additional recording-wide bad electrodes were identified through this process, we proceeded to search for trial-wide bad electrodes; these were defined to be any electrodes for which at least 10% of voltage magnitudes exceeded 50 μV within a given trial. The rows in the data frame corresponding to these electrodes were removed from the data frame for only the trial(s) in which the electrode was flagged. The final preprocessing steps were performed on a trial-by-trial basis: We first removed noisy transients from the data by setting EEG samples whose magnitude voltage exceeded four standard deviations of its channel's mean power to NaNs, with the procedure repeated four times in an iterative fashion. Next, missing rows in the data matrix corresponding to recording- or trial-wide bad electrodes were re-constituted and filled with NaNs, ensuring that all data frames contained the same number of rows and that rows corresponded to the same electrodes across recordings. Next, we appended a row of zeros to the data matrix, representing the temporal activation of the vertex electrode, and converted the resulting data matrix to average reference by subtracting, from each time point (column), the mean instantaneous amplitude across all electrodes.[7] Cleaned data frames were thus of size 125-by-time. The

[6] This action was chosen over simply removing the electrode at this point because we have found that removing bad electrodes in early preprocessing steps (prior to ICA) improves ICA performance and leads to more effective removal of ocular artifacts.

[7] Performing the average referencing step after removing transients introduces a risk of slight discontinuities in the data for time points where data samples were set to NaN, especially across several electrodes;

cleaned data epochs corresponding to trials of a given recording were stored in a cell array.

It was necessary to collect data from 58 participants in order to obtain 48 sets of usable data. Unusable datasets were identified during data collection and preprocessing and were excluded for the following reasons: Gross noise artifacts during data collection (3 participants); 20 or more bad electrodes (4 participants); participant did not follow instructions—eyes closed (1 participant); we learned, during the experimental session, that the participant did not meet eligibility criteria for the experiment (2 participants). The stimuli and stimulus orderings assigned to these participants were re-assigned to subsequent participants for collection of replacement data.

Once data from all recordings were preprocessed, the cleaned data frames for a given stimulus and listen were aggregated across participants into a single three-dimensional electrodes-by-time-by-participants matrix. As the experiment comprised 16 stimuli with two listens per stimulus and 12 participants assigned to each stimulus, the complete dataset comprised 32 such matrices, where the number of electrodes was always 125, the number of time samples varied according to the stimulus, and the number of participants was always 12. These cleaned and aggregated datasets have been anonymized and made publicly available for download through the Stanford Digital Repository (Kaneshiro et al., 2016a).[8]

Data Analysis

For the present computation of reliable components, we utilized the publicly available RCA codebase released with Dmochowski et al. (2015).[9] The RCA procedure used here is as outlined in Dmochowski et al. (2012) and in the previous chapter. For our primary analyses and for computation of ISCs, we computed RCA over the full set of 32 response matrices (all participants, stimuli, and listens). For comparison of scalp topographies only, we additionally computed RCA separately for each stimulus condition and listen (eight RCA runs comprising four stimuli apiece). We computed the first 5 RCs for all RCA computations. Thus, the output time-by-RC-by-participants matrix for a given stimulus was of size $T \times 5 \times 125$, where T varied according to the stimulus.

however, we felt that this was preferable to converting the data to average reference before removing transients, and possibly propagating large transients across several electrodes.

[8] `http://purl.stanford.edu/sd922db3535`

[9] `https://github.com/dmochow/rca`

Inter-Subject EEG Correlations

Time-resolved inter-subject EEG correlations for individual stimuli were computed in the component subspace of single RCs. ISCs were computed using a 5-second correlation window that advanced in 1-second increments, for an effective temporal resolution of 1 Hz. Within each windowing frame, the cross-correlation of every participant pair of data (effective sample size of $\binom{12}{2} = 66$ pairwise comparisons for 12 participants) was computed. We report the mean correlation across the subject pairs for every time window. Each time-resolved ISC is plotted and interpreted at the midpoint of its temporal window (e.g., the ISC computed from 0–5 seconds is mapped to 2.5 seconds).

3.2.6 Extraction of Stimulus Features

We used the publicly available LabROSA `myspecgram` Matlab implementation[10] to compute spectrograms of the stimuli for visualization purposes (Figure 3.1). In order to compare the time course of the EEG-ISCs with the amplitude envelopes of the stimuli, we extracted the amplitude envelope of each stimulus using the MIRtoolbox, version 1.5 (Lartillot and Toiviainen, 2007).[11] Amplitude envelopes were extracted at a sampling rate of 1 Hz in order to match the sampling rate of the ISC time series. Song parts of the stimuli were human-annotated.

3.2.7 Statistical Analyses

Because of autocorrelation characteristics of the ISC time series (Sturm et al., 2014), statistical significance of these results over the course of a given stimulus was assessed via permutation test (Fisher, 1971). For the present experiment, the following procedure was performed 500 times:

1. Participants' data frames were partitioned into non-overlapping 5-second windows, and the windows for each participant's data were shuffled independently;
2. Time-resolved pairwise ISCs were calculated over the collection of shuffled time series.

We used a threshold of $\alpha = 0.05$ to assess significance over all of the permutation iterations; thus, any temporal window in which the ISCs of the intact data exceeded the 0.95 quantile across the 500 permutation iterations was deemed to contain a statistically significant ISC.

[10] `http://labrosa.ee.columbia.edu/matlab/sgram/`
[11] `https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox`

The proportion of significant ISCs across an entire stimulus, then, was the proportion of ISC frames for which the mean value of the intact-data ISCs strictly exceeded the significance threshold from the permutation iterations.

In comparing ISC time series from the first versus the second listen of a given stimulus, it was possible to perform a paired test over the vector of differences in mean ISCs between the listens at every time point. Past findings point to an exposure effect, whereby the ISC time course is generally lower on the second exposure (Dmochowski et al., 2012). To test this hypothesis, we subtracted, for each stimulus, the mean ISC time series of the second listen from the mean time series of the first listen. We then performed a one-tailed Wilcoxon signed-rank test (Lehmann, 2006) to determine whether the values in the difference vector come from a distribution with median greater than zero. We then controlled for False Discovery Rate (FDR) (Benjamini and Yekutieli, 2001) over the vector of resulting p-values from these tests across all 16 stimuli. We report p-values of individual tests and specify which are statistically significant ($p < 0.05$ after FDR correction) or marginally significant ($0.05 \leq p < 0.1$ after FDR correction).

In relating the stimulus amplitude envelopes with the ISC time series, we correlated ISCs with amplitude envelopes, as well as with the rectified difference of the envelope, which represents the magnitude change in amplitude between time samples. The rectified difference envelope DE of amplitude envelope AE can be expressed as

$$DE_i = |AE_{i+1} - AE_i|, i < \text{length}(AE) \tag{3.1}$$

Each resulting collection of p-values was then corrected for FDR. We report p-values from all correlations, along with whether the p-value is statistically significant ($p < 0.05$ after FDR correction) or marginally significant ($0.05 \leq p < 0.1$ after FDR correction).

3.3 Results

3.3.1 Behavioral Ratings

The behavioral ratings of the Hindi stimuli are plotted in Figure 3.2. Responses have been aggregated across participants and songs for each stimulus condition and are separated by question (row) and listen (column). For every question and listen, the original stimuli (blue) were always rated highest overall, while the phase-scrambled stimuli (red) always received

the lowest ratings. Reversed and measure-shuffled stimuli received middling ratings. These results, while not the main focus of our present analysis, validate the intended impact of stimulus manipulations on perceptual judgments of the songs.

Figure 3.2: Behavioral ratings of Hindi songs. Ratings are aggregated across all participants and stimuli and separated by question (row) and listen (column). Original versions of songs (blue) receive highest ratings of pleasantness, musicality, orderedness, and interestingness overall, while phase-scrambled versions (red) receive lowest ratings for all four questions.

3.3.2 EEG Results

RC1 Topographies

The forward-model projected topographies of RC1–RC3 are shown in Figure 3.3. RC1 presents a fronto-central topography. This topography is similar to the RC1 topography derived from responses to intact Hindi stimuli in our pilot study (Kaneshiro et al., 2014); it is also roughly consistent with PC1 topographies from single-trial classification studies using shorter excerpts of naturalistic music (Schaefer et al., 2011), topography of grand-averaged ERPs around 100–200 msec after beat onsets (Stober et al., 2015) as well as the MUSIC component derived by Sturm et al. (2015) in response to naturalistic music excerpts. RC2

appears to highlight right-lateralized temporal and medial parietal electrode sites, while RC3 presents a medial parietal topography that is roughly similar to a subset of PCA and tensor components derived in Schaefer et al. (2011) and Schaefer et al. (2013).

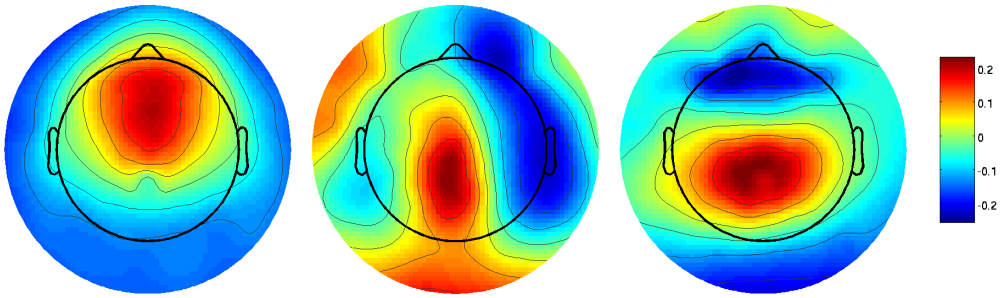

Figure 3.3: RC1–RC3 topographies. RCA was computed across the full set of responses, incorporating all stimuli and both stimulus presentations. All subplots are scaled to the same colorbar. RC1 presents a fronto-central topography, while RC2 and RC3 implicate temporal and parietal electrodes.

In our pilot study (Kaneshiro et al., 2014), we found that the RC1 topography derived from responses to phase-scrambled stimuli lacked physiological plausibility. To investigate whether that topography would be replicated here, and to assess the component topographies for the other stimulus conditions across repeated listens, we performed separate RCA computations for each stimulus condition and listen. The resulting RC1 topographies, shown in Figure 3.4, are consistent across stimulus conditions and listens for all responses except those to the phase-scrambled stimuli. The phase-scrambled RC1 topographies are not only inconsistent with those of the other conditions; they differ within-condition from the first to the second listen. Thus, it appears that all stimulus conditions retaining musically informed temporal variations produce consistent RC1 topographies.

Inter-Subject Correlations

We next computed time-resolved RC1 and RC2 ISCs for the first-listen responses to all stimuli. The time-resolved RC1 ISCs for the four conditions of the first-listen responses to Song 1 are shown in Figure 3.5. Any portion of an ISC time course exceeding the top boundary of the shaded gray area exceeds the 0.95 quantile of the permutation iterations and is considered statistically significant ($\alpha = 0.05$). As can be appreciated by an inspection of the plot, the first three stimulus conditions (rows) produce statistically significant ISCs

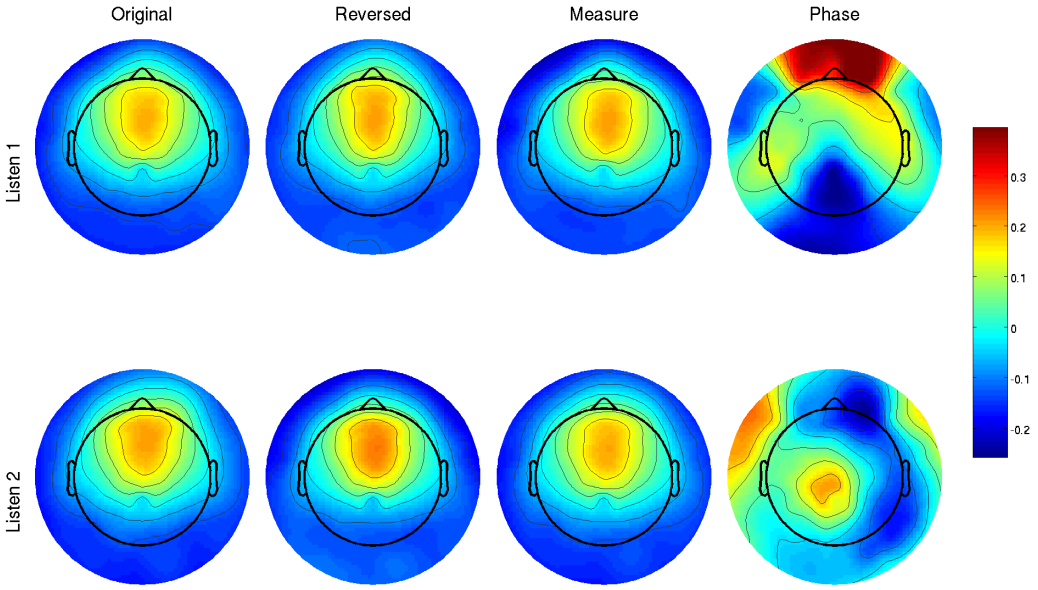

Figure 3.4: RC1 topographies by stimulus condition (columns) and listen (rows). Subplots share a consistent color scale. Responses to all stimulus conditions except the phase-scrambled stimuli reproduce the fronto-central RC1 that was derived across the full collection of responses. RC1 topographies for the phase-scrambled stimuli, however, differ both from those of the other stimuli and from one another across listens.

over the course of the stimuli; percentages of significant ISCs for these conditions range from 20.23% (reversed condition) to 46.18% (measure-shuffled condition) for this song. In contrast, only 6.11% of ISCs are statistically significant for the phase-scrambled version for Song 1.

We note also that the proportions of significant ISCs (right-hand plots) for Song 1 are lower for RC2 than RC1 (left-hand plots) for the three stimulus conditions (original, reversed, measure shuffled) whose RC1 topographies were similar to the aggregate RC1 topography, demonstrating that mutual correlation was successfully maximized in the first RC for these conditions. For the phase-scrambled condition, ISCs are slightly higher for RC2 than RC1, highlighting the effect of subjecting these data to a spatial filter that was not maximizing mutual correlation for this stimulus condition. The time-resolved ISCs for the other three songs are included in §A.2.1.

Dmochowski et al. (2012) showed that temporally resolved ISCs of different RCs reached

Figure 3.5: Time-resolved RC1 and RC2 ISCs for Song 1. EEG records were transformed from electrode space to RC space, and temporally resolved ISCs were computed in the subspace of the first two most reliable components. Results shown are for RC1 (left) and RC2 (right) for the four stimulus conditions of Song 1.

statistical significance at different times for a given original stimulus. The authors interpreted this finding as reflecting that the different RCs perhaps reflect processing of different stimulus features. For the present study, RC1 and RC2 ISCs are plotted together for each of the original songs in Figure 3.6. Here we see some variability among the temporal activations of the RCs for a given song, for example between 3:30–4:00 of Song 1. However, there do emerge some regions where the ISCs of the two RCs are enhanced at the same time, for example, around 3:10 of Song 4. As can be seen from the bar plots on the right-hand sides of the figures, the proportion of significant ISCs for original versions of songs is always higher for RC1 than for RC2. Overlaid RC1/RC2 plots for the other three versions of the songs can be viewed in §A.2.2.

The proportion of significant RC1 and RC2 ISCs for first-listen responses to all stimuli are summarized in Figure 3.7. Here we see that indeed, the proportion of statistically significant RC1 ISCs is higher than RC2 ISCs for all stimulus conditions except for the phase-scrambled stimuli, where the proportion of significant ISCs is higher for RC2 than for

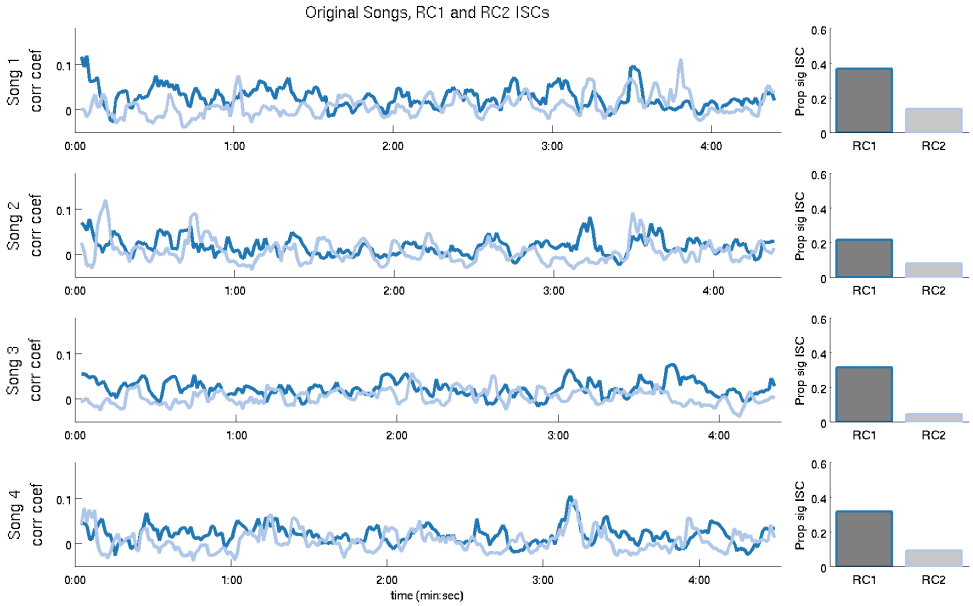

Figure 3.6: Time-resolved RC1 (dark blue) and RC2 (light blue) ISCs for all original songs. Left: ISC peaks of the RCs for a given song are sometimes disparate, and sometimes co-occur. Right: For all original songs, RC1 produces a higher proportion of significant ISCs than does RC2.

RC1 for three of the four songs. Again, this likely reflects the fact that the RC1 topography specific to these responses to phase-scrambled stimuli differs from the RC1 topography derived from the full set of responses.

First Versus Second Listen

In their initial EEG-ISC study Dmochowski et al. (2012) revealed an exposure effect—that is, the proportion of significant ISCs decreased upon second viewings of audiovisual film excerpts. We were interested to determine whether this would be the case also for music. Considering that repetition—both of structural and thematic elements within a piece of music and through repeated listens of complete works—is a more prominent feature of composition and consumption for music than for literal narrative works such as films, we conjectured that ISCs for the present experiment would not necessarily decrease in the second exposure. The proportion of significant RC1 ISCs across the first and second listens of each stimulus are summarized in Figure 3.8. Here we can see that the proportion of

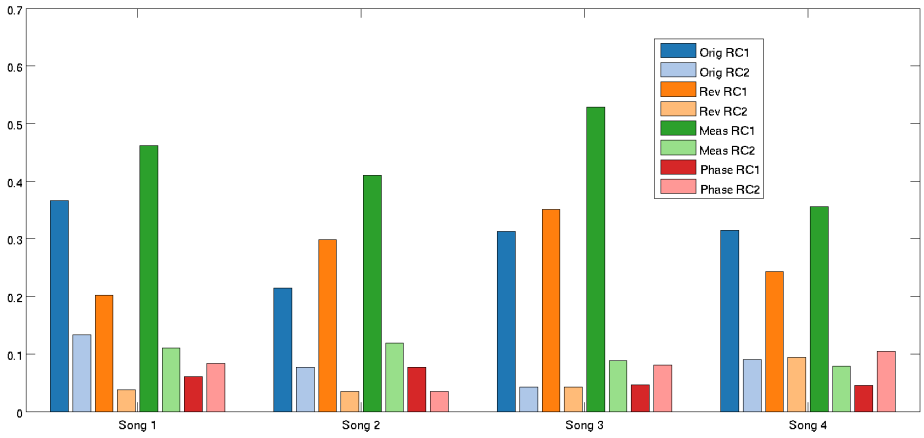

Figure 3.7: Proportion of significant RC1 and RC2 ISCs, first listen. Measure-shuffled stimuli always produce the greatest proportion of statistically significant RC1 ISCs. Proportions of statistically significant ISCs are higher for RC1 than for RC2 for all but the phase-scrambled stimulus conditions.

significant ISCs decreases from the first listen to the second for some stimuli, but does not do so consistently across the set.

The above results provide insight into the change of ISCs from the first to second listen across entire songs. Since both listens of a given stimulus involve responses over time to identical content from identical populations and RCs, we can also plot the first- and second-listen time series together, as shown for the original versions of the four songs in Figure 3.9. We note here that in certain regions of the songs, for example, around 1:30 and 3:30 in Song 1 and the opening, 1:50, and 3:20 of Song 4, the ISC activations are somewhat aligned between listens. In other regions, relation between time series is less clear. First- versus second-listen RC1 ISC time series for the other stimulus conditions are included in §A.2.3

As first- and second-listen ISC time courses reflect a common underlying stimulus, we can quantify the differences between the time series in a temporally resolved fashion. For each stimulus we performed a one-tailed Wilcoxon signed-rank test on the difference of the first- and second- listen RC1 ISC time series. The results are summarized in Table 3.2. Here we observe a statistically significant drop in ISCs in the second listen for some of the stimuli. Notably, all but the phase-scrambled versions of Song 3, and all of the reversed stimuli except for Song 4, show significant drops. However, this exposure effect does not generalize across all songs.

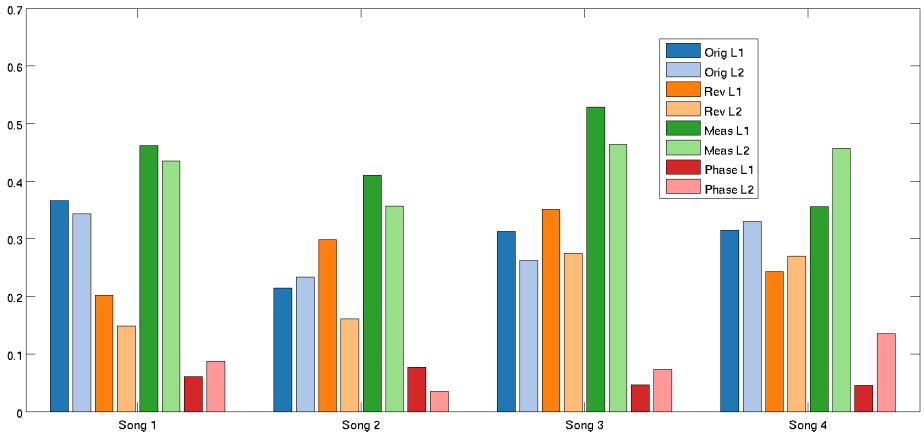

Figure 3.8: Proportion of significant RC1 ISCs, first versus second listen.

		Stimulus Condition			
		Orig	Rev	Meas	Phase
Song	1	0.0436*	0.0000**	0.0067**	0.9852
	2	0.1405	0.0000**	0.2054	0.0476*
	3	0.0000**	0.0117**	0.0002**	0.1855
	4	0.7142	0.1557	0.9411	0.8804

Table 3.2: Wilcoxon test results comparing listen-1 and listen-2 RC1 ISCs. For every stimulus, we performed a one-tailed Wilcoxon signed-rank test to determine whether the first-listen ISCs were strictly greater than the second-listen ISCs across time. Statistically significant results suggest that ISCs collected in the first listen were higher. Two asterisks '**' denote statistical significance after correction for FDR ($p < 0.05$); one asterisk '*' denotes marginal statistical significance after correction for FDR ($0.05 \leq p < 0.1$)

Relating ISCs to Stimulus Features

Finally, while Dmochowski et al. (2012) did not systematically analyze links between ISC peaks and stimulus features, they anecdotally noted relations between peaks and salient (for example, suspenseful or arousing) events in the corresponding film stimulus. To see whether similar insights pertaining to musical structure might emerge here, we plotted the time-resolved first-listen RC1 ISCs for the original versions of the songs over human-annotated structural elements (song parts). As shown in Figure 3.10, ISC peaks are not linked to a single specific song part, but occur throughout the songs. Interestingly, many of the statistically significant peaks occur around transitions between song parts, for example

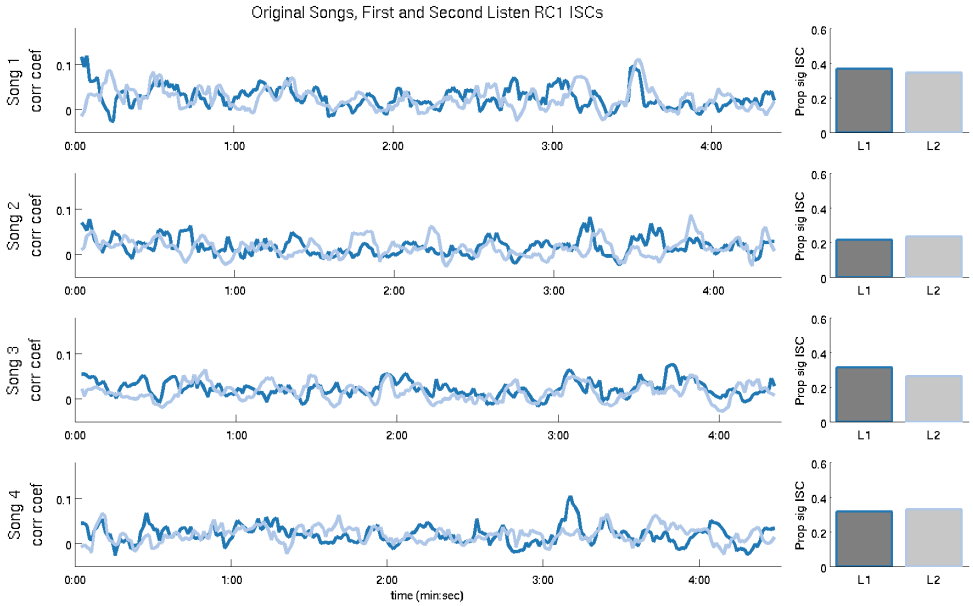

Figure 3.9: RC1 ISCs of original stimuli, first versus second listen. We wished to assess whether ISCs are lower in the second presentation of the stimulus. One-tailed Wilcoxon signed-rank tests of first-listen minus second-listen ISCs reveal a statistically significant exposure effect for Song 3, and a marginally significant effect for Song 1 (Table 3.2)

around 1:07, 1:20, 3:00, and 3:30 for Song 1; 0:40 and 3:35 for Song 2; and around 3:10 for Song 3 and Song 4.

In considering responses to music, sound intensity is known to affect the amplitude of auditory evoked responses (Mulert et al., 2005). In a supplementary analysis in their single-trial classification study, Schaefer et al. (2011) found varying degrees of correlation between PC1 (first Principal Component) activations and amplitude envelopes of short, naturalistic-music stimuli. We performed a similar analysis, except on the time course of the ISCs rather than on the RC1 activations themselves. For the present experiment, we note that the measure-shuffled stimuli, which contained noticeably abrupt changes in amplitude envelope, led to the greatest proportion of significant EEG-ISCs overall. While increased *reliability* of neural responses does not necessarily imply increased *amplitude* or vice versa, we explored this matter further by correlating each RC1 ISC time course with the both the amplitude envelope and the rectified difference envelope of the corresponding stimulus.

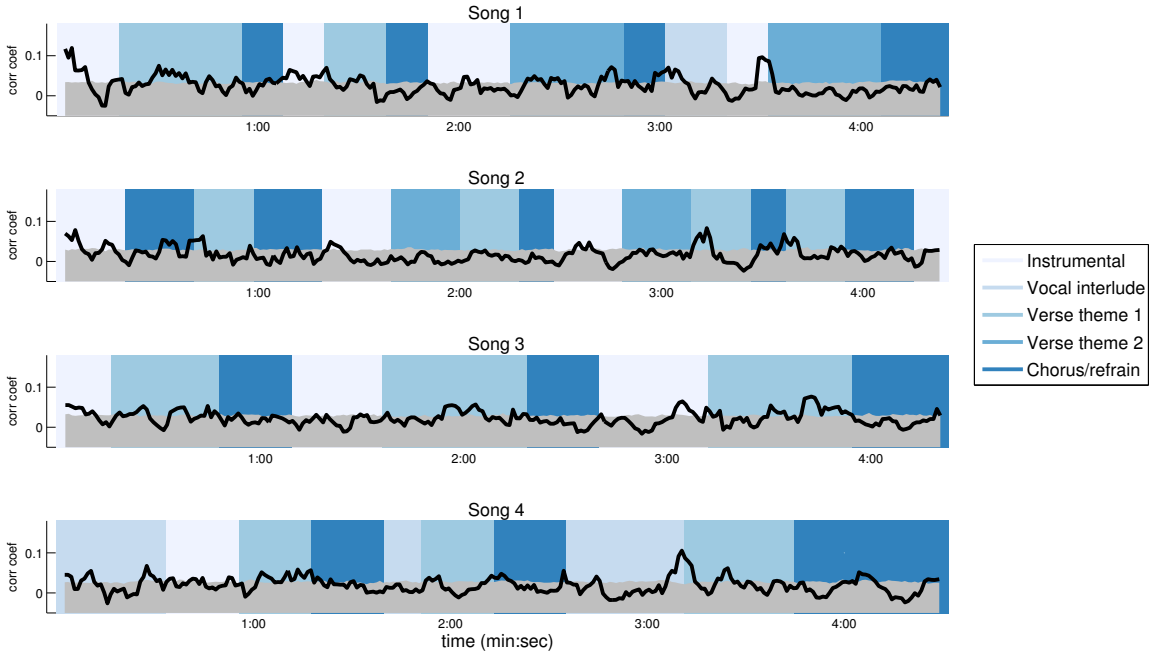

Figure 3.10: First-listen RC1 ISCs of original stimuli, plotted over song parts. ISC peaks occur during various song parts, and occasionally around transitions between song parts.

Figure 3.11 shows the time-resolved RC1 first-listen ISCs for the four versions of Song 4, a song that shows some of the more striking relationships between ISCs and amplitude envelope. Normalized ISC time series are plotted in color, while normalized amplitude envelopes are plotted in black, and normalized rectified difference vectors in gray. As we can see from this figure, there are several points at which ISC peaks correspond to fluctuations in amplitude envelope. For example, notable effects of negative amplitude fluctuations are visible around 3:10 for the original condition, 3:40 and 4:00 in the reversed condition, and 0:20, 1:15, 2:05 in the measure-shuffled condition. However, drops in amplitude do not always drive ISC peaks, as at 1:20 for the reversed stimulus and 4:00 in the measure-shuffled stimulus. The correlation coefficients for all of the stimuli and listens are summarized in Table 3.3. Interestingly, the highest correlations between EEG-ISCs and amplitude envelopes are produced by the original stimuli, not the measure-shuffled stimuli as we had guessed. Amplitude-envelope plots for the other three songs are included in §A.2.4.

As a follow-up analysis in relating ISCs to the amplitude envelopes of their stimuli, our final analysis took a first look at the effect of temporal reversal on ISC time series derived from responses to identical acoustical content. If ISCs are driven solely by 'instantaneous'

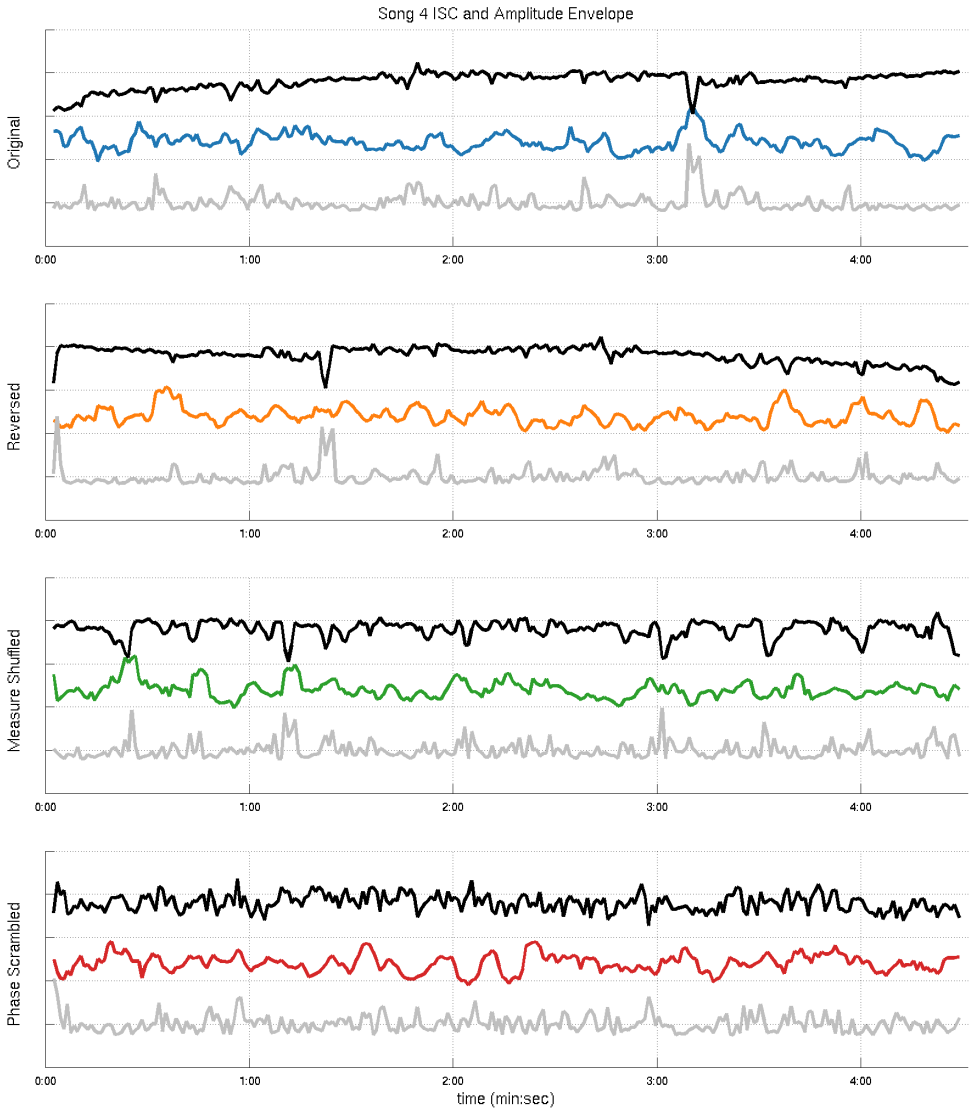

Figure 3.11: RC1 ISCs of Song 4 first-listen responses, plotted with stimulus amplitude envelopes. Each subplot shows the scale-free amplitude envelope (black), ISC time series (color), and rectified difference envelope (gray) for a given stimulus condition. As shown in Table 3.3, the original and measure-shuffled ISC time series for this song are significantly correlated with both the amplitude and rectified difference envelopes.

	Song	Listen 1 Env corr	Listen 1 Diff corr	Listen 2 Env corr	Listen 2 Diff corr
Orig	1	-0.1440**	-0.0596	-0.2026**	-0.0215
	2	-0.1853**	-0.0240	-0.1820**	0.0802
	3	-0.1804**	.1302	-0.1586**	-0.0232
	4	-0.2456**	0.3378**	0.0185	0.1775**
Rev	1	-0.1849**	-0.0020	0.0605	0.1421
	2	-0.1076	0.0304	0.0605	-0.1214
	3	-0.1000	-0.0149	-0.1137	-0.0182
	4	0.0105	0.0016	0.0433	0.1063
Meas	1	-0.0477	-0.0305	0.0566	0.0587
	2	-0.0804	-0.0121	-0.1369	0.1154
	3	-0.3553**	0.1422	-0.2642**	0.1209
	4	-0.1565**	0.2643**	-0.2600**	0.2925**
Phase	1	0.0787	-0.0113	-0.0665	-0.1076
	2	0.0521	-0.0117	-0.0375	-0.0096
	3	0.0241	0.0170	0.1828**	0.0828
	4	0.0686	-0.0112	-0.1367*	-0.0093

Table 3.3: ISC-amplitude envelope correlation information. Two asterisks '**' denotes statistical significance after correcting for FDR ($p < 0.05$); one asterisk '*' denotes marginal statistical significance after correcting for FDR ($0.05 \leq p < 0.1$).

acoustical features, then an original stimulus will produce an ISC time course that is roughly equivalent to the reversed ISC time course produced by the reversed version of that stimulus. However, as we saw in the previous analysis, the impact of amplitude envelope on ISC time course is to some extent context dependent (e.g., in Figure 3.11, ISCs at 3:10 of the original version differ from those at 1:20 in the reversed version). ISCs of original versions of the four songs are plotted over the flipped ISCs from the corresponding reversed version in Figure 3.12. As can be seen in the plot, there are some regions of agreement between the time series for a given stimulus. This result may point to regions of cortical synchrony that are driven by acoustical information in the moment, independent of the larger musical context. This is broadly similar to findings of Regev et al. (2013), who analyzed fMRI responses to forward and reversed silent films. The summary of correlation coefficients and p-values are shown in Table 3.4. Shared acoustical content, independent of order, produces statistically significant correlations for Song 1 and Song 2.

Figure 3.12: RC1 ISCs for original stimuli plotted with flipped ISCs for reversed stimuli. Correlation coefficients of the two time series for each song are reported in Table 3.4

Song	Correlation coefficient	p-value
1	0.3176	$< 10^{-16}$**
2	0.1688	0.0063**
3	0.1078	0.0833
4	0.0036	0.9537

Table 3.4: Correlation of original and flipped reversed ISCs. Two asterisks '**' denotes statistical significance after correcting for FDR ($p < 0.05$); one asterisk '*' denotes marginal statistical significance after correcting for FDR ($0.05 \leq p < 0.1$).

3.4 Discussion

In this study, we have validated the use of combined RCA and EEG-ISCs to study cortical responses to complete musical works. We presented participants with intact and temporally scrambled auditory stimuli derived from engaging yet novel musical works. Maximally correlated components were derived across the set of unique participant pairs for each stimulus and the projected response data were used in the computation of ISCs over time.

The RC topographies derived by Dmochowski et al. (2012) from responses to audiovisual film excerpts were fairly consistent across stimulus conditions, with RC1 likely reflecting visual processing and RC2 reflecting auditory processing. Our aggregate RC1 in response to musical excerpts is in agreement with component topographies derived by other means in previous studies of naturalistic music processing. This topography was found to be consistent across all stimulus conditions that retain musical features such as beat, meter, melody, and recognizable instruments, while the phase-scrambled RC1 topographies were anomalous by comparison and not consistent across stimulus exposures.

This division of stimulus conditions into broadly 'musical' and 'non-musical' categories is apparent in the behavioral and ISC results as well. Aggregate behavioral ratings were always lowest for phase-scrambled stimuli, as were the proportions of significant ISCs. Among the musical stimuli, however, there was some disparity between the behavioral and cortical measures, with original stimuli receiving highest behavioral ratings and measure-shuffled stimuli producing the greatest proportion of significant ISCs. The cortical results do not appear to be solely a result of following the amplitude envelope; whether this result reflects lower-level startle or orienting responses, or in fact reflects increased engagement (with attention and interest in future events) will be an interesting topic to study in more detail in the future. The measure-shuffled results also differ from the EEG-ISCs derived by Dmochowski et al. (2012) for an analogous stimulus manipulation (scrambling at the level of scenes), in that scene scrambling produced a significant *decrease* in proportion of significant ISCs. The present findings, taken together with the occasional observed synchrony of original and flipped reversed ISC time courses, may suggest that EEG-ISCs may be a useful approach to studying further the processing of music on multiple time scales. We acknowledge that the songs used for the present study, being highly repetitive with little tonal or timbral variation, are probably not ideal for exploring manipulation of listener expectations over longer time scales; in fact, several participants reported at the end of their

session that they did not realize the (unfamiliar) reversed and measure-shuffled stimuli had been manipulated at all.

The approach taken in this study was intended primarily to validate the use of RCA and ISCs to analyze responses to full-length, naturalistic music excerpts collected in single listens. Our stimulus manipulations and experimental design have provided preliminary insights into the level of temporal coherence in the stimulus needed to drive temporal reliability in the brain response. Future studies can build and improve upon the present approach to bring the focus more specifically on *engagement*. While we imposed signal-level manipulations on naturalistic excerpts, future studies could consider compositional manipulations that vary specific attributes of a musical excerpt (e.g., tonality, expressivity). As we have yet to disentangle the role of the amplitude envelope in driving synchronous responses to intact and scrambled excerpts, we propose a future control manipulation that phase-scrambles the stimulus while preserving the original amplitude envelope.

While we analyzed the ISC results here in relation to amplitude envelope and high-level structural segmentation, other stimulus attributes could also be explored. For example, while the songs were not in English, the mere presence of vocals (or a distinct instrumental melodic line) may impact the reliability of the audience response. Musical variety brought about by juxtaposition of phrase-level and thematic elements within a song part may also keep listeners engaged (and may in fact clarify some of the ISC peaks during verse sections in Song 3). Lyrics impart important information to many listeners but were intentionally obscured here; this attribute could form an important component for future studies.

Other interesting questions that could be addressed in future studies include effects of repetition and exposure. Does repetition of musical material within a song help to drive cortical synchrony? Does it matter whether repeated structural elements are distributed throughout a song, or occur one after the other? Will the degree of cortical synchrony over repeated listens reflect the inverted-U curve of musical preference over time (Hargreaves, 1984)? Finally, all of our EEG analyses were performed over all participant pairs; however, *intra*-subject correlations (IaSCs) can also be computed (Dmochowski et al., 2012). Therefore it may be interesting to use this approach to explore personal preference—for example, through responses to one's personal favorite music, or to study the relationship between subjective behavioral ratings and degree of cortical synchrony with other audience members.

Chapter 4

Physiological and Behavioral Measures

In the previous chapter we examined the impact of structural coherence of music on the temporal reliability of EEG responses across an audience of listeners. How might these initial findings be informed by other types of continuous responses? To gain some insight into this matter, we now review selected literature on physiology and behavior. Physiological responses have been used in music cognition research to study the arousal—and often by extension, emotional (Rickard, 2004)—aspects of music processing. Continuous behavioral responses have been used to investigate a variety of dimensions along which music can be characterized, including arousal, valence, familiarity, and, importantly, engagement. In examining the main findings from these studies, we will consider how these responses might be combined with the EEG methodology introduced in Experiment 1 to provide more insight into characterizing musical engagement.

4.1 Physiological Responses

Physiological responses provide a useful measure in music perception and cognition research. These responses are fairly inexpensive and easy to collect; and, like brain responses, they are for the most part beyond the conscious control of the listener, and may thus be considered objective. Compared to brain responses, physiological responses are generally easier and more inexpensive to collect; the response features of interest are also usually immediately

discernible from visual inspection of the raw data, even prior to data cleaning and preprocessing. The apparatus for collecting physiological responses has also been described as less cumbersome and potentially less distracting than those used for measuring brain responses (Bracken et al., 2014).

The responses most commonly used in music research include heart rate (HR), computed from inter-beat intervals of the ECG; respiratory responses, which may be collected by belts worn around the chest and abdomen and from which breathing rate and amplitude may be derived; and galvanic skin response (GSR), which typically refers to the transient, phasic response over small time windows (also known as the skin conductance response (SCR) or electrodermal activity (EDA)) but can refer also to tonic levels over a long duration (also known as skin conductance level (SCL)) (Khalfa et al., 2002). Other, less frequently used measures include skin or finger temperature (Krumhansl, 1997; Lundqvist et al., 2009; Salimpoor et al., 2009; Tsai et al., 2014), electromyography (EMG)—for instance, to measure facial muscle activity related to smiling or frowning (Grewe et al., 2007a; Egermann et al., 2013; Russo et al., 2013)—and cortisol levels (Rickard, 2004). The experience of musical 'chills' generally does not refer to a specific physiological response, and physiological correlates are often sought among a suite of responses (see, for example, Grewe et al. (2007b, 2010)). Self-reports rating some attribute of the stimulus, retrospectively or in real time, are often collected along as well, for comparison with the physiological responses.

The analysis of physiological responses to music dates back several decades. Acknowledging the need for objectivity in measuring effects of music on a listener, Phares (1934) collected both GSR and behavioral ratings of mood, enjoyment, and attention. While this study did demonstrate a relationship between skin conductance and affective intensity, the author deemed the response 'of little value' in the study of musical appreciation. Even so, other researchers have subsequently continued this line of research.

Many music physiology studies address, directly or implicitly, the question of whether music *evokes* emotional responses in listeners (the 'emotivist' view), or whether it merely *represents* emotions, which listeners may recognize but not experience themselves (the 'cognitivist' view) (Krumhansl, 1997; Rickard, 2004; Lundqvist et al., 2009). In a fundamental physiology study, Krumhansl (1997) focuses specifically on this question, analyzing a suite of 12 physiological responses to musical excerpts chosen for the emotional states they represent (sad, fear, happy). Her results did show an effect of musical emotion on listeners'

physiological responses, lending support to the emotivist view. In fact, of the studies reviewed here, only Grewe et al. (2007a) interpreted their results as explicitly supporting the cognitivist view, and in that case it was unclear whether the conclusion was due to a positive result in that direction, or to inconclusive physiological results.

4.1.1 Experimental Approaches

Stimulus Selection

While most music cognition research relies upon stimuli that are selected or composed by the experimenter, varied approaches to stimulus selection can be found in the music physiology literature. One reason for this may have to do with the fact that physiological responses, often parameterized as a combination of arousal and emotion factors, cannot be evoked by just any stimulus. As a result, stimuli used in these studies are often naturalistic music excerpts that were pre-selected by the experimenters according to their representation, and possible elicitation, of an arousal or emotional characteristic of interest. Phares (1934) selected eight excerpts for each of the four moods of gay, melancholy, triumphant, and tragic. Another early study by Zimny and Weidenfeller (1963) focused more on arousal characteristics of the stimuli, selecting three pieces which they deemed to be exciting, neutral, or calming. The aforementioned set of emotions analyzed by Krumhansl (1997) was expanded by Khalfa et al. (2002) to include a category of excerpts characterized by peacefulness. A more recent study by Grewe et al. (2007a) focused more on incorporating a broad range of musical *styles*, and thus selected music from genres ranging from classical to pop to death metal. Experimenter-selected stimuli in physiology studies involving music need not be limited to music; for example, Gomez and Danuser (2004) used short excerpts of both music and everyday noises that spanned the valence and arousal continua. Grewe et al. (2010) later expanded the scope of stimulus modalities further, including, along with music, stimuli that they hypothesized would stimulate participants in visual, tactile, and gustatory domains.

Some researchers have acknowledged that the likelihood of arousing physiological responses from experimental participants might be higher if participants selected their own stimuli. This approach was taken by Rickard (2004); here, participants were instructed to 'choose a piece of music that is emotionally powerful or moving, and personally meaningful to you,' with the reasoning that such self-selected excerpts would reliably arouse the

person who chose them. As these analyses were not time-resolved, the experimenter could simply compare a preselected time window of each response to a baseline to assess changes in physiological response levels, presumably to the arousal characteristics of the stimuli. A variation of this approach was used by Grewe et al. (2007b), who combined a stimulus set from another of their studies (Grewe et al., 2007a) with 5–10 pieces selected by each participant on the basis that the self-selected excerpts would arouse strong emotions. The authors identified the excerpts that most effectively induced chills in this study, and used them as experimenter-selected stimuli in a subsequent experiment (Grewe et al., 2010).

Stimulus Manipulations

For the most part, music physiology studies have used naturalistic music stimuli in their original form and—unlike most music perception and cognition studies, which assess responses to experimental stimuli in comparison to responses to control stimuli—focus more on the extent to which responses to experimental stimuli differ from a baseline measure. However, a few studies have examined physiological responses to manipulated stimuli. The topics of interest here concern harmonic unexpectedness and dissonance, and whether these properties induce physiological manifestations of emotion. Steinbeis et al. (2006) used excerpts from Bach chorales with unexpected harmonic events in their original forms, and also in manipulated versions that increased or decreased the harmonic unexpectancy at the same points in the music. A later study used original (consonant, pleasant) versions of instrumental dance tunes in conjunction with control (dissonant, unpleasant) conditions that added pitch-shifted versions of the piece to the original, at dissonant intervals (Sammler et al., 2007). In a more recent study, taking a similar approach to Steinbeis et al. (2006), Koelsch et al. (2008) selected excerpts from classical piano sonatas, which contained irregular chords in their original compositions. The authors created alternate versions of the excerpts whereby the irregular chords were 'corrected' and also made more irregular. An additional control condition eliminated expressive variations in tempo.

It is interesting to note that the studies employing stimulus manipulations were all analyzing physiological responses in conjunction with EEG responses. As the EEG analyses were fairly conventional (averaging-based ERP and spectral power analyses), it makes sense that the authors chose stimuli that were amenable to their planned EEG analyses. We note also, however, that the manipulated stimuli were all derived from pre-existing compositions or musical recordings, and thus these stimulus sets may constitute a compromise between

the fully controlled and synthesized stimuli often used in music EEG experiments, and the fully naturalistic excerpts used in other music physiology studies.

Coding of Stimulus Features

Over the course of the research, characterization of stimulus features has grown more nuanced and has incorporated computational approaches. While the earlier studies reviewed here focused on musical excerpts deemed by humans to express, overall, a particular mood (Phares, 1934; Krumhansl, 1997; Khalfa et al., 2002), arousal level (Zimny and Weidenfeller, 1963), or style (Grewe et al., 2007a), later approaches came to include more fine-grained characterizations of the stimuli. For example, Gomez and Danuser (2007) had three musical experts characterize a stimulus set using such musical features as tempo, rhythm, and rhythmic articulation. Computational analyses of stimulus features were introduced by Grewe et al. (2007b), who extracted psychoacoustic features (loudness, sharpness, roughness, and fluctuation) from a set of 190 excerpts. Recently, Egermann et al. (2013) characterized their stimuli using the information dynamics of music model (IDyOM), a computational model of auditory expectation, and compared the model output over time to subjective ratings and physiological responses.

Baseline Recordings

A practical consideration in physiology experiments is the pre-stimulus baseline. Group-level analyses of physiological responses are more effective when individual differences in resting-state arousal are controlled for. Thus, it is fairly common practice to subtract, from the responses to experimental stimuli, a mean or median baseline measure (one study scales the experimental responses by the baseline to examine percent change from baseline measure (Rickard, 2004)). Including a baseline period prior to every stimulus can further serve to correct physiological changes over the course of an experimental session (Krumhansl, 1997).

Reported baseline periods ranged in duration from 5 seconds (Lundqvist et al., 2009) to 5 minutes (Salimpoor et al., 2009). Most reported baseline periods ranged in duration from 15–90 seconds (Krumhansl, 1997; Gomez and Danuser, 2004, 2007; Grewe et al., 2007a,b; Sammler et al., 2007). Some studies used only latter portions of a longer baseline: Egermann et al. (2013) used 40 of 45 seconds of baseline preceding a stimulus, while Russo et al. (2013) used the final 20 seconds of a 30-second baseline. Baseline durations can also be tailored

to specific physiological responses; for example Grewe et al. (2010) used only a 10-second baseline for heart rate, but used a full 1-minute baseline for the slower respiratory responses.

4.1.2 Analysis Approaches

Temporal Resolution of Analyses

The literature contains various approaches to the temporal resolution of the analyzed responses. Some studies compute an average measure of physiological responses over the stimulus or some portion thereof, which are compared to the baseline measures. For example, Khalfa et al. (2002) averages across 7-sec stimuli; Gomez and Danuser (2004) and Gomez and Danuser (2007) use the final 15 seconds of their stimulus, while Rickard (2004) uses the middle portion of the stimuli, whose regions of interest range from 2–5 minutes in length. One analysis in the Sammler et al. (2007) study compares the first and second halves of the response, while Russo et al. (2013) takes the mean of standardized values across a 30-second trial. Tsai et al. (2014) take a more inventive approach to defining the temporal region of interest, focusing on a temporal window around the entrances of first and second choruses of various pop songs—responses are pooled across all subjects and occurrences of song parts and compared to baseline measures.

Other studies analyze physiological responses as they evolve over time and collections of time-sampled responses. Zimny and Weidenfeller (1963) analyzed the trajectory of mean responses computed over minutes of participants' responses. Krumhansl (1997) and Lundqvist et al. (2009) use greater temporal resolution, analyzing responses over 1-second and 5-second blocks, respectively. Egermann et al. (2013) present some of the most temporally resolved results, with lowpass-filtered responses sampled at 256 Hz over the duration of full pieces ranging from 38 seconds to 3.5 minutes in length. A few studies have taken a hybrid approach to temporal resolution, varying the window length for averaging depending on the response of interest (Steinbeis et al., 2006; Grewe et al., 2007a; Koelsch et al., 2008).

As can be inferred from this review thus far, for the most part physiological responses are analyzed in reference to a full stimulus or to some pre-selected portion of the stimulus. Some researchers have exercised some flexibility in this approach, for example by letting participants select their own stimuli (Rickard, 2004) or by pooling responses across repeated instances of a song part of interest (Tsai et al., 2014). A contrasting approach, focused more on characterizing the response than seeking to evoke it reliably, was taken by Grewe et al.

(2009). In this study, participants were instructed to indicate whenever they felt a chill, for as long as the chill lasted. The researchers then focused their analyses on the physiological activity during these reported chill events, regardless of the corresponding musical content, to characterize the physiological indices of the chill response itself. The researchers also found that individual listeners experienced chills at different points in time, which may point to the usefulness of assessing physiological responses to music on an individual level.

Descriptive Versus Predictive Analyses

In general, music physiology researchers have sought to characterize listeners' responses to different types of music—this can be considered a *descriptive* approach. A few studies have taken a *predictive* approach, in which the researchers attempt to predict stimulus characteristics from listeners' physiological responses. Kim and André (2008), interested more generally in predicting emotional states underlying physiology and finding music to be a suitable means of evoking emotions, classified participants' physiological responses in a four-class problem using the quadrant-based arousal-valence model. More recently, Russo et al. (2013) used linear regression and neural networks to predict, from physiological responses, listeners' experienced valence and arousal to musical excerpts. Finally, Shin et al. (2014) proposed a stress-relieving music recommendation system around the finding that the sympathovagal balance index (SVI), a measure derived from heart-rate variability, could predict participants' musical preferences.

4.1.3 Summary of Current Findings

In summary, the music physiology literature spans a variety of response modalities, stimuli, experimental procedures, and analysis techniques. While the general consensus from this literature is that there does exist a relationship between emotional content of the stimuli and physiological responses of the listener (supporting the emotivist view), the degree of consensus across studies varies from response to response.

Findings appear to be in highest agreement for GSR, sometimes analyzed in conjunction with the 'chill' response. This response is generally reported to increase along with arousal (Gomez and Danuser, 2004; Rickard, 2004), excitement (Zimny and Weidenfeller, 1963), in conjunction with reported chills (Grewe et al., 2007b, 2010), in response to faster tempos and more staccato musical events (Gomez and Danuser, 2007), over predefined regions of interest in pop songs (Tsai et al., 2014), and for musical expectation violation (Steinbeis

et al., 2006; Koelsch et al., 2008; Egermann et al., 2013). Increased GSR has also been observed along with happiness or pleasure (Lundqvist et al., 2009; Salimpoor et al., 2009), though there is less consensus on this point: Khalfa et al. (2002) found GSR to be higher for fearful or happy excerpts over sad or peaceful excerpts, while Phares (1934) found GSR to increase according to intensity of emotion regardless of reported *quality* of the emotion, and Russo et al. (2013) and Krumhansl (1997) report a negative relation between valence and skin conductance *level*, though these last two measures were computed over a 30-second interval and averaged over 1-second time intervals, respectively. Finally, Grewe et al. (2007a) found no significant effect of emotion, while Russo et al. (2013) found no impact of arousal on reported skin conductance levels. Chills were found to increase in frequency in response to more arousing stimuli (Rickard, 2004) and show a high correlation with perceived pleasantness (Grewe et al., 2007b).

Respiratory rate findings are also fairly consistent. Krumhansl (1997) reported increased respiratory rate for all musical excerpts, with a most pronounced effect for fearful and happy stimuli. Gomez and Danuser (2004) found respiratory rate to increase for positive valence and heightened arousal, as did Russo et al. (2013). Gomez and Danuser (2007) report an increase in respiratory rate for stimuli with faster tempos and more staccato articulations. However, Egermann et al. (2013) found no correlation between respiratory rate and the computational assessment of musical expectation violation, though they did find that respiratory rate decreased concurrently with participant-reported expectedness, and increased with reported unexpectedness of the musical stimuli. Grewe et al. (2010) report no effect of chill experiences on respiratory rate.

The literature has provided some insights into the impact of arousal and emotion in music on heart rate. Krumhansl (1997) reports a decrease in heart rate for all stimuli, especially sad ones; conversely, Salimpoor et al. (2009) reports an increased heart rate during pleasurable listening experiences. Sammler et al. (2007) and Egermann et al. (2013) report a decrease in heart rate for dissonant excerpts and unexpected musical events, respectively; Russo et al. (2013) and Rickard (2004) report increased heart rate for more arousing stimuli, though in the case of Rickard, this effect was not significant. However, beyond this, a number of studies report no relation between heart rate and mood/valence (Zimny and Weidenfeller, 1963; Lundqvist et al., 2009; Russo et al., 2013), arousal (Gomez and Danuser, 2007), valence and arousal (Gomez and Danuser, 2004), expectation violation (Steinbeis et al., 2006; Koelsch et al., 2008), or expressiveness (Koelsch et al., 2008) of music. In addition,

Grewe et al. (2010) found no relation between heart rate and chills experienced in response to music (though they did for other stimulus modalities).

Finally, stimulus characteristics not related to an emotional correlate of arousal may influence physiological responses. Respiratory activity, for example, may entrain to a musical rhythm or tempo—a finding reported by Haas et al. (1986) and exploited in a clinical setting by Cui et al. (2010).

4.1.4 Reliability of Physiological Responses

Music studies have assessed responses to stimuli using response measures such as mean or median deviation from baseline. However, we could find no music studies that assessed the temporal *reliability* of physiological responses—that is, the metric of interest applied to EEG responses in the previous chapter using ISCs. However, this approach has been attempted in studies using non-musical stimuli. For example, Bracken et al. (2014) computed ISCs of heart rate and skin conductance levels. Motivated directly by the approaches of Hasson et al. (2004) and Dmochowski et al. (2012), the researchers recorded physiological responses while 163 experimental participants watched a 100-second donation solicitation video of a father telling the story of his 2-year-old son who is dying of brain cancer. Each participant had the opportunity, at the end of the experimental session, to donate some or all of his participant payment to St. Jude's Children's Research Hospital. ISCs were computed over 5-second windows. HR remained synchronized for a longer period of the video for donors than for non-donors. However, SCL was correlated for more time windows for non-donors than for donors. A more recent study, though not using ISCs, appears to have used the same physiological responses to predict donation behavior from HR, SCL, HR variability, and hormonal levels (Barraza et al., 2015). Here, heart responses and SCL significantly predicted the decision to donate.

4.2 Continuous Behavioral Responses

The use of physiological responses in music research is motivated largely by the objectivity of those responses, and by the ability to analyze them in a time-resolved fashion. Continuous behavioral responses offer complementary advantages: They can also be analyzed in conjunction with time-resolved musical events. The more important implication of this feature is that the behavioral responses are delivered in real time, as the stimulus plays,

and not retrospectively at the conclusion of the stimulus (Gregory, 1989). Additionally, as a listener's assessment presumably varies over the course of a musical excerpt, a continuous behavioral response will capture these variations over time and allow the results to be compared across experimental participants (Madsen et al., 1993).

4.2.1 Response Collection Interfaces

While today it is trivial to devise an interface to collect and store continuous behavioral responses—for example, using a mouse or joystick connected to a computer, or using the screen of a mobile device—collection of these responses originally required the fabrication of a custom interface. One of the first continuous-response systems for music research was the Continuous Response Digital Interface (CRDI) developed by Gregory (1989). This was a mechanical device, in which voltages from a potentiometer were digitized and sent to a data-acquisition computer. The CRDI was introduced in two models, a horizontal slider and a dial interface—both of which were validated in music experiments. The reliability of the CRDI was confirmed in a later study (Gregory, 1995). Gregory (1989) emphasizes the need for a response collection interface that is unobtrusive, cost effective, and easy to use (not requiring specialized skill or dexterity on the part of the experimental participant). The CRDI was next used by Madsen and Geringer (1990) to assess whether musicians attend to different musical features than do nonmusicians.

It is interesting to note that in these early applications, the slider interface was actually used to collect categorical responses—that is, which feature or dimension of music, such as dynamics, rhythm, or melody, the participant was currently focused on—rather than a response that varied along a continuum. However, Madsen et al. (1993) later used the CRDI to measure the degree of aesthetic experience over time, in response to an opera excerpt. These results indicated high agreement among participants in timing of critical drops and peaks in reported aesthetic response across the excerpt, showing promise in the idea of attaining an empirical measure of 'aesthetic experience', even when a definition of the term was deliberately withheld from experimental participants.

4.2.2 Dimensions of Self-Report

The fact that ratings of something as abstract as aesthetic experience can be generalized across a participant population points to the possibilities of what types of musical features

may be self-reported with continuous behavioral responses. While the physiology studies focused primarily on arousal, valence, mood, and expectation, here it is possible to ask the participants to report on specific assessments of the music. For example, Krumhansl (1996) analyzed ratings of musical tension over time in conjunction with participant-identified segmentation boundaries in the music, and uncovered a relationship between slowing of tempo, judgments of section boundaries, and tension ratings—specifically, tension peaks occurred at ends of structural segments. Tension was also shown to covary with melodic pitch height and density of notes. McAdams et al. (2004) later conducted a large-scale experiment with over 200 participants in a live concert setting with a novel musical piece and collected continuous ratings of familiarity (that is, recognition of repeated content from the same work) or emotional force from a given participant. The authors found here a relationship between familiarity and structural elements of the piece, and a fairly consistent global contour of emotional force over the course of the piece.

The studies reported so far all operated over one-dimensional responses (while two dimensions are reported in McAdams et al. (2004), only one was assigned to any given participant). This was a matter of necessity for the early CRDI experiments—in fact, studies around the time of its introduction would use two separate dials (one for each hand) to collect responses along two dimensions simultaneously (Gregory, 1995; Madsen, 1998). However, the CRDI eventually evolved to a two-dimensional model, which permitted a less complicated task for the participant. Madsen (1998) used this interface—a mouse interface connected to a television monitor—to collect ratings of Haydn's Symphony no. 104; participants reported arousal and valence ratings, here defined as the ranges of 'exciting' to 'relaxing' and 'ugly' to 'beautiful', respectively. The researchers found a correlation of -0.58 between the dimensions, and also found a strong correlation between the arousal ratings and ratings of tension collected in a separate study. In a later exploratory study, Schubert (2004) used a different two-dimensional response interface to collect responses of arousal and valence (here labeled with facial expressions) to four musical excerpts which the authors presumed would occupy various quadrants of the arousal-valence space. They then used a regression approach to determine how much of variance of time-differenced arousal and valence responses was explained by temporally resolved musical features of each piece.

4.2.3 Reliability of Continuous Behavioral Responses

The studies discussed thus far have looked at the activity of continuous behavioral responses over time. This is a similar approach to that taken in the music physiology studies, although here we consider position over a range of responses, rather than deviations from a baseline measure. However, recent studies have also looked specifically at the reliability of participant responses. An early application of this approach was taken by Krumhansl (1996), who computed ISCs of tension ratings over time, although these were not time resolved. An approach closer to our approach with time-resolved EEG-ISCs is taken in two recent studies, which also happen to be interested specifically in engagement.

The first, by Schubert et al. (2013), is actually a study of audience engagement with a live dance performance. In this study, however, the authors make the important distinction between quantifying the *amount* of engagement and the *agreement* in the engagement response. The authors use an analysis technique based on the standard deviation of responses, across participants, over the course of the piece, and in fact discover that periods of high engagement differ from periods of good agreement regarding engagement. They differentiate 'gem moments'—periods where engagement rises suddenly, often in response to a surprising event—from moments of high agreement, which they surmise might occur more during periods where expectations have been established and are not interrupted.

In a subsequent study, Olsen et al. (2014) collected continuous self-reports of engagement to a set of classical and electroacoustic musical excerpts and attempted to use reported levels of engagement, alone and alongside time-varying acoustical intensity and spectral flatness, to predict time-varying ratings of arousal and valence from a previous study. The findings suggest that engagement may mediate the relationship between an excerpt's acoustical features and subjective ratings of arousal and valence.

4.2.4 Experimental and Analytical Approaches

As the literature on continuous behavioral responses grows, so too does the sophistication of data analysis. While insights from earlier studies were made largely from inspection of the data (Madsen et al., 1993) or correlations of responses (Krumhansl, 1996; Madsen, 1998), more recent analyses are motivated by time-series approaches (Schubert, 2004; Olsen et al., 2014). For multiple comparisons such as are made across a collection of time samples, an awareness of correcting the significance measures is appropriate—this is mentioned by

McAdams et al. (2004), who use a lower p-value threshold and acknowledge the need for more careful consideration of this matter.

4.3 Discussion

There exists a rich literature on the physiology of music, and a growing literature using continuous self-reports. The physiology studies encompass a broad range of responses, stimuli, and analysis techniques, while the behavioral studies stand out in terms of the variety of dimensions that can be reported upon. Analysis approaches to these data range from anecdotal to statistical, and responses are interpreted against stimuli from broad categorizations such as mood to fine-grained variations in acoustical features and dynamic computational models of musical expectation. It should be pointed out, too, that several of the physiology studies utilized continuous behavioral responses, against which the physiological responses were compared.

There are still several avenues of research to be explored for both of these modalities. In particular, while self-reports of engagement show promise, both in terms of differentiating activity from agreement, and in relating engagement to valence and arousal (popular dimensions of interest for physiology studies), there exist no conclusive findings on the physiology of *engagement*. Therefore, the study of engagement could well benefit from a combined analysis of physiological and continuous behavioral responses.

4.3.1 Considerations

There are several points that must be considered regarding these response modalities. First, there remains the issue of group-level versus individual analyses. Various physiology studies have benefited from diverging from traditional experimental designs, showing that participant-selected stimuli (Rickard, 2004; Grewe et al., 2007b) and response-locked rather than stimulus-locked analyses (Grewe et al., 2009) may be appropriate for studying these types of responses. In fact, Salimpoor et al. (2009) found that physiological responses to music were not observed in participants who reported no pleasure response to the musical content. Individual differences will arise in continuous behavioral responses as well. For example, Madsen et al. (1993) reported experience effects in their participant ratings of aesthetic experiences; some participants would report peak experiences only in relation to their own instrument or vocal range. These events would likely not generalize across the

participant population, but still present meaningful findings.

For both types of responses, there arises the problem of multiple comparisons when responses are analyzed across time samples. This issue is not always addressed in the literature. In the previous chapter, we assessed statistical significance of the EEG-ISCs using a permutation test, and corrected other p-values using FDR. A similar procedure will need to be implemented in the analysis of the present proposed results.

There are other practical considerations to keep in mind with these types of responses. Physiological responses will susceptible to orienting behavior around the time of stimulus onset (Grewe et al., 2007a; Sammler et al., 2007; Koelsch et al., 2008; Lundqvist et al., 2009). Continuous behavioral responses additionally suffer from reliability issues, both at stimulus onset (Schubert, 2013) and as an 'afterglow' effect after a stimulus ends (Schubert, 2013) or following periods of peak experience in the music (Madsen et al., 1993).

Furthermore, EEG responses typically occur within 500 msec of corresponding stimulus events—for example, Schaefer et al. (2011) report a maximal EEG-to-amplitude envelope correlation at a time lag of 100 msec, while Sturm et al. (2015) incorporate stimulus-to-EEG response lags of up to 300 msec in their analysis. EEG responses may thus be treated as effectively instantaneous responses, especially in the case of ISC analyses which aggregate the data over temporal windows lasting many seconds. For physiological and behavioral responses, however, there exist lags of varying lengths between musical events and corresponding responses. For example, a delay of 1–5 seconds between musical events and psychological or physiological reactions is proposed (Schubert and Dunsmuir (1999), reported in Grewe et al. (2007a)). In examining physiological correlates of reported chills, Grewe et al. (2009) found that an increase in skin conductance preceded the report of a chill by around 2 seconds, while heart rate increased after chill onset; both responses peaked 4–5 seconds after the reported chill onset. For non-musical narratives, Bracken et al. (2014) report an estimated delay of 5 seconds between stimulus events and corresponding cardiac responses.

In terms of behavioral responses, Sammler et al. (2007) report an estimated delay time of no more than 3 seconds, while Egermann et al. (2013) assume a lag of 2–3 seconds. Krumhansl (1996) found that shifting the continuous ratings of tension 2–3 beats earlier produced a good alignment with theoretical predictions of tension, though she posits that this could be tied to time course of repetition of musical content. Lags have also been found to vary over the course of a stimulus: Schubert (2004) report a 1–3 second lag in general,

but also shorter lags in arousal ratings of 0–1 seconds after sudden changes in loudness.

It is also important to point out once more that while physiological responses may be considered objective, continuous behavioral responses are not. Therefore, while self-reports are more direct, and—as we have seen—enable more specific aspects of the musical experience to be probed, these types of response may be affected by a desire to deliver the 'correct' response (Lundqvist et al., 2009), or by the participant's ability to accurately assess his or her experience (Madsen et al., 1993). The act of responding could affect the response; McAdams et al. (2004) actually received several spontaneous reports from experimental participants that the task of delivering a continuous behavioral response enriched their listening experience because they were more focused. However, the opposite result could also occur, for example from having to consciously reflect on the listening experience (Russo et al., 2013).

Despite the potential complications and shortcomings of these responses, physiology and behavior will provide potentially useful complementary data to elucidate cortical findings. In the next chapter, we collect these responses in tandem with EEG in a combined analysis focused on responses to musically salient events in a complete naturalistic excerpt from the classical genre.

Chapter 5

Experiment 2

5.1 Introduction

In Experiment 1, we provided a first validation of the EEG-ISC method for analyzing responses to intact and scrambled Hindi pop songs. Stimuli that preserved short- and long-term temporal structural coherence of the music produced consistent RC1 topographies and higher ISCs than did the phase-scrambled control.

As a further validation of using EEG-ISCs to study musical engagement, we now broaden the scope of our responses to include measures derived from continuous physiological and behavioral responses. In this second experiment, we record dense-array EEG, electrocardiogram (ECG), and respiratory inductive plesthmyography (chest and abdomen respiratory activity) concurrently while participants hear original and reversed conditions of a classical music excerpt. In a separate experimental block, participants deliver continuous behavioral measures of engagement with both stimuli. We compute the synchrony (ISCs) of the cortical responses, along with both the level (deviation from baseline) and synchrony of the physiological and continuous behavioral responses. This combined set of responses allows a first look at interpreting EEG-ISC results among other measures of engagement (continuous behavioral responses) as well as arousal measures (physiological responses), and also among other objective (physiological) as well as subjective (continuous behavioral) responses delivered in time with the stimulus as it plays. Selected results of this study are reported in Kaneshiro et al. (2016b).

5.2 Methods

5.2.1 Ethics Statement

As with the previous experiment, this experiment was approved by Stanford University's Institutional Review Board as part of the study IRB-28863: Studies of Musical Learning and Expectation Using Behavioral, Physiological, and Scalp-Recorded EEG Responses. All participants delivered written informed consent prior to their involvement in the experiment.

5.2.2 Stimuli

Song Selection and Stimulus Manipulation

Stimuli for this experiment were derived from the first movement of Edward Elgar's Cello Concerto in E Minor, Op. 85 (1919). Selected for the large fluctuations in arousal that occur across the movement, this piece has also been shown to induce frisson ('chills') in past experiments (Grewe et al., 2007b, 2010). The version used here is the recording of the 1965 Jacqueline du Pré performance with Sir John Barbirolli and the London Symphony Orchestra, considered to be a definitive and influential performance of the piece (Solomon, 2009).

We purchased a digital version of the EMI Records, Ltd. digitally remastered version of the recording from iTunes.[1] The .m4a stereo recording was remixed to mono and exported to .wav format using Audacity recording and editing software, version 2.0.3.[2]. Subsequent stimulus processing steps were performed in Matlab. First, the mono audio file was loaded and a linear fade-in and fade-out was applied to the first and last 1000 msec of the recording, respectively. We then created a second, reversed version of the ramped mono stimulus. As in the previous experiment, a second audio channel was added to each mono stimulus to deliver intermittent timing clicks for correcting the stimulus onset time during data preprocessing. The resulting stereo signals were written out to .wav files with a sampling rate of 44.1 kHz.

Musical Events of Interest

The selected excerpt is around eight minutes long and has an ABA' structure, with minor A sections and a primarily major B section. The sections are linked through rhythmic and

[1] `https://itunes.apple.com/us/album/elgar-cello-concerto-in-e/id693718997`

[2] Audacity® is copyright ©1999–2016 Audacity Team, `http://audacity.sourceforge.net/`

melodic variation of the primary motivic element of the movement. The A and A' sections each include climactic events characterized by sharp melodic and textural/dynamic rises, which culminate in cadences in E minor.

In descriptive music analysis, salient events are often designated as structurally relevant. These events can include climactic 'highpoints' (Agawu, 1984); the introduction and reprise of principal thematic materials; a sudden, generally unexpected pause; and significant changes of timbre or texture, such as a change in orchestration or the entrance of a soloist. We identify a set of salient events in the original (forward), which we conjecture will produce reliable responses in our population of listeners; event onsets are annotated in Figure 5.1. First is the first solo entrance of the main theme in the cello in the A and A' sections. In particular, we conjecture the re-entrance of this theme (A2) may serve to draw in the listener, as it denotes a return to section A'. Next, we highlight regions of rapidly rising tension that culminate in the A and A' highpoints (B1 to C1; B2 to C2). Finally, as a contrasting event of interest due to its lack of activity and trajectory, we demarcate point D, which marks a break in activity between sections A and B of the movement. We surmise that stimulus reversal, which disrupts the contextual salience of these events (e.g., highpoints at C1, C2 will now drop rapidly in intensity to B1, B2; A1 and A2 will denote the exit, rather than entrance, of the solo instrument), will impact the effectiveness of these events in driving reliable responses.

5.2.3 Participants

For this experiment we sought healthy, right-handed participants between 18–35 years old with normal hearing, who were fluent in English and had no cognitive or decisional impairments. As formal musical training has been shown to enhance cortical responses to music (Pantev et al., 1998), we sought participants with at least five years of formal training in classical music, which could include private lessons, AP or college-level music theory courses, and composition lessons. Years of training did not need to be continuous. Because we wished to avoid involuntary motor activations found to occur in response to hearing music played by one's own instrument (Haueisen and Knösche, 2001), we recruited participants who had no training or experience with the cello. Finally, since we are studying engagement, we sought experimental participants who enjoyed listening to classical music, at least occasionally.

We collected usable data from 13 participants (four males, nine females) ranging from

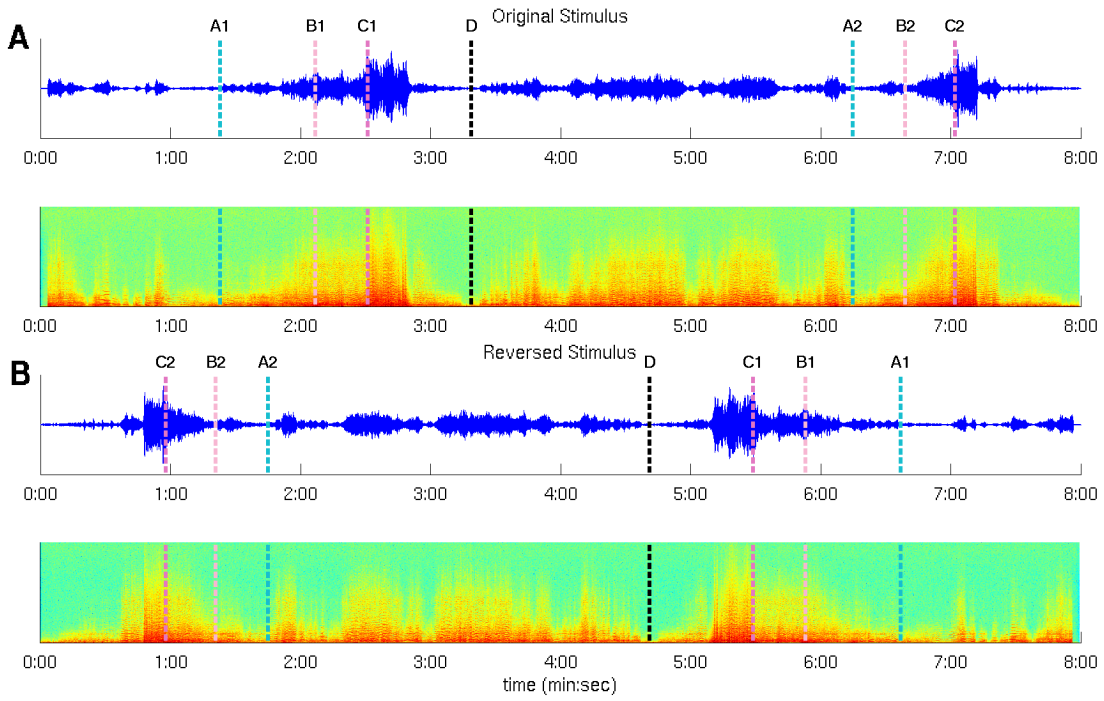

Figure 5.1: Elgar stimulus waveform and spectrogram. Stimuli were derived from the first movement of Elgar's Cello Concerto in E Minor, Op. 85. Panel A: We identify a set of musical events of interest in the original version. A1 and A2 mark the entrance of the main theme in the solo cello. B1 and B2 commence periods of increasing tension, which culminate in highpoints at C1 and C2. Point D designates the end of the first section of the movement—a period of low activity in the excerpt. Panel B: Reversed stimulus. Musically salient events are acoustically intact, but contextually disrupted.

18–34 years of age (mean = 23.08 years). All participants met the eligibility requirements for the experiment. Years of formal musical training ranged from 7–17 years (mean = 11.65 years). Eight participants were currently involved in musical activities at the time of their experimental sessions. Music listening ranged from 3–35 hours per week (mean = 13.96 hours). Data from three additional participants who took part in the experiment were excluded during preprocessing due to gross noise artifacts (§5.2.5).

5.2.4 Experimental Paradigm and Data Acquisition

Experimental Overview

After a participant delivered written informed consent for the experiment, he or she filled out a demographic and musical experience questionnaire. The experiment was structured in two blocks. The first block involved the simultaneous EEG and physiological recordings. The participant was given an interactive overview of this block before net and sensor application. In this block, the participant heard each of the two stimuli once in random order. Each stimulus was preceded by a one-minute baseline, during which time low-amplitude pink noise was presented; the physiological responses collected over this interval are used to establish individualized baseline levels for each participant. The participant was instructed to sit still, avoid any movement, and view a fixation image presented on the monitor 57 cm in front of him while any auditory stimuli were being presented. At the conclusion of each stimulus, the participant delivered ratings, using number keys on a keyboard, for the following questions using a nine-point Likert scale (adapted from the previous experiment):

1. How pleasant was this excerpt, on a scale of 1 (not pleasant at all) to 9 (very pleasant)?
2. How arousing was this excerpt, on a scale of 1 (not arousing at all) to 9 (very arousing)?
3. How much of the excerpt was interesting, on a scale of 1 (none of it) to 9 (all of it)?
4. How predictable was this excerpt, on a scale of 1 (not predictable at all) to 9 (very predictable)?
5. How familiar were you with this excerpt, on a scale of 1 (not familiar at all) to 9 (very familiar)?

A sixth question was asked only after the presentation of the original (forward) stimulus:

6. How often do you listen to this genre of music, on a scale of 1 (never) to 9 (all the time)?

Each baseline and stimulus presentation was preceded by a break, the length of which the participant controlled via key press.

Following the EEG/physiology block, the participant was escorted out of the booth, and the electrode net and physiological sensors were removed. Once the participant was ready to proceed with the second experimental block, the experimenter presented an overview of the continuous behavioral response interface. Here, the participant was instructed to use the mouse to control a horizontal slider to indicate his level of engagement with each stimulus as it played. In this block, the forward and reversed stimuli were again presented

once in random order.

We used the definition of engagement drawn from past studies using continuous behavioral responses to measure audience engagement (Schubert et al., 2013; Olsen et al., 2014): 'DEFINITION: Engagement—being compelled, drawn in, connected to what is happening, and interested in what will happen next'. This definition, along with participant instructions 'YOUR TASK: You will continuously rate your level of engagement with an excerpt as it plays.' were presented on-screen to the participant prior to the presentation of each stimulus.

Once the participant was ready to begin the trial, he pressed the space bar and transitioned to the pre-trial screen, which instructed him to reposition his right hand over the mouse and press the space bar once more with his left hand when he was ready to begin the trial. During the trial, the current position of the slider was displayed on-screen, along with the instruction 'Rate your level of engagement as the excerpt plays.'

In summary, the experiment session always comprised the following seven sections, in order: Forms and questionnaires; overview of EEG/physiology block; net and sensor application; EEG/physiology block; removal of net and sensors; training of continuous behavioral block; and continuous behavioral block. On average, a complete session lasted between 1–1.5 hours.

Continuous behavioral responses were always collected in the second block, after the EEG/physiology recording. The reason for this was that we did not want participants to have a specific definition of engagement, or a behavioral task, in mind while the neurophysiological responses were collected, as these responses are meant to involve no cognitive effort on the part of the participant beyond attending to the stimuli. Consequently, the continuous behavioral responses were always collected as the second listen to each stimulus. We note that Steinbeis et al. (2006), who also employed separate blocks for EEG/physiology recordings and continuous behavioral responses (rating tension and emotion), presented the continuous behavioral block first. However, participants in that study were given a different task in the neurophysiology block (comparing lengths of stimuli), so they too were likely not impacted by the task of the behavioral response block.

Apparatus and Equipment

Neurophysiological responses collected during the first experimental block were recorded using the Electrical Geodesics, Inc. (EGI) GES 300 platform. The EEG sensors, amplifier,

and acquisition computer are as described in the previous experiment. ECG responses were recorded using a two-lead configuration with adhesive Covidien Kendall H135SG hydrocel Ag-Cl snap electrodes. Respiratory activity was measured using thoracic and abdominal belts that plugged into a Z-Rip Belt Transducer Module. The ECG and respiratory sensors and apparatus were obtained from EGI and are approved for use with their Polygraph Input Box (PIB). The electrode net provided the ground for these physiological inputs. We additionally attempted to measure GSR. As EGI provided no approved sensor set for measuring this response, we attempted to integrate a custom apparatus into the PIB, using two Velcro snap electrodes attached to the distal phalange of the index and middle fingers of the participant's non-dominant (left) hand. We later determined that these responses were not being properly recorded, and therefore exclude these responses from analysis. The ECG leads, output leads of the Z-Rip Module, and output leads of the GSR apparatus were plugged into the EGI PIB, which in turn plugged into the Net Amps amplifier. The EEG and physiological responses were therefore recorded simultaneously, and synced precisely to the auditory stimuli using the click track we embedded in the second audio channel. Physiological sensor placement is shown in Figure 5.2.

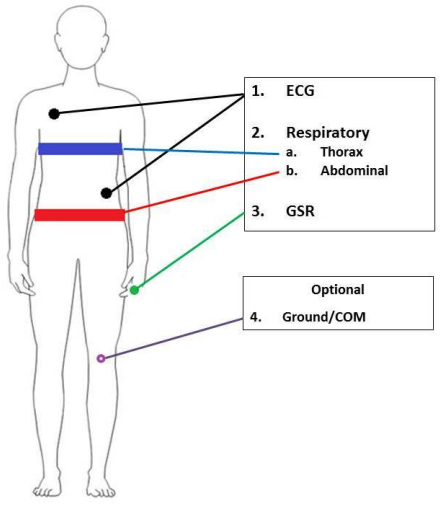

Figure 5.2: Physiological sensor configuration. ECG electrodes were affixed directly to the skin, while respiratory belts were worn over the participant's clothing. Sensors were grounded by means of the EEG net, rather than through the optional electrode affixed to the knee.

EEG and physiological responses were acquired at a sampling rate of 1 kHz. EEG data were referenced to the vertex with no filtering at acquisition. We incorporated a custom .xml file into the EGI acquisition template for this experiment, specifying filter settings to apply to physiological responses at acquisition. ECG and respiratory responses were highpass filtered at 0.1 Hz and lowpass filtered at 100 Hz, with a notch filter at 60 Hz. GSR responses were highpass filtered at 0.05 Hz and lowpass filtered at 3 Hz, with a notch filter at 60 Hz.

Experiments for both blocks of the experiment were programmed using Matlab's Psychophysics Toolbox (Brainard, 1997) on Matlab software, version R2013b, on the same stimulus computer used previously. The hardware configuration for presenting stimuli and delivering stimulus trigger events to the EGI amplifier are as described in the previous experiment.

For the continuous behavioral block, we used a custom slider interface implemented within the Psychophysics Toolbox scheme to simultaneously play the stimuli, display the current state of the slider, and collect the behavioral responses. While the EGI system was not used in this block of the experiment, we used the same audio playback configuration as in the neurophysiological block, and maintained the interface to the EGI amplifier. Continuous behavioral responses were recorded onto the stimulus laptop at a sampling rate of approximately 50 Hz.

5.2.5 Data Analysis

All analyses were performed using Matlab software, version R2013b.

EEG Responses

Each EEG recording was first zero-phase bandpass filtered between 0.3–50 Hz, temporally downsampled by a factor of 8 (to a sampling rate of 125 Hz), and exported to .mat file format using EGI's Net Station software. Following that, the same EEG preprocessing pipeline described in Experiment 1 was used here. Output data variables from this procedure included the behavioral ratings of the stimuli as well as four electrodes-by-time data matrices pertaining to the two trials and their respective baseline recordings (the EEG baseline recordings are not analyzed further).

Three of the 16 participants' data were excluded during preprocessing on the basis of gross EEG artifacts (having either 20 or more bad electrodes, or 20% or more of the data

flagged as transients in the final preprocessing stages), leaving 13 usable datasets. We note that the EEG recordings for this experiment were much noisier overall than the data from Experiment 1. We believe this was caused by the GSR apparatus, which was not designed for use with the EGI system. We have since collected data from a 17th subject with no GSR sensors, and data quality appears to have improved substantially.

Cleaned EEG data frames for each stimulus were aggregated across the set of usable participants into a single 3D time-by-channels-by-participants matrix as input to RCA. For an eight-minute stimulus with a sampling rate of 125 Hz, the data frames are thus $60001 \times 125 \times 13$ samples in size. RCA was computed across the two trial matrices of data as described in Chapter 2, with an effective sample size of $\binom{13}{2} = 78$ participant pairs. Subsequent ISC analysis is performed on the one-dimensional subspace of the RC1 data projection only. We also computed RCA on responses to each stimulus separately in order to assess the RC1 topographies; these RCs are included for visualization purposes only and were not used for ISC analysis. Due to the noisiness of the current data, we lengthened the ISC window from 5 seconds to 10 seconds. Thus, time-resolved ISCs were computed on RC1 data using a 10-second correlation window advancing in 1-second increments.

Continuous Behavioral Response Preprocessing

The continuous behavioral (CB) responses were aggregated across subjects into a time-by-participants matrix. The length of the response vector varied slightly in length across recordings (typically by one or two time samples). In order to analyze the responses in aggregate, we truncated each response to the length of the shortest response vector, which was 9,589 samples. Thus, the effective sampling rate was 19.975 Hz. Continuous behavioral ratings are known to suffer from reliability issues at the start and end of a stimulus (Olsen et al., 2014). To account for this, and for any possible startle responses at the onset of the stimulus (Salimpoor et al., 2009), we discarded the first and last 10 seconds of response from each continuous behavioral response. Once these potentially transient portions of the data were removed, we z-scored the remainder of the data for each participant, so that each response had a mean of zero and standard deviation of one.

ECG Preprocessing

The Net Station waveform tool used to bandpass filter the EEG data prior to downsampling does not apply filters to physiological recordings. Since to simply apply the subsequent

downsampling step to these physiological responses would introduce a risk of aliasing (low-pass filtering to 100 Hz but downsampling by a factor of 8), we performed a second .mat export of the neurophysiological data with no filtering or downsampling, and performed all preprocessing procedures for the physiological responses in Matlab.

Our first step was to extract the stimulus triggers and derive the corrected trial onset times, according to our eventual sampling rate of 125 Hz. Next, we used Matlab's `decimate` function to lowpass filter and downsample the ECG data across the entire recording, first applying an 8th-order Chebyshev Type I filter with a cutoff frequency of 50 Hz, and then resampling the data to 1/8th the original sampling rate, to $fs = 125$ Hz. Performing this operation over the entire recording (before epoching) avoids filter artifacts during baseline and trial epochs. Following this, we epoched the data into baseline and stimulus trials and saved the output. Cleaned ECG responses for each stimulus were aggregated across subjects into time-by-participants matrices.

Next, we converted the ECG activity to HR over time, measured in BPM. As the ECG waveforms sometimes exhibited low-frequency activity over the course of a stimulus, we used the continuous wavelet transform to obtain the R-peak coefficients from the ongoing ECG, from which the timings of the peaks could then be derived. Time-resolved HR was then computed from temporal intervals between successive R peaks: For peaks R_i and R_{i+1}, the instantaneous BPM $60/(R_{i+1} - R_i)$ was mapped to the midpoint of that time interval, $(R_i + R_{i+1})/2$. Finally, the vector of instantaneous BPM over time was spline-interpolated to $fs = 125$ Hz so that all responses could be analyzed over a common time axis.

The purpose of the baseline recordings during the experimental sessions was to acquire a level against which to measure each participant's physiological activity during the subsequent trial. For each participant and trial, we computed a baseline BPM value as the mean BPM from 25–55 seconds (inclusive) of the 60-second baseline preceding that trial. Using the latter half of the baseline period gave the participant adequate time to relax and reach a steady state of cardiac activity, while avoiding spline-interpolation artifacts at the very end of the epoch.

Using the mean baseline values computed above, we subtracted, from each participant's interpolated trial data, the mean baseline BPM. Thus, subsequent analyses consider *deviation* from baseline over the course of a stimulus. Due to possible orienting responses at the start of a trial (Lundqvist et al., 2009), as well as spline-interpolation artifacts at the start and end of each trial, we discarded the first and last 5 seconds of HR data for each trial.

Respiratory Preprocessing

The respiratory responses were preprocessed in a similar fashion as the ECG responses. The main difference here is that the frequency range of interest for respiratory responses is lower, since our computations are based on the rise and fall of breathing activity and not the sharp R peaks of the ECG. Thus, for the chest and abdomen respiratory responses we lowpass filtered the data across the entire recording, using a zero-phase 8th-order Butterworth filter with a cutoff frequency of 1 Hz. Following that, we temporally downsampled the data by a factor of 8, again to a sampling rate of $fs = 125$ Hz. Baseline and trial epochs were then aggregated across subjects into time-by-participants matrices. We observed during preprocessing a significant correlation between the chest and abdomen respiratory activity. Therefore, we focused our analysis solely on the chest respiratory responses, which tended to show more variation in amplitude.

While for the ECG responses we were interested only in BPM over time, for the respiratory responses there are two response features of interest: Respiratory *rate* and *amplitude* over time (RRate and RAmpl). Our first step in computing these time-resolved measures was to identify the positive and negative peaks across each response. To do this we employed a peak-finding algorithm with default peak magnitudes of zero and default inter-peak interval of 2 seconds (separate intervals for positive and negative peaks). We then verified the correctness of this procedure by confirming that positive and negative peaks were interleaved; if they were not, we manually checked and corrected the problematic peaks.

Once the positive and negative respiratory peak times were identified, we computed time-resolved RRate using the same procedure used for the ECG responses: The time difference between adjacent peaks was mapped to instantaneous breaths per minute, and this value was mapped to the midpoint between the two peak times. Temporally resolved RAmpl was computed using the amplitude difference between successive positive-negative (peak-to-trough) and negative-positive (trough-to-peak) peak values, with magnitude distances between peaks mapped to the midpoint in time between peaks. As the RAmpl calculation utilized both positive and negative peaks, this response vector contained twice as many data points as RRate over the same range of time. Finally, both response vectors were spline-interpolated to a sampling rate of 125 Hz.

As with the ECG responses, we computed mean baseline RAmpl and RRate values from 25-55 seconds, inclusive, of each baseline epoch, and subtracted the baseline measure from the response vectors for the trial. As above, we discarded the first and last 5 seconds of

response for each trial.

Analysis of Continuous Behavioral and Physiological Responses

After preprocessing, the aggregated continuous behavioral, ECG, and respiratory responses were stored in 2D time-by-participants matrices. Continuous behavioral matrices were 9189×13 after trimming the first and last 10 seconds of response, while the physiological response matrices were 58751×13 after trimming the first and last 5 seconds of response. For plotting results pertaining to response activity, we present median values of responses across participants.

Statistical Analyses

Participants rated each stimulus along five dimensions (pleasantness, arousal, interestingness, predictability, and familiarity) in the first experimental block. As we had no incoming expectation that the responses would be unidirectionally higher or lower for original versus reversed stimuli, we performed paired two-tailed t-tests on the responses to these questions across all participants, using the Bonferroni correction for multiple comparisons (McDonald, 2014). For the sixth question of genre, which was answered only for the original stimulus, we report the median rating across subjects.

To assess the statistical significance of the mean ECG, respiratory rate, respiratory amplitude, and continuous behavioral responses, we performed two-tailed non-parametric Wilcoxon signed-rank tests across the population of responses at every time point, which assessed whether said responses reflected a zero-median population. We corrected for multiple comparisons using FDR (Benjamini and Yekutieli, 2001).

Statistical significance of all temporally resolved ISCs (EEG, CB, HR, RAmpl, RRate) was assessed via permutation test. As described in the previous experiment, we partitioned the data into non-overlapping 5-second windows, which were permuted independently for each participant. ISCs were then computed over the set of permuted data records. This procedure was repeated 500 times. Time points at which ISCs of intact responses exceed the 0.95 quantile of all permutation permutations are considered statistically significant.

To quantitatively compare the proportion of significant ISCs for each response across stimulus conditions, we performed the non-parametric Chi-squared test of proportions (DeGroot and Schervish, 2002). Here we again applied the Bonferroni correction for multiple comparisons across the nine tests (McDonald, 2014).

5.3 Results

5.3.1 Behavioral Ratings

We first summarize the behavioral ratings collected during the EEG/physiology block. Summary boxplots of responses, overlaid with individual ratings for each question, are shown in Figure 5.3 along with p-values from the two-tailed paired t-tests. With a Bonferroni-corrected p-value threshold of $p_B = 0.01$, only the dimensions of pleasantness and predictability vary significantly according to stimulus condition. In the both cases, the original (forward) stimulus receives higher ratings overall. We therefore conclude that the original stimulus was perceived to be more pleasant and more predictable, but not necessarily any more arousing, interesting, or familiar, than the reversed version. We note that familiarity with the stimuli was low overall, though two of the 13 participants reported that they were at least moderately familiar with the reversed stimulus. Finally, reported genre exposure ranges from 2–9 with a median value of 5, verifying that all participants met the inclusion criterion of listening to classical music at least occasionally.

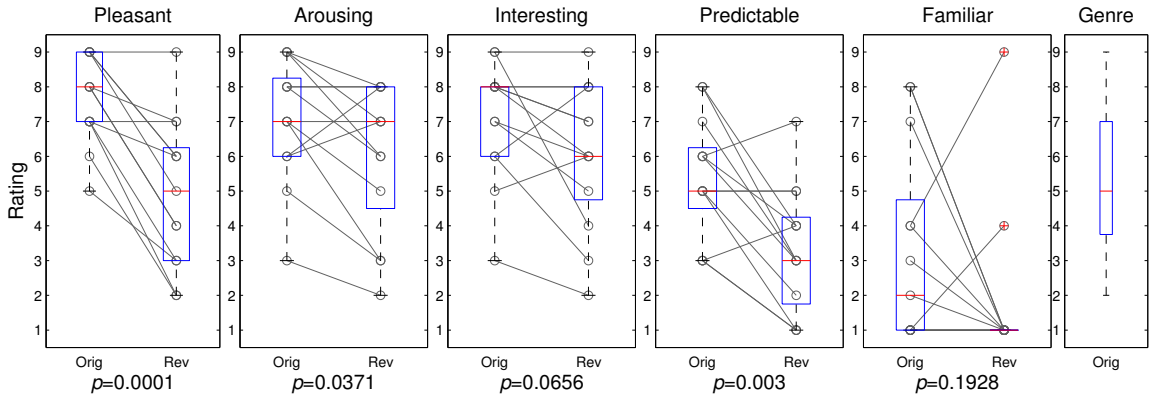

Figure 5.3: Behavioral ratings of Elgar stimuli. Participants rated the pleasantness, arousal, interestingness, predictability, and familiarity of each stimulus after it played. Participants reported how often they listened to this genre of music for the original stimulus only. Only pleasantness and predictability ratings vary significantly by stimulus condition after Bonferroni correction.

5.3.2 EEG Responses

RC1 Topography

We performed RCA over responses to both stimuli together for ISC analyses, as well as over responses to each stimulus separately for visualization purposes. The forward-model projected component topographies are shown in Figure 5.4. We note that these topographies are similar to the RC1 topographies from Experiment 1 (Figure 3.3, Figure 3.4), and that the present topographies are consistent whether derived from one or both stimuli. However, the present topographies are also less smooth and symmetric than those derived in the previous experiment. We believe this is likely connected to the increase in data artifacts observed during preprocessing.

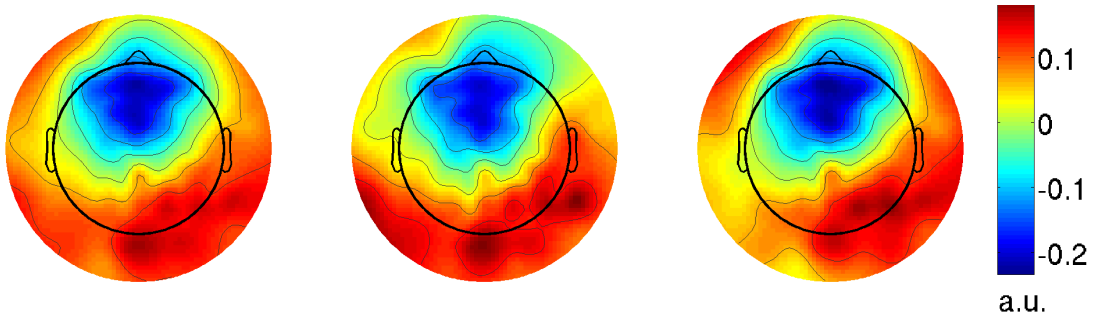

Figure 5.4: RC1 topographies for responses to Elgar stimuli. RCA was computed over responses to both stimuli (left), as well as over responses to the original (center) and reversed (right) stimulus only. Topographies are roughly consistent with RC1 topographies from Experiment 1, though less smooth and symmetric.

EEG-ISCs

Next, we computed ISCs from the RC1 EEG responses, using a 10-second window advancing in 1-second increments, with statistical significance assessed over 500 permutation iterations using a 5-second partition window. Results are plotted in Figure 5.5. Here, the proportion of significant ISCs is 29.30% for the original stimulus and 31.42% for the reversed stimulus—not a statistically significant difference (see Figure 5.12).

In terms of designated musical events of interest, we note that ISCs reach statistical significance during both regions that build to structural highpoints (B1–C1, B2–C2). However, synchrony is not significant at or after the highpoints themselves. As we had conjectured,

EEG-ISCs are also significantly high at A2, the return of the cello theme. For the reversed stimulus, significant ISCs span broader temporal intervals around each highpoint (C2, C1), as well as an extended interval around what was the second entrance of the solo cello theme in the original version (A2).

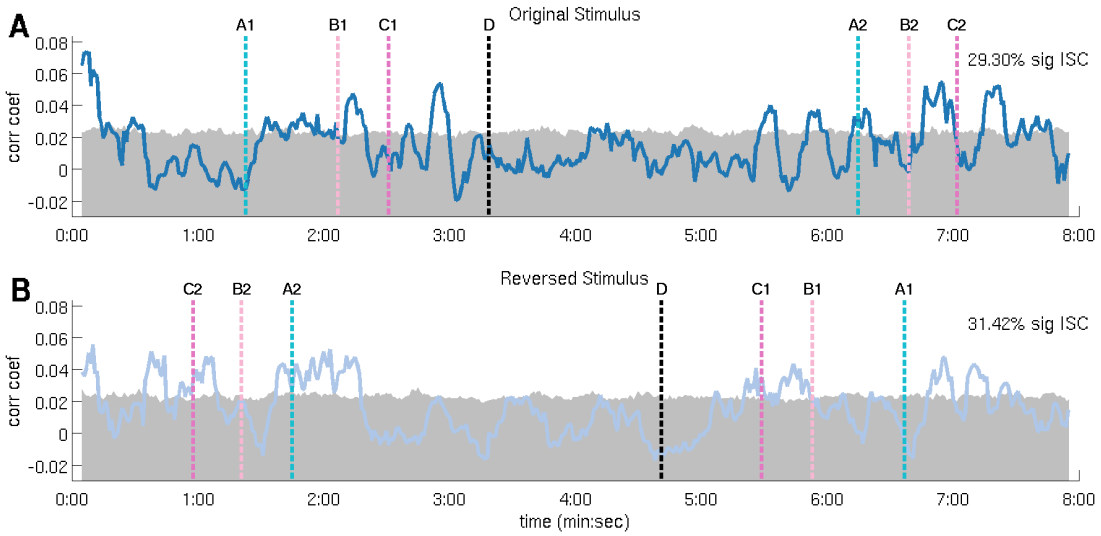

Figure 5.5: Elgar EEG-ISCs. Panel A: ISC peaks in response to the original stimulus occur while tension builds to highpoints (B1–C1, B2–C2), and at the re-entrance of the solo cello theme (A2). Panel B: The reversed stimulus significant ISCs at and around both highpoints (C2, C1), and around the second entrance (now the first exit) of the solo cello theme (A2). The proportion of significant EEG-ISCs does not differ significantly across stimulus conditions (Figure 5.12).

5.3.3 Continuous Behavioral Responses

For the EEG responses, we have focused solely on the synchrony of the responses across participants, and not on the voltage amplitudes themselves. However, for all other responses collected in this experiment, we may assess both the activity—or level, in terms of deviation from baseline, and the synchrony—measured with ISCs—over the course of the stimuli.

Both forms of the CB results are shown in Figure 5.6. For the original stimulus, regions of statistical significance of both the median z-scored activity and the ISCs relate to our musical events of interest. Interestingly, median level of engagement is significantly *below*

zero around the entrance of the first cello theme (A1), but later reaches significantly positive levels, most notably at and after both of the highpoints of the excerpt (C1, C2). Synchrony of reported engagement with the original stimulus peaks around the start of the first buildup of tension (B1), at both highpoints (C1, C2), and also shortly after the drop in activity (D), but does not show the extended periods of significance displayed by engagement level following the structural highpoints. The reversed stimulus produces no significant level of reported engagement; however, synchrony of engagement for this condition does reach statistical significance after the solo cello entrance (exit) at A2, as well as highpoint C1. Proportions of CB level, but not synchrony, are significantly affected by stimulus condition (Figure 5.12).

5.3.4 ECG Responses

Music physiology studies exhibit a lack of consensus regarding the impact of musical emotion and arousal on HR deviation from baseline. Our present results, shown in Figure 5.7, indicate that neither stimulus produces significant HR deviation from baseline across participants (top subplots of Panel A and B). However, we do observe some significant results in HR synchrony. For the salient events in the original stimulus, this appears at the drop in activity (D). For the reversed stimulus, HR synchrony is significant after the solo cello entrance (exit) at A2, as well as following the drop at D and after what would have been the first tension build at B1. Proportions of significant HR response measures do not vary significantly by stimulus condition (Figure 5.12).

5.3.5 Respiratory Responses

For respiratory responses, we know from previous studies that respiratory rate has been found to increase for music expressing fearful or happy emotions, as well as heightened arousal, tempo, or staccato articulations (Krumhansl, 1997; Gomez and Danuser, 2004; Russo et al., 2013; Gomez and Danuser, 2007). We had no specific expectations, based on past findings, with regard to respiratory amplitude. RAmpl results are shown in Figure 5.8. For the original stimulus (Panel A), respiratory amplitude is significantly different from baseline shortly before the first entrance of the cello (A1) and also in the winding down of activity leading to event D. In both cases, median RAmpl is shown to be below baseline. Synchrony of RAmpl implicates both structural highpoints of the excerpt. For the reversed

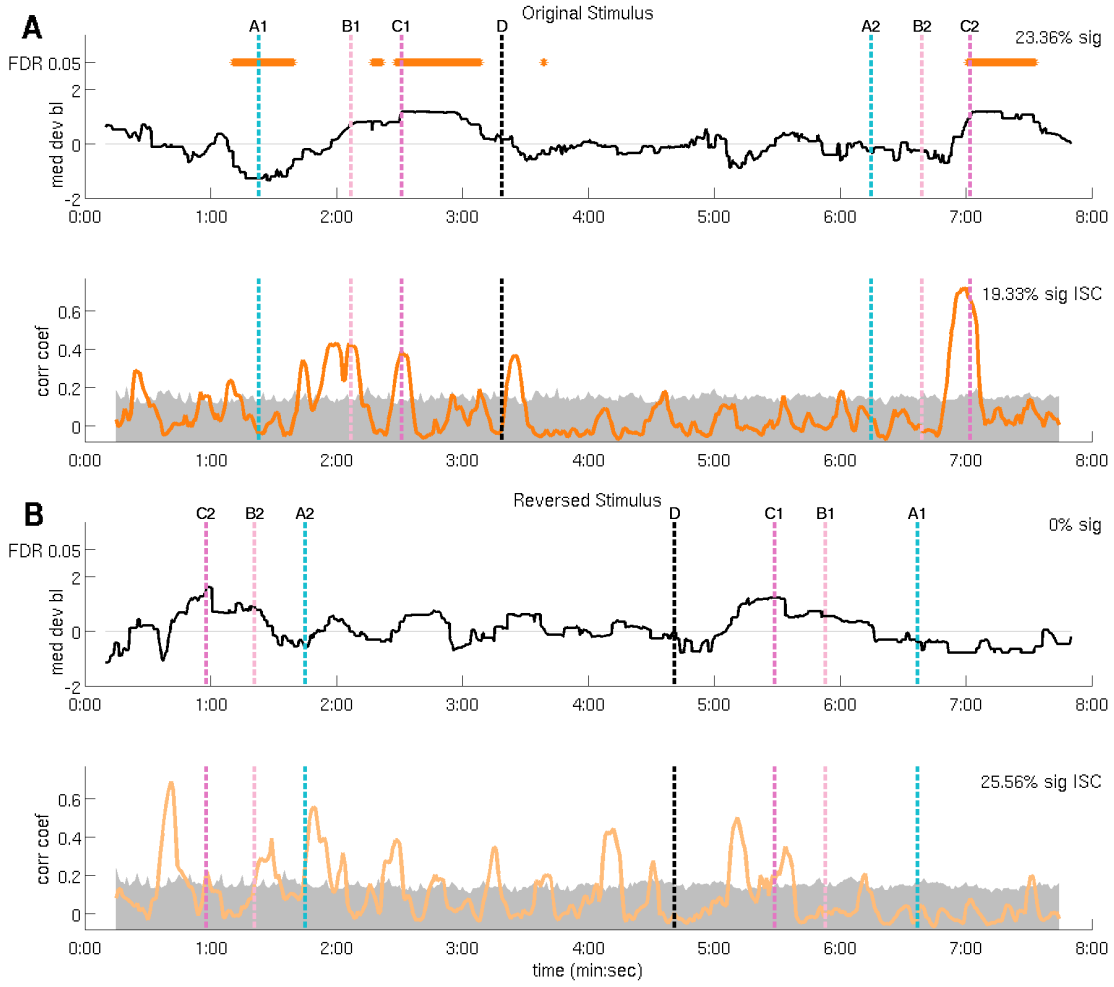

Figure 5.6: Continuous behavioral responses. Panel A: Median level (top) and ISCs (bottom) in response to the original stimulus. Panel B: Median level (top) and synchrony (bottom) in response to the reversed stimulus. Asterisks denote statistically significant results for level; regions of the curve exceeding the shaded gray area denote statistically significant ISCs.

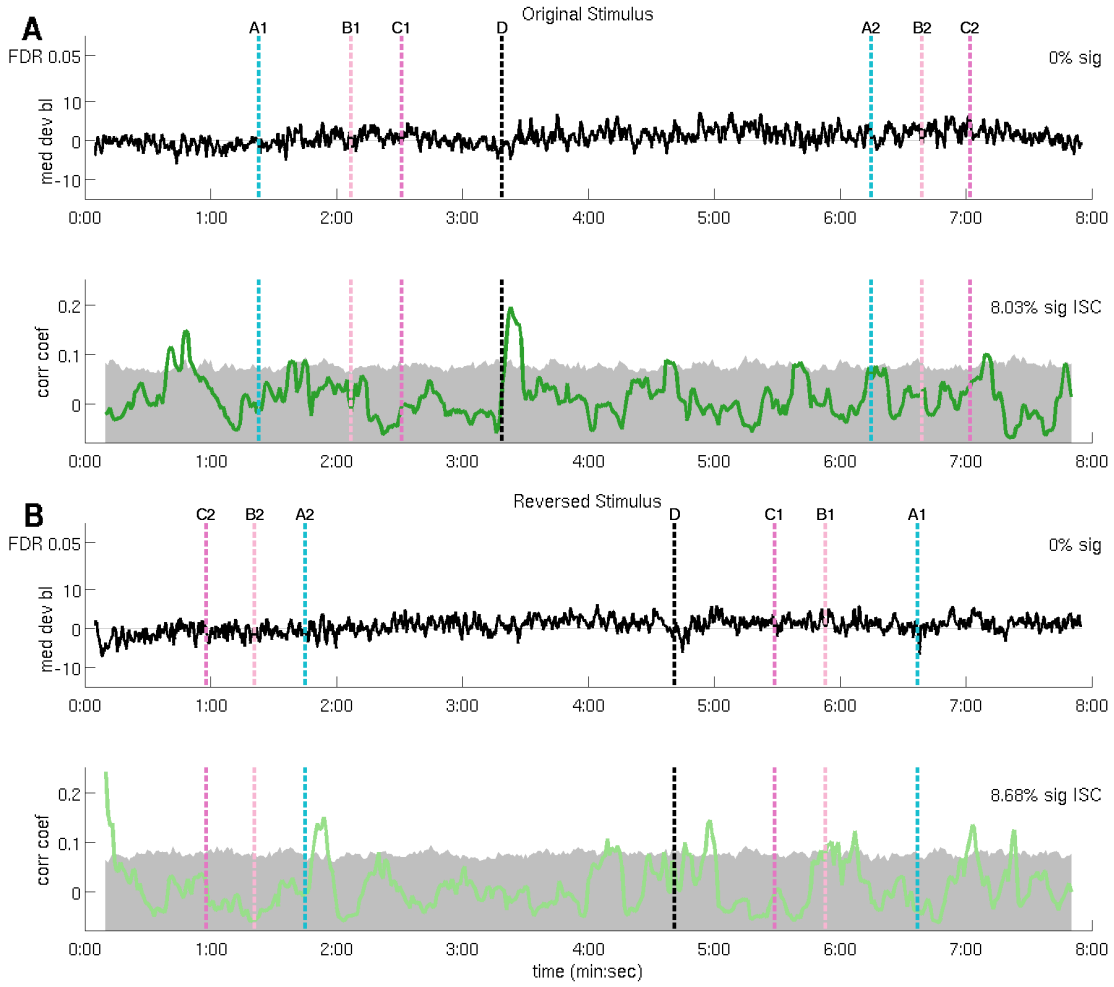

Figure 5.7: HR activity and synchrony over time. Panel A: The original stimulus brings about no significant deviation from baseline (top), though synchrony is significant around the drop in musical activity (bottom, D). Panel B: There is again no significant deviation from baseline for the reversed stimulus (top), but synchrony is significant around A2, D, and B1 (bottom).

excerpt (Panel B), there are no regions of statistically significant RAmpl deviation from baseline, while RAmpl-ISCs are significant at what was the second solo cello entrance (A2).

While we do not see a significant increase in RRate (Figure 5.9) at the structural highpoints (assumed to be the points of highest arousal in the excerpt), we do see a brief period of significant rate in the second buildup (Panel A, top, after B2). RRate synchrony, however, is significant at the first cello entrance (A1), the first highpoint (C1), and the second buildup (B2). For the reversed stimulus (Panel B), RRate is significantly above baseline during the periods leading up to what were the structural highpoints (C2, C1). RRate synchrony here is significant at only one of our designated salient events (highpoint C1).

Overall, the proportion of significant RAmpl and RRate levels, but not synchrony, differs by stimulus condition (Figure 5.12). The original stimulus brings about a larger proportion of RAmpl deviation from baseline, while the reversed brings about a larger proportion of RRate deviation from baseline.

5.4 Discussion

5.4.1 Main Findings

In this study, we supplemented EEG-ISCs with physiological and continuous behavioral responses in order to better understand the role of cortical synchrony as a measure of engagement. We used a full musical excerpt characterized by dramatic fluctuations in arousal, and predetermined a set of musically salient events to guide our interpretation of results. For the new responses, we evaluated both deviation from baseline and ISCs, as these measures have been shown to highlight different facets of engagement and expectation (Schubert et al., 2013).

Indeed, our results for the present experiment have shown that level and synchrony of a given response can highlight different stimulus events. The CB responses to the original stimulus in particular show that level may implicate periods of high excitement (Figure 5.6), while synchrony may relate more to specific events of shorter duration. These findings provide an interesting complement to what Schubert et al. (2013) termed 'gem moments'. Recall that those authors found engagement level to be high in response to surprising events, but engagement agreement (or synchrony) to be high when expectations have been established. Here we find level to be high once a highpoint has occurred, but synchrony to be high for shorter intervals around a variety of musically salient events. For the physiological

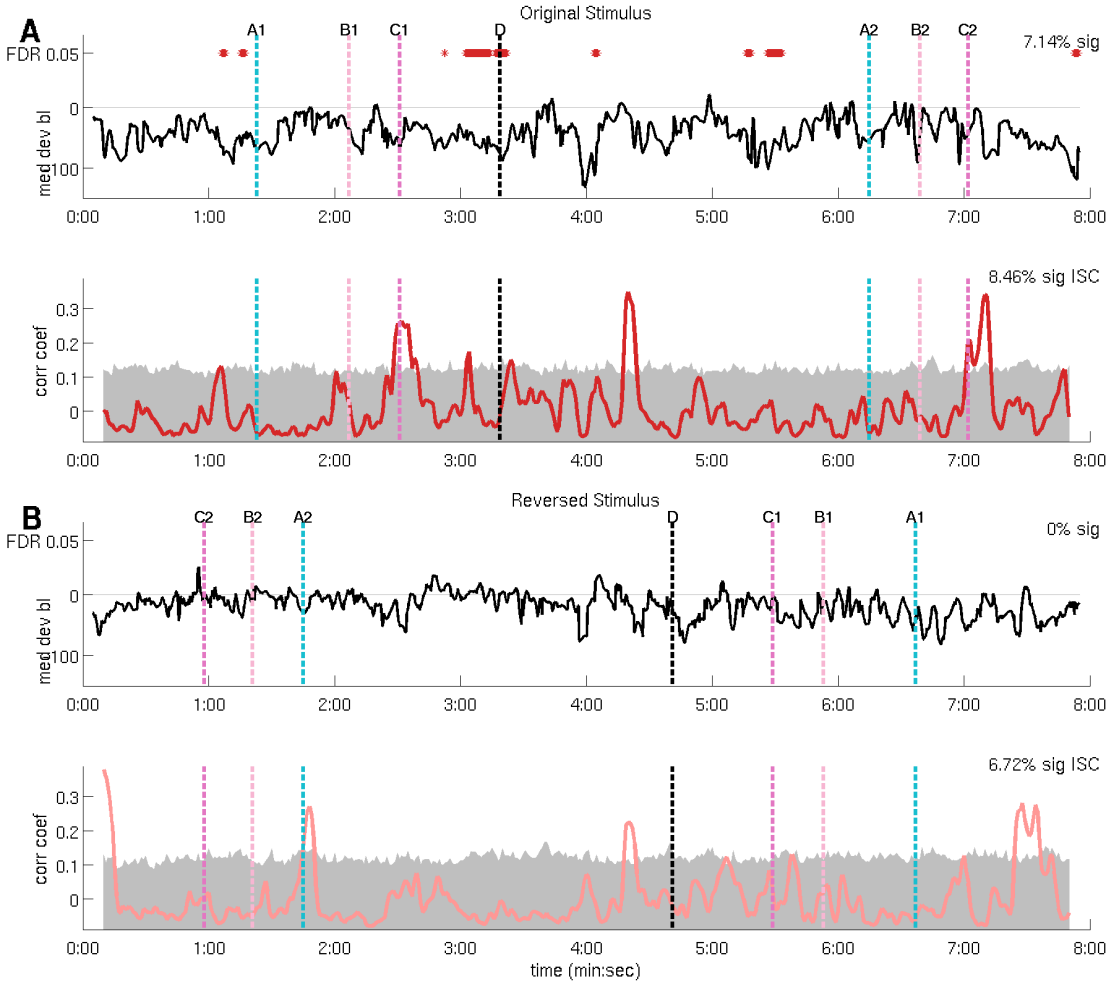

Figure 5.8: Respiratory amplitude over time. Panel A: Amplitude level (top) is significantly below baseline leading up to the first solo cello entrance (A1) and the drop in activity (D) for the original stimulus. Synchrony of respiratory amplitude (bottom) is significant around both highpoints (C1, C2). Panel B: There is no significant deviation from baseline for the reversed stimulus (top), but synchrony is high around the second entrance (first exit) of the cello theme (bottom, A2).

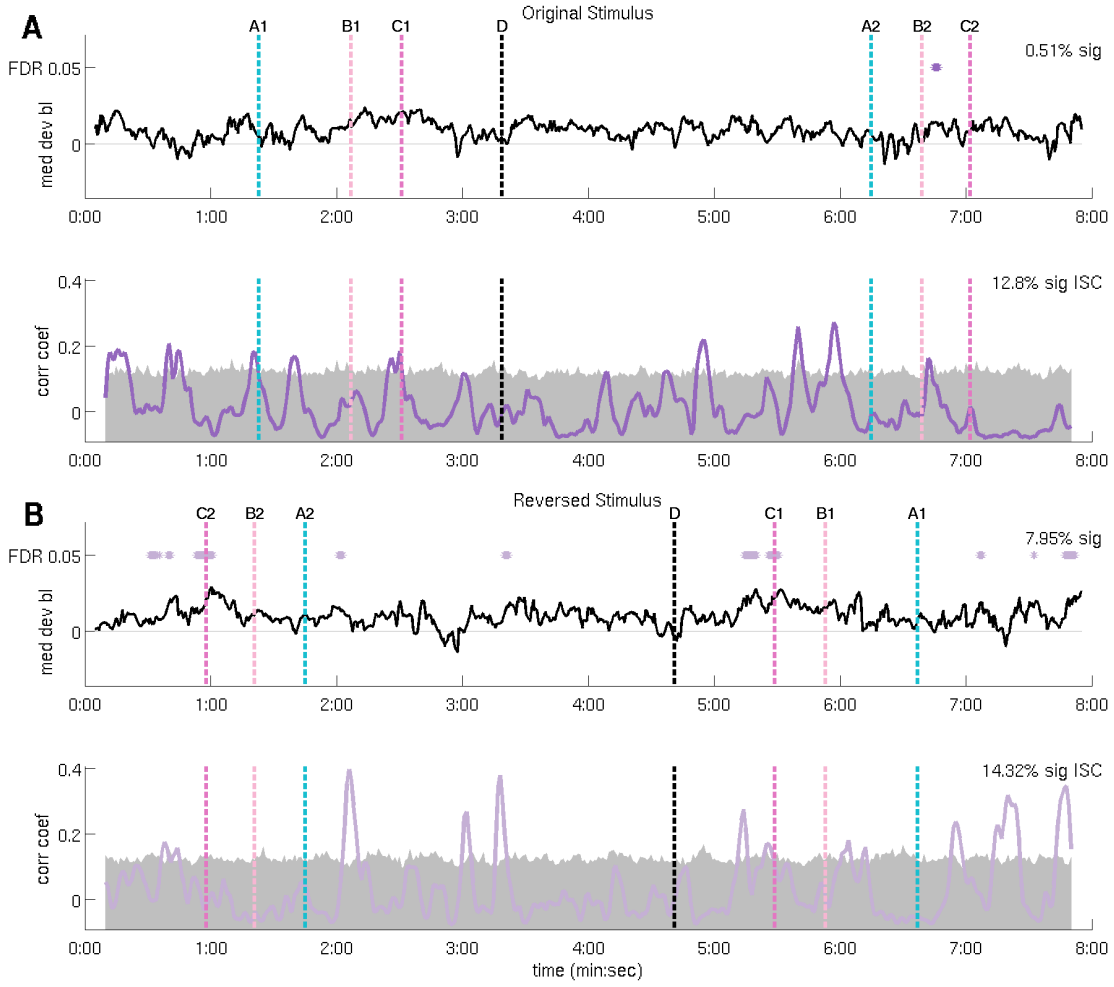

Figure 5.9: Respiratory rate over time. Panel A: Deviation from baseline for the original stimulus is significant only briefly during the second buildup to a structural highpoint (top, B2). Synchrony (bottom) is significant at a few demarcated musical events. Panel B: For the reversed stimulus, respiratory rate is significantly above baseline leading up to structural highpoints (top, preceding C2, C1). Synchrony of respiratory rate co-occurs with only one stimulus event (bottom, C1).

responses, too, we found that response synchrony often implicated different musical events than did level, and was especially informative in cases where there were no significant deviations from baseline, for example with HR (Figure 5.7) and RAmpl (Figure 5.8).

We find also that some specified musical events bring about a greater number of significant response measures than others. For the original stimulus, whose responses are plotted together in Figure 5.10, both structural highpoints are associated with a number of significant responses: C1 brings about significant CB level and ISC, as well as Rampl and RRate ISCs, while significant EEG-ISCs, CB level and ISCs, HR-ISCs, and RAmpl-ISCs occur at or near C2. Interestingly, the moment noted for its lack of activity, D, is associated with high ISCs for CB, HR, and RAmpl, as well as RAmpl level.

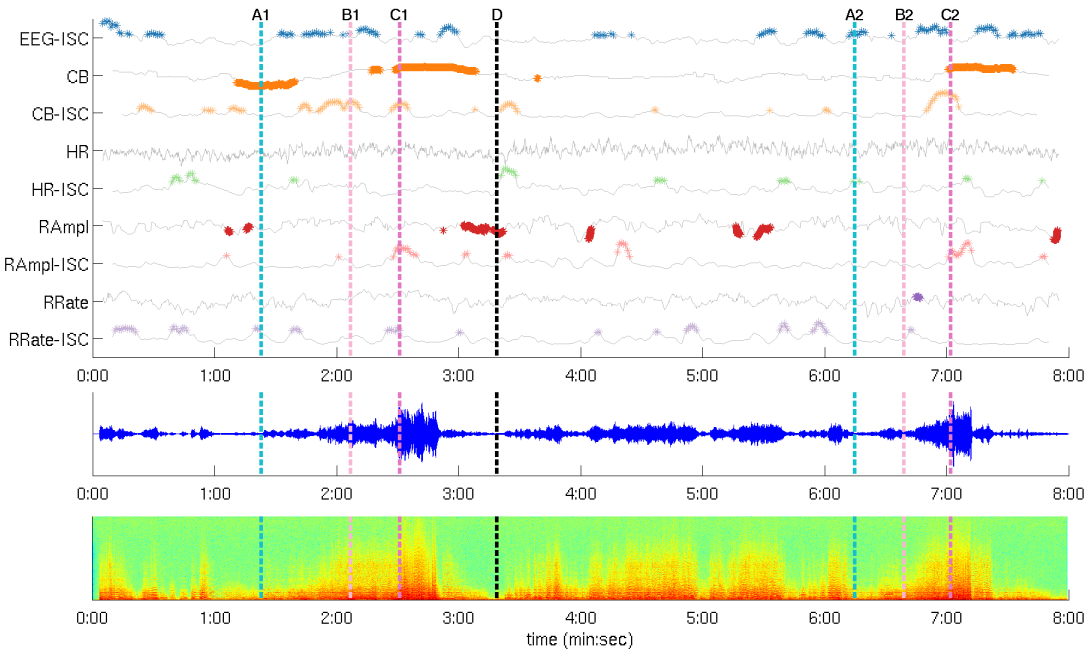

Figure 5.10: Aggregate responses, original stimulus.

For the reversed stimulus, whose responses are aggregated in Figure 5.11, the musical events producing numerous significant responses are what was the re-entrance of the cello theme in A2, which here would be the exit of that theme. Here, ISCs of EEG, CB, HR, and RAmpl are statistically significant. The second notable event for this stimulus is what was the first highpoint, C1, which is accompanied by statistically significant EEG, CB, and

RRate ISCs, as well as RRate level.

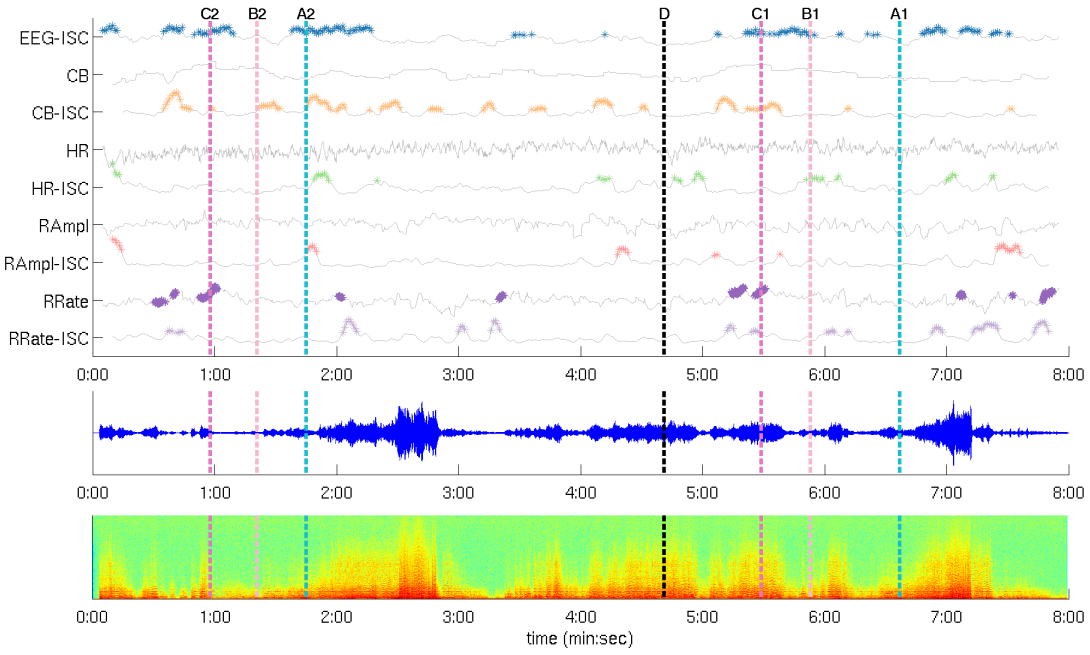

Figure 5.11: Aggregate responses, reversed stimulus.

Finally, we can assess whether the stimulus condition impacted the proportion of significant results for each response measure. As can be seen in Figure 5.12, there is no systematic effect of stimulus condition on the proportion of significant results. After applying the Bonferroni correction to the output of non-parametric tests of proportions, we find that the only measures significantly impacted by stimulus condition are CB level and RAmpl level (both higher for the original than the reversed stimulus), as well as RRate level (higher for the reversed stimulus).

5.4.2 Considerations

The combined level and synchrony analysis of neurophysiological and behavioral responses employed here is a promising approach toward achieving a better understanding of musical engagement; the relation of engagement to arousal; and the distinction between objective and subjective responses. There are several interesting directions this research could take—for example, to assess the effectiveness of music on mediation of stress (Labbé et al., 2007)

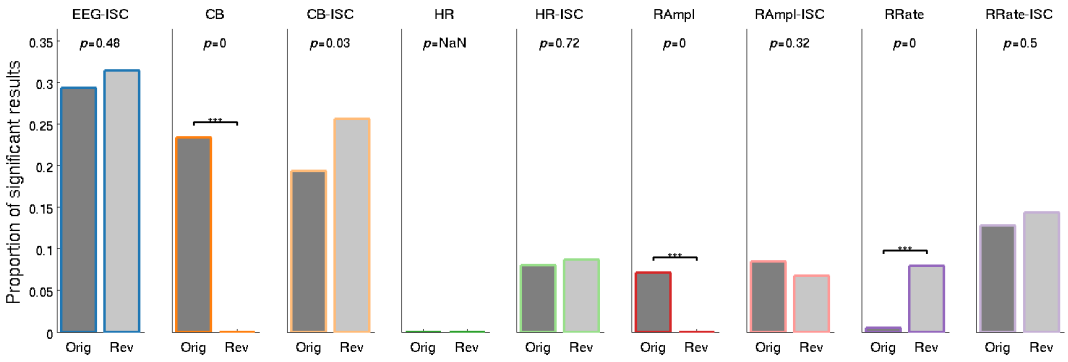

Figure 5.12: Summary of proportion of significant results for each response. Horizontal bars with asterisks denote statistical significance ($p < 0.05$) after applying the Bonferroni correction to the stated p-values.

or develop physiology-based recommendation systems (Shin et al., 2014). However, we must also acknowledge several areas that could be improved in future studies of this kind.

Experimental Design

First, there was a potential confound of familiarity with the stimulus. Since we used a well-known piece from the classical repertoire and sought participants with musical training and exposure to classical music, we knew it was a possibility that some or all participants would already know the original excerpt. Thus, while median familiarity with the original version was low (2 on a scale of 1–9), there were varying degrees to which a participant knew the original better than the reversed version, which could have played a role in the differential responses between stimulus conditions (van den Bosch et al., 2013). This could be avoided in the future by using musical excerpts by lesser-known composers (an approach taken by Sridharan et al. (2007) and Abrams et al. (2013)). Another approach would be to recruit nonmusician participants, though this could also impact the likelihood that they would find music from this genre engaging.

It has also been noted that participant-selected stimuli may more reliably evoke physiological responses. Along these lines, we could invite participants to bring in their own excerpts and focus the analysis on personal, rather than aggregate responses, as has been done in previous studies (Rickard, 2004; Grewe et al., 2005, 2007b; Salimpoor et al., 2009).

We acknowledge that always placing the continuous behavioral block second in the experiment may have imposed confounds of familiarity or fatigue. However, we felt that

this was preferable to collecting the neurophysiological responses from participants who had already been exposed to a specific definition or task regarding an experience of engagement. An alternate approach would be to divide the participants into two groups, each of which would complete only one of the two experimental blocks.

While we were interested in the buildup to the structural highpoints in the original excerpt, especially in terms of how those events unfold over time, we now feel that stimulus reversal may not have been the best control condition. As can be seen in Figure 5.11, reversed highpoints C2 and C1 are still accompanied by statistically significant responses. However, it is hard to assess whether those responses are driven by the actual musical content between the C and B demarcations (what was the buildup in the original version), or whether responses are driven more by 'afterglow' effects that have been observed after peak musical events (Madsen et al., 1993). Therefore, in the next iteration of this experiment, we will likely employ the amplitude-preserving phase-scrambling procedure we outlined at the end of Experiment 1, which will enable us to better distinguish between the impact of amplitude envelope and underlying musical content—or lack thereof—on listener responses.

Analysis Considerations

We noted earlier that the EEG data were unusually noisy for this experiment, likely due to the GSR apparatus. While it is disappointing that we were not able to collect usable GSR responses (since, as noted in the previous chapter, there are fairly consistent findings using this response), we will likely exclude this response in the future, if continuing to work with the EGI PIB apparatus.

ISC results for this experiment always provide a sample-to-sample temporal resolution of 1 Hz, due to the 1-sec hop size of the ISC analysis window. However, the response vectors, which are used to produce the deviation-from-baseline results, have higher temporal resolution and are therefore longer in length. For example, the CB response vector has a sampling rate of roughly 20 Hz and is 9,189 time samples in length, while the physiological response vectors have a sampling rate of 125 Hz and are 58,751 samples in length. This might be excessive temporal resolution, given the time scale over which these responses are thought to occur. Therefore, we may consider adding a binning step to the level-based analyses in the future, for example by averaging each participant's responses in 1-sec windows. Such a procedure would result in the same temporal resolution as the ISC analyses, while also reducing the number of multiple comparisons in our FDR procedure.

We used ISCs to assess synchrony of all responses. Other measures of synchrony could also be considered, for example the approach used by Schubert et al. (2013) and Schubert (2013), based on the standard deviation of participant responses over time.

Interpretation of Results

Our present analysis mapped the results directly to the point in the stimulus at which they occurred. However, we must keep in mind that varying temporal lags are inherent to every response measure. As noted in the previous chapter, physiological responses are thought to occur up to five seconds after the corresponding stimulus event, with an estimated lag of up to three seconds for continuous behavioral responses. Thus, the reported timing of level results may need to be shifted back in time for these responses. The fact that ISC results are mapped to the midpoint in time of a 10-sec analysis window further complicates the interpretation of results. Therefore, future attempts should consider adjusting these responses accordingly if attempting precise mapping of responses to stimulus events.

For this initial attempt, we interpreted the collection of responses within the framework of predetermined musical events. However, results could also be interpreted in a more data-driven fashion, with high agreement among responses used to highlight musical events of interest. For example, in responses to the original stimulus (Figure 5.10) we note that CB activity and ISC, RAmpl activity and ISC, and RRate ISC are all significant in the general area of 3:00 (during a more subdued solo cello passage). A number of response measures are also significant around 5:45, which implicates a period of heightened tension through an extended dominant. Points of agreement among the responses can also be found for the reversed stimulus (Figure 5.11).

Finally, as in our present design we collect all responses from all participants, it may be interesting to consider how the collection of responses can be aggregated and analyzed at once. For example, a variation of RCA that could operate over combined cortical, physiological, and behavioral responses could derive aggregate component weightings that could clarify the contribution of each response to reliable audience experiences.

Chapter 6

Conclusion

In this thesis we have presented two applications of a novel EEG analysis technique to the study of responses to music. Drawing from a cortical-synchrony theory of engagement, along with experimental approaches employed in studies of engagement in other stimulus domains, in our first experiment we validated the use of RCA and EEG-ISCs in response to full-length, naturalistic stimuli. Here we found that the temporal organization of acoustical events into music plays a significant role in driving reliable cortical responses across listeners—thought to be a key indicator of focused engagement.

In a second experiment, we extended existing research on physiological and continuous behavioral responses to music and related stimuli, analyzing EEG-ISCs in conjunction with other continuous measures of arousal and engagement. Results from this study suggest that cortical, physiological, and behavioral responses may together provide new insights into characterizing the experience of musical engagement.

6.1 A Narrative Framework for Musical Engagement

The state of focused engagement explored in this thesis can be interpreted within the transportation/cognitive elaboration framework for narrative engagement. This framework is thought to characterize distinct states of response to story-based works such as films or novels (Green and Brock, 2000). Here, transportation is defined as a state of absorption or immersion—of being 'lost in a story' (Green and Brock, 2000). Researchers consider this state to be similar to, but different from, enjoyment (Green et al., 2004), a key difference

being that a transported audience will have been transformed by their experience with the work (Green and Brock, 2000; Green et al., 2004). The state of immersion has also drawn some comparisons to 'flow' (Busselle and Bilandzic, 2009).

In contrast, cognitive elaboration implies critical attention rather than immersion (Green and Brock, 2000). Here, each audience member experiences the narrative differently, and interprets and evaluates incoming information through self-referencing of opinions, knowledge, experiences, memories, and beliefs (Green and Brock, 2000; Escalas, 2007).

Effective narratives are often linked to transportation. In an advertising setting, for example, transportation is thought to evoke positive feelings in lieu of analytical thought, while cognitive elaboration may produce more critical thoughts and fewer positive emotions (Escalas, 2004, 2007). Transportation is considered a state of convergent processing across audience members (driven by immersion in the stimulus), and is thus the state that would produce heightened ISCs. Cognitive elaboration, on the other hand, is considered divergent (each audience member has a distinct experience) (Green and Brock, 2000). Therefore, if effective narratives drive transported engagement, and transportation implies synchronous processing, then ISCs may serve to index engagement.

While some musical works are programmatic or reflect narrative content of lyrics, most are at best referential in allusions to extra-musical elements. In considering whether transportation and elaboration are applicable to musical engagement, we reason that musical features that project temporal trajectories and goals—such as cadential formulae in functional tonal music, or performed tempo changes signifying approaches to or departures from salient events—have the capacity to manipulate listener expectations in a manner analogous to narrative devices. As suggested by current results, there appears to be some relationship between heightened cortical ISCs and structural segmentation boundaries between song parts, periods of building tension, and structurally relevant repetitions of musical motives. Further investigation into the role of such events in driving reliable audience responses may help to clarify musically induced states of transportation.

6.2 Future Work

There exist several possible extensions and modifications of the current experimental approaches. Broadly speaking, in the present work we sought to identify salient musical

attributes and events that drive temporally reliable cortical responses across audience members. This approach relates to the perceptual 'locate' research proposed by Honing (2010), and would be interesting to generalize further to other, real-world forms of data that objectively denote interest in specific musical events. Such approaches, especially if applied to the prediction of large-scale musical preferences (Dmochowski et al., 2014; Falk et al., 2012), have potential applications in the field of Neuromarketing (Ariely and Berns, 2010). Davies (2014) has proposed that narratives are 'a primitive kind of virtual reality, making us forget our physical surroundings and feel as though we are transported into the world' of the narrative—a description well aligned with the aforementioned state of transportation. Approaches to quantifying engagement may find novel application in assessing user experiences in actual VR settings in coming years. EEG-ISCs could also be prove to be a useful tool for assessing cortical processing in clinical and rehabilitative settings; for instance, using fMRI-ISCs, Hasson et al. (2009) gained valuable insights into idiosyncratic processing of audiovisual film excerpts by adults with autism.

The use of full, naturalistic works that required no more than one presentation presents a significant advance in the ecological validity of music-EEG experiments. However, the listening setting—sitting still in a darkened room, listening passively with neurophysiological sensors attached—is still somewhat misaligned with the experience of music in real life. As music listening, when it occurs, is often not the main activity (Sloboda et al., 2001; Cunningham et al., 2007), it may be useful to devise experiments to better understand how we engage with music as it plays in the background. Music listening in a shared setting is also lost in the traditional experimental setting (Sloboda et al., 2001), but plays an important role in the listening experience (McAdams et al., 2004; Schubert et al., 2013). Advances in portable and mobile EEG systems have been proposed for future research in music information retrieval studies (Kaneshiro and Dmochowski, 2015), and could facilitate the study of cortical responses collected in a live concert setting, similar to the physiological approach employed by Egermann et al. (2013).

Other facets of engagement are open to cortical investigation as well. As pointed out by Hasson et al. (2008b), low ISCs do not necessarily imply low audience engagement. Rather, they simply reveal that audience members were not processing the stimulus in a reliable fashion (hence our present emphasis on focused engagement). It will be interesting to consider how the present ISC approach might be extended to study cortical representations

of other forms of engagement, including those that would be classified as cognitive elaboration, such as music-invoked autobiographical memories (which have been successfully studied using fMRI (Janata et al., 2007; Janata, 2009)). Another approach could be to analyze EEG responses in the time-frequency domain rather than the time domain (as was done by Dmochowski et al. (2012)) to assess more broadly the state of listeners, rather than processing of specific stimulus events.

6.3 Closing Remarks

The study of musical engagement is a challenging task. It centers on a concept that is not only hard to define but difficult to measure, particularly through the modality of EEG responses. The methodological and empirical contributions of this thesis point to many exciting directions for the study of cortical correlates of engagement and, more broadly, for EEG research on music perception and cognition. It is hoped that this work establishes a strong foundation for future research in musical engagement.

Appendix A

Experiment 1 Supplement

A.1 Stimulus Figures, Songs 2–4

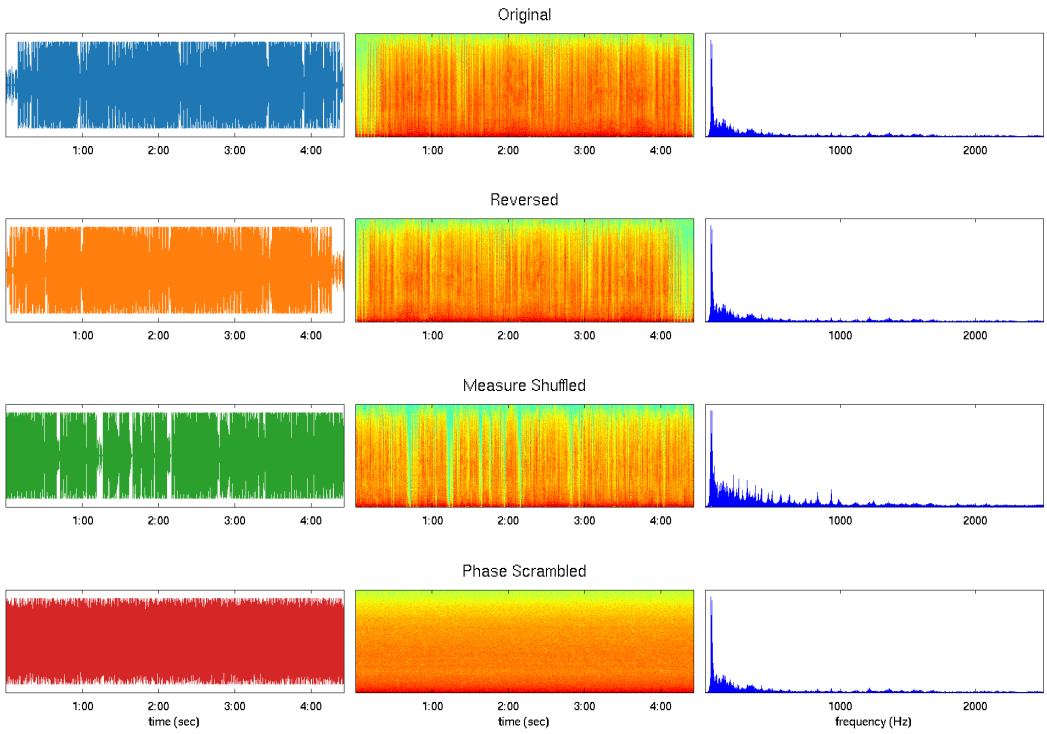

Figure A.1: Waveforms, spectrograms, and magnitude spectra of Song 2 stimuli.

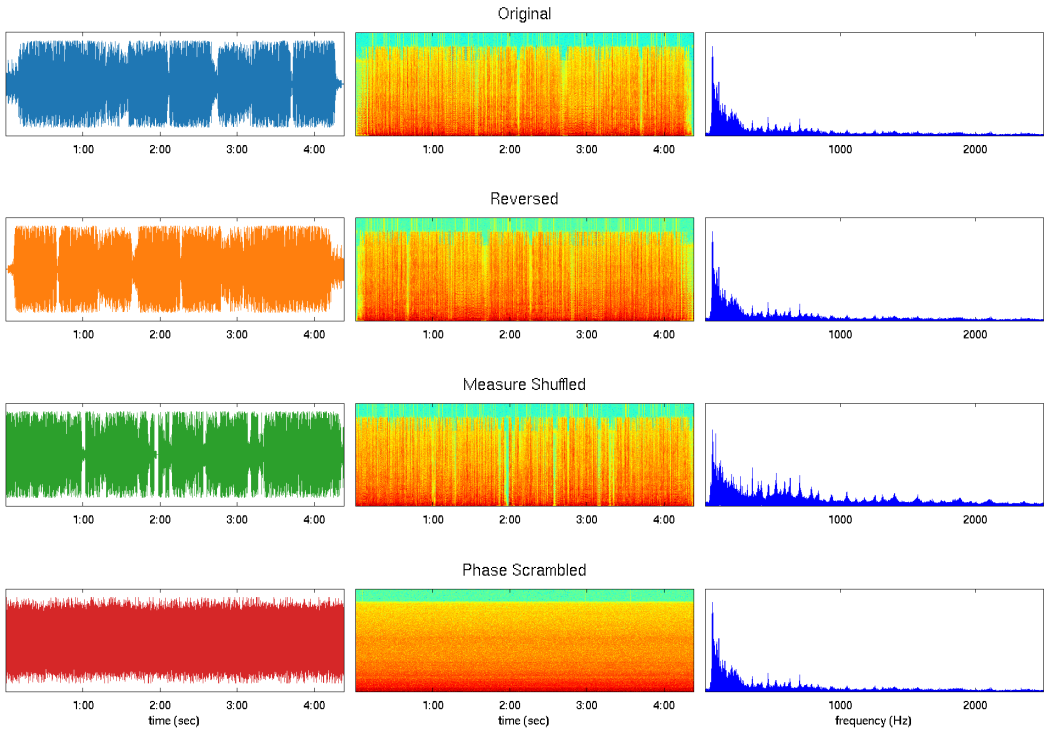

Figure A.2: Waveforms, spectrograms, and magnitude spectra of Song 3 stimuli.

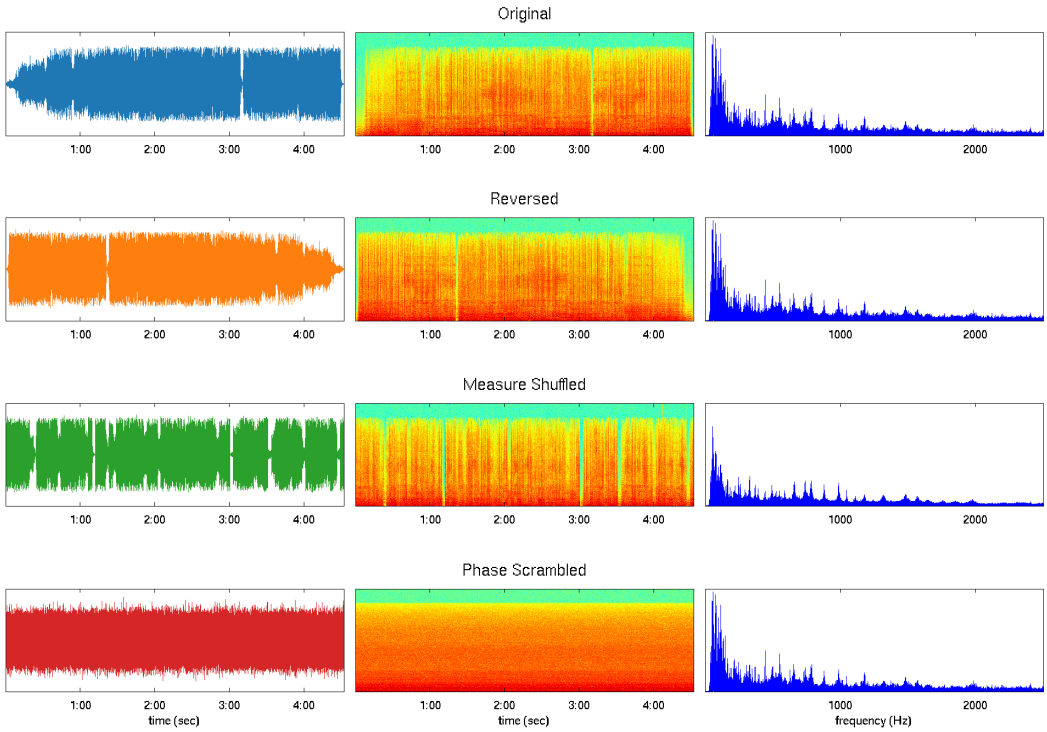

Figure A.3: Waveforms, spectrograms, and magnitude spectra of Song 4 stimuli.

A.2 Inter-Subject Correlations

A.2.1 RC1 and RC2 ISCs, Songs 2–4

Figure A.4: Time-resolved RC1 and RC2 ISCs for Song 2.

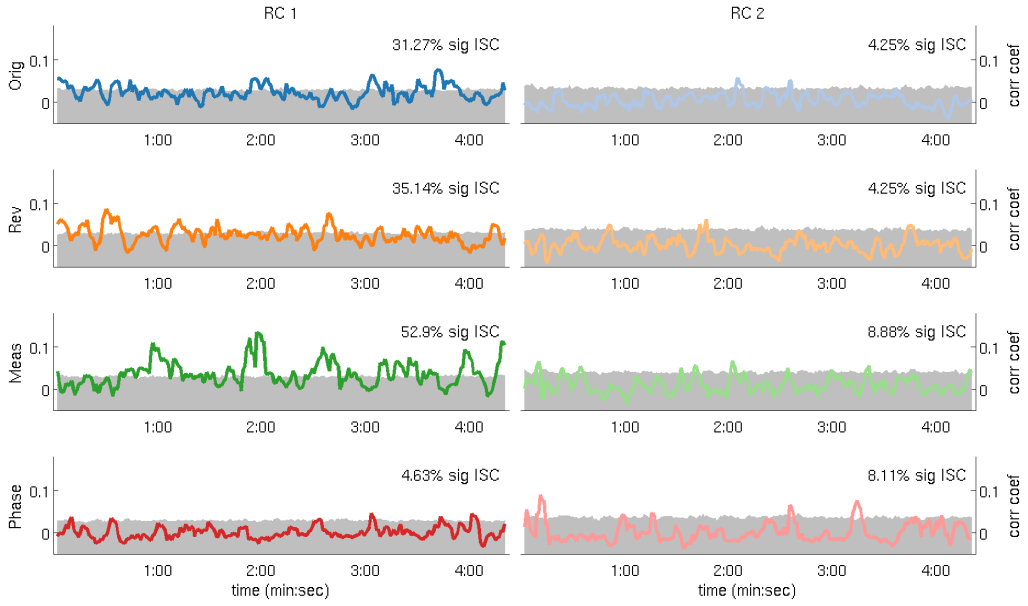

Figure A.5: Time-resolved RC1 and RC2 ISCs for Song 3.

Figure A.6: Time-resolved RC1 and RC2 ISCs for Song 4.

A.2.2 RC1 and RC2 ISCs for Manipulated Stimuli

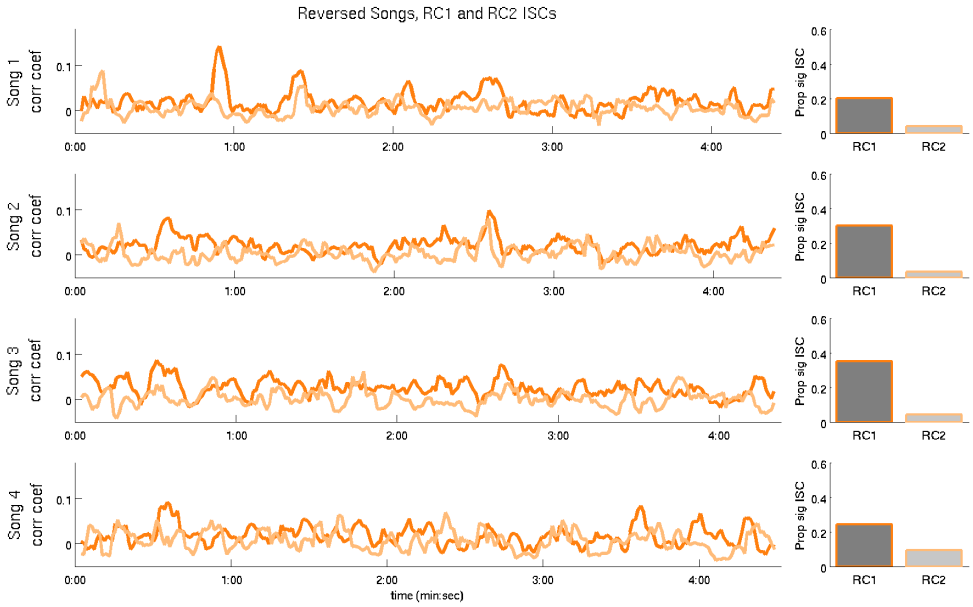

Figure A.7: Time-resolved RC1 and RC2 ISCs for all reversed songs.

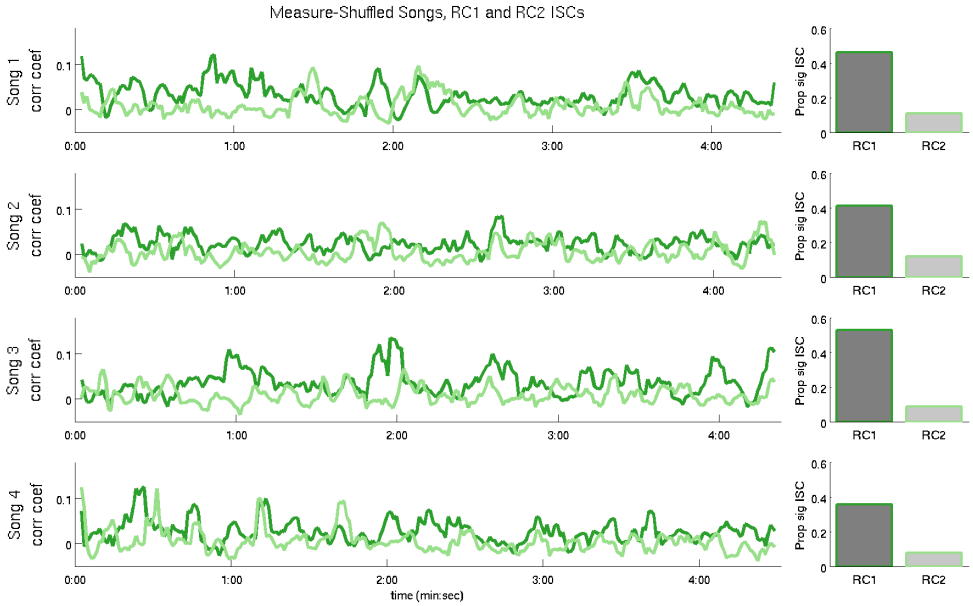

Figure A.8: Time-resolved RC1 and RC2 ISCs for all measure-shuffled songs.

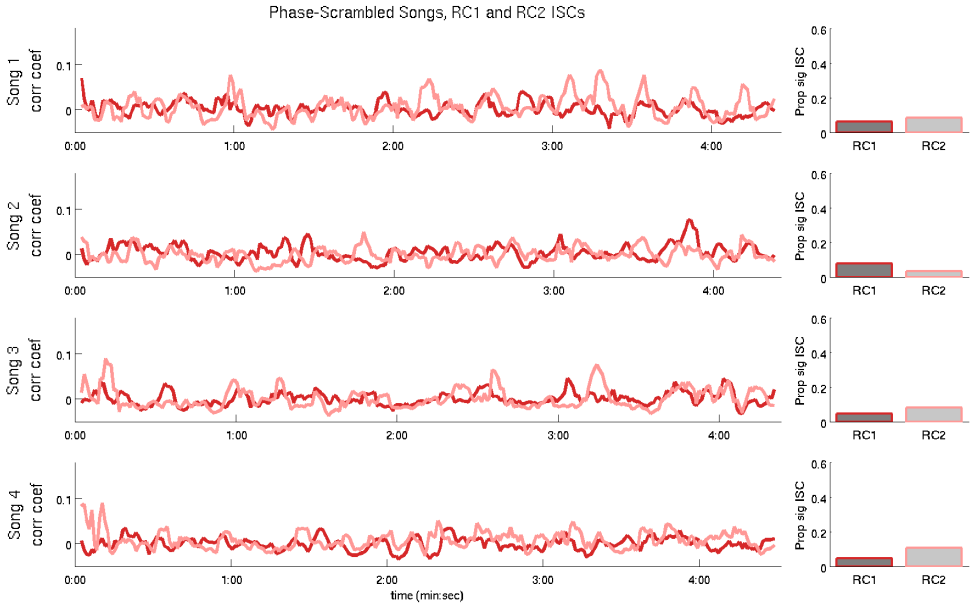

Figure A.9: Time-resolved RC1 and RC2 ISCs for all phase-scrambled songs. Note that the proportion of significant ISCs is not strictly lower for RC2, likely because the RC component weights used here did not correspond to those derived specifically in response to the phase-scrambled stimuli.

A.2.3 First- and Second-Listen RC1 ISCs

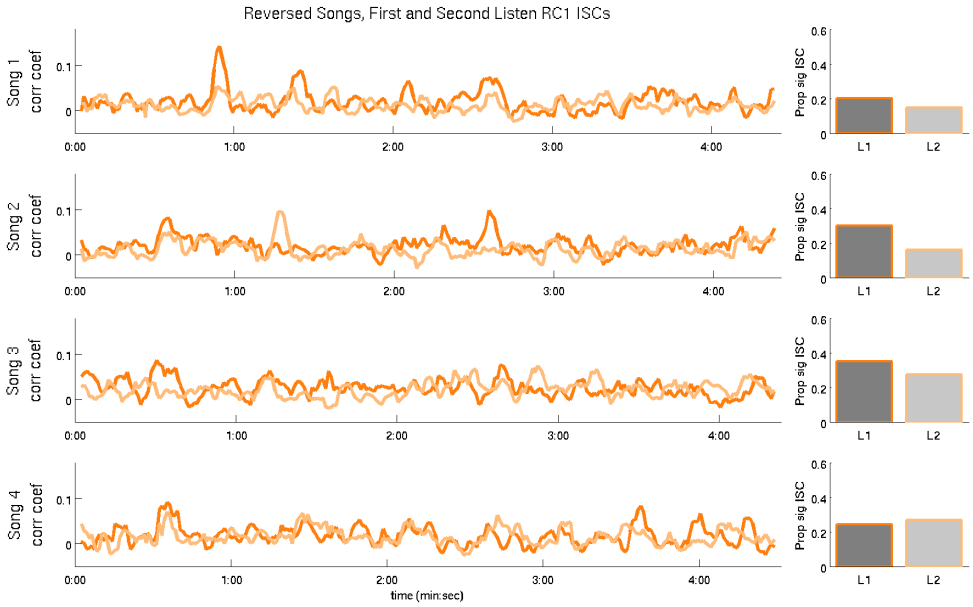

Figure A.10: RC1 ISCs of reversed stimuli, first versus second listen. The barplots on the right suggest that across the entire song, the proportion of significant ISCs is higher for the first listen than the second listen for the first three songs. Wilcoxon signed-rank tests on the difference of the ISC time series (Table 3.2) indicate that ISCs for this stimulus condition are significantly higher for all three of these songs.

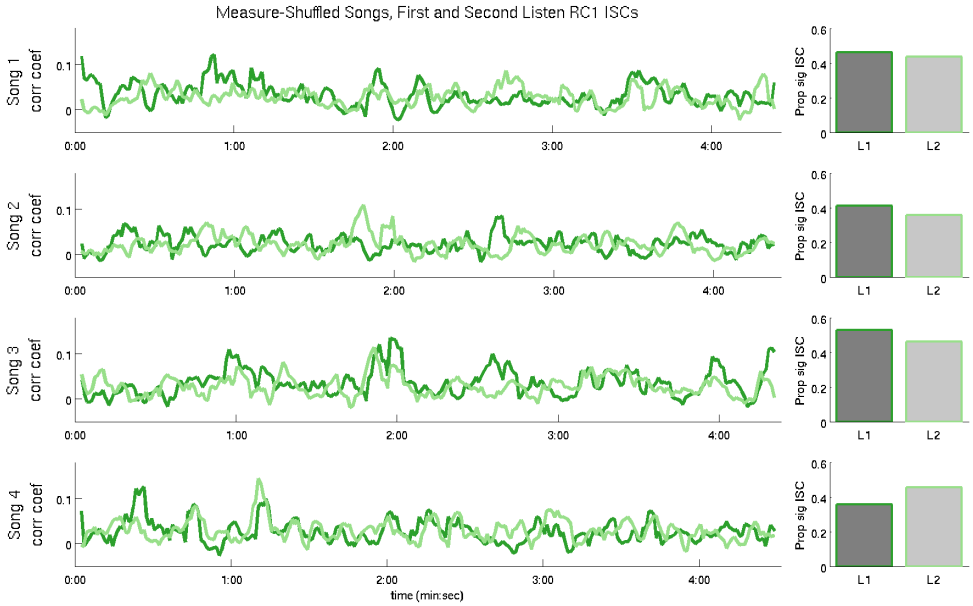

Figure A.11: RC1 ISCs of measure-shuffled stimuli, first versus second listen. While the first three songs show a lower proportion of significant ISCs over the second listen (right), the difference in proportions is statistically significant only for Songs 1 and 3 (Table 3.2).

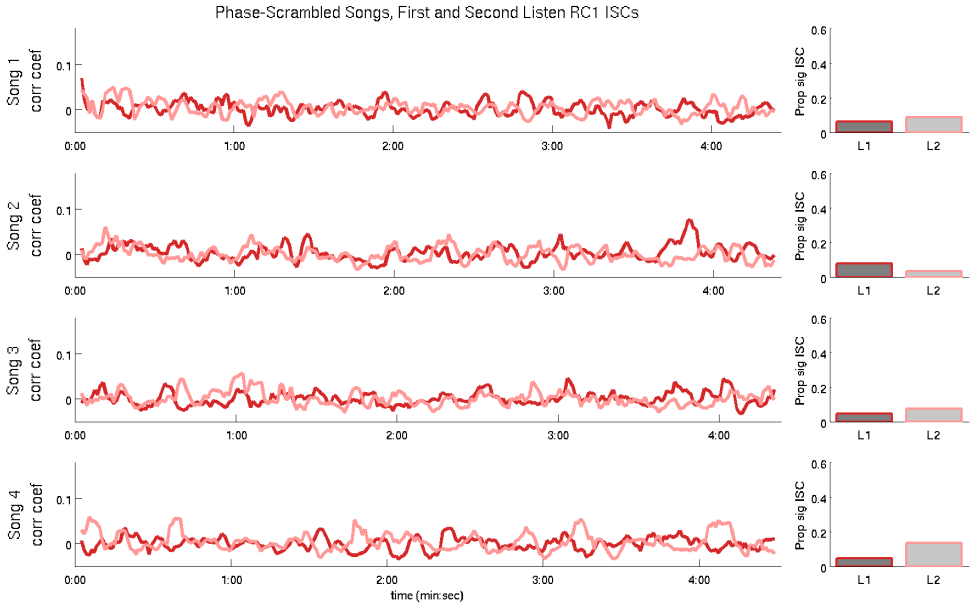

Figure A.12: RC1 ISCs of phase-scrambled stimuli, first versus second listen. Song 2, the only song for which the proportion of significant ISCs across the song is lower for the second listen (right) has only a marginally significant drop in ISCs from the first to the second listen (Table 3.2).

A.2.4 ISC-Amplitude Envelope Plots

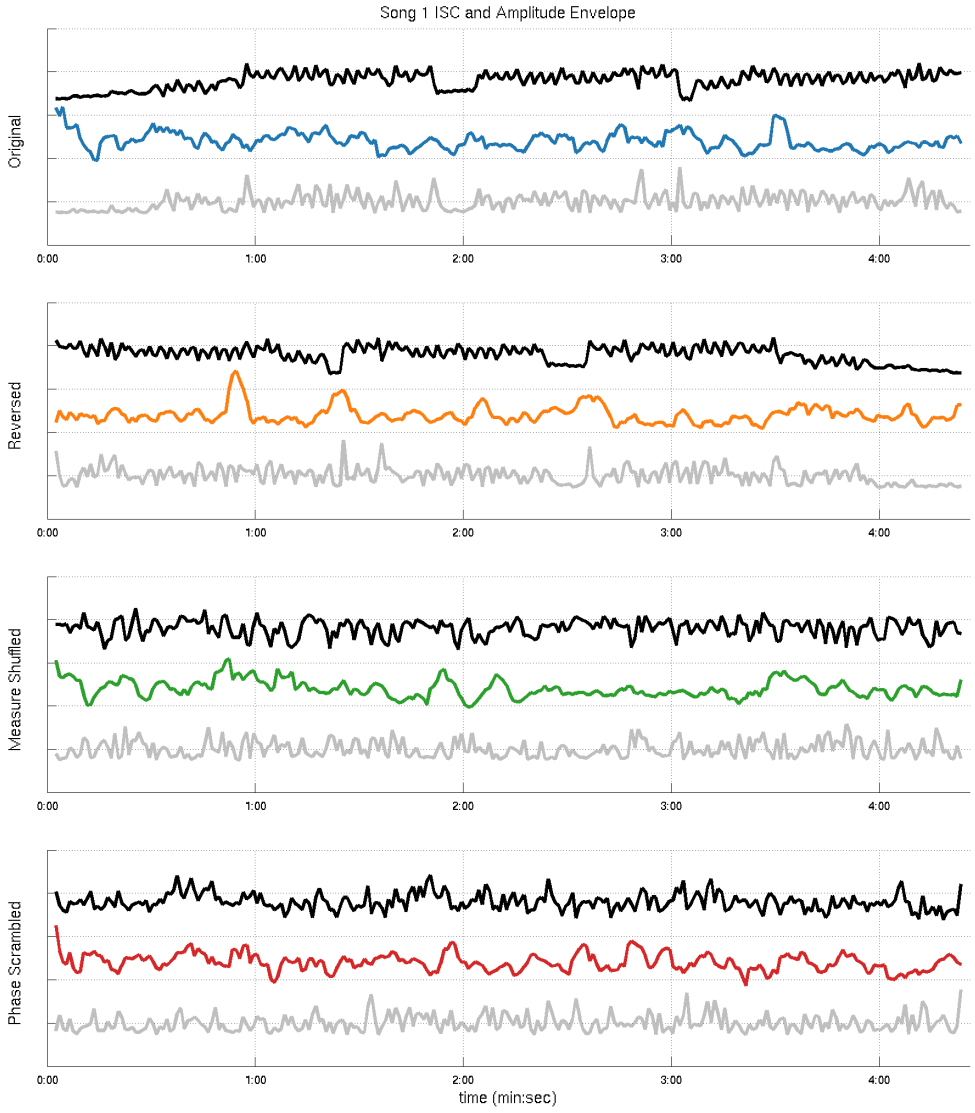

Figure A.13: RC1 ISCs of Song 1 first-listen responses (color), plotted scale-free with stimulus amplitude envelopes (black) and rectified difference envelopes (gray). For this song, ISCs produced by the original (blue) and reversed (orange) versions are statistically significantly correlated with the amplitude envelope (Table 3.3).

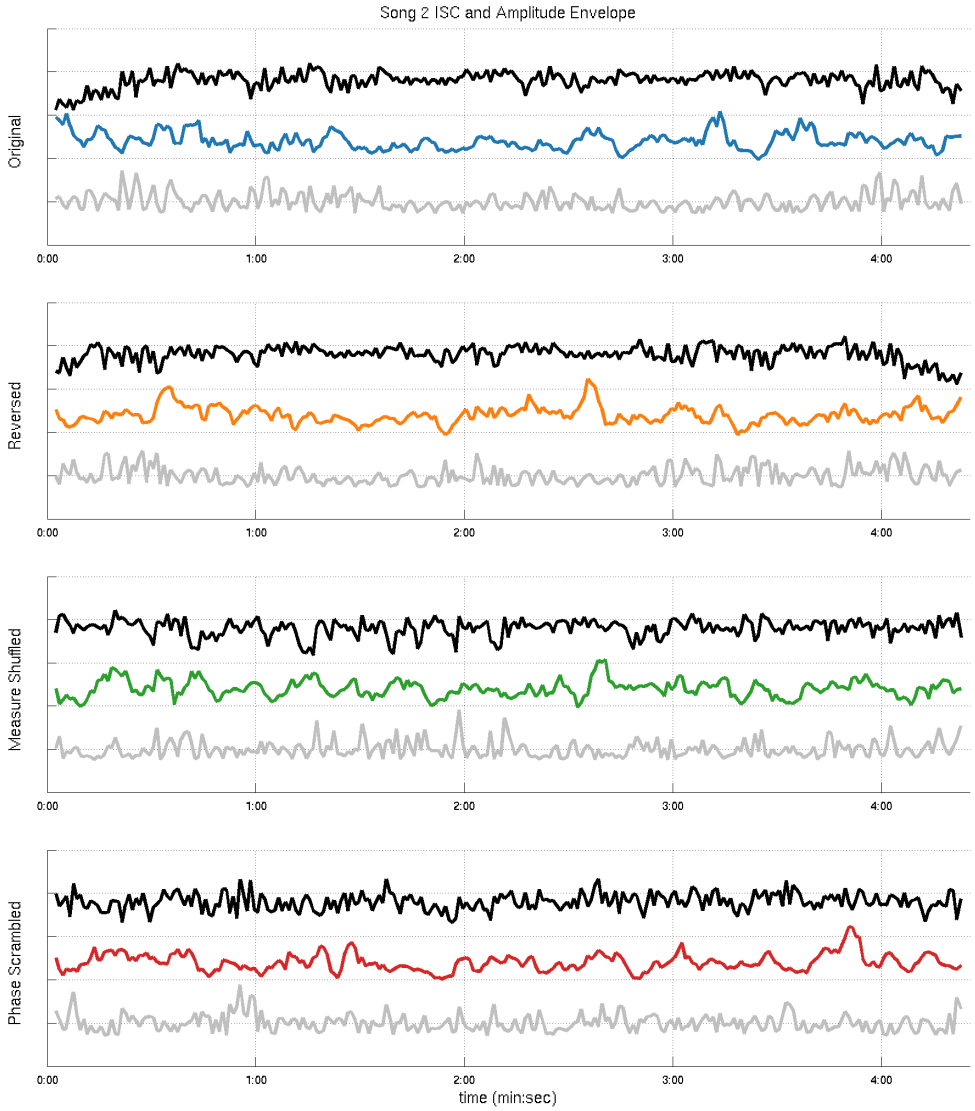

Figure A.14: RC1 ISCs of Song 2 first-listen responses (color), plotted scale-free with stimulus amplitude envelopes (black) and rectified difference envelopes (gray). Only the original version (blue) produces a statistically significant ISC correlation with the stimulus amplitude envelope (Table 3.3).

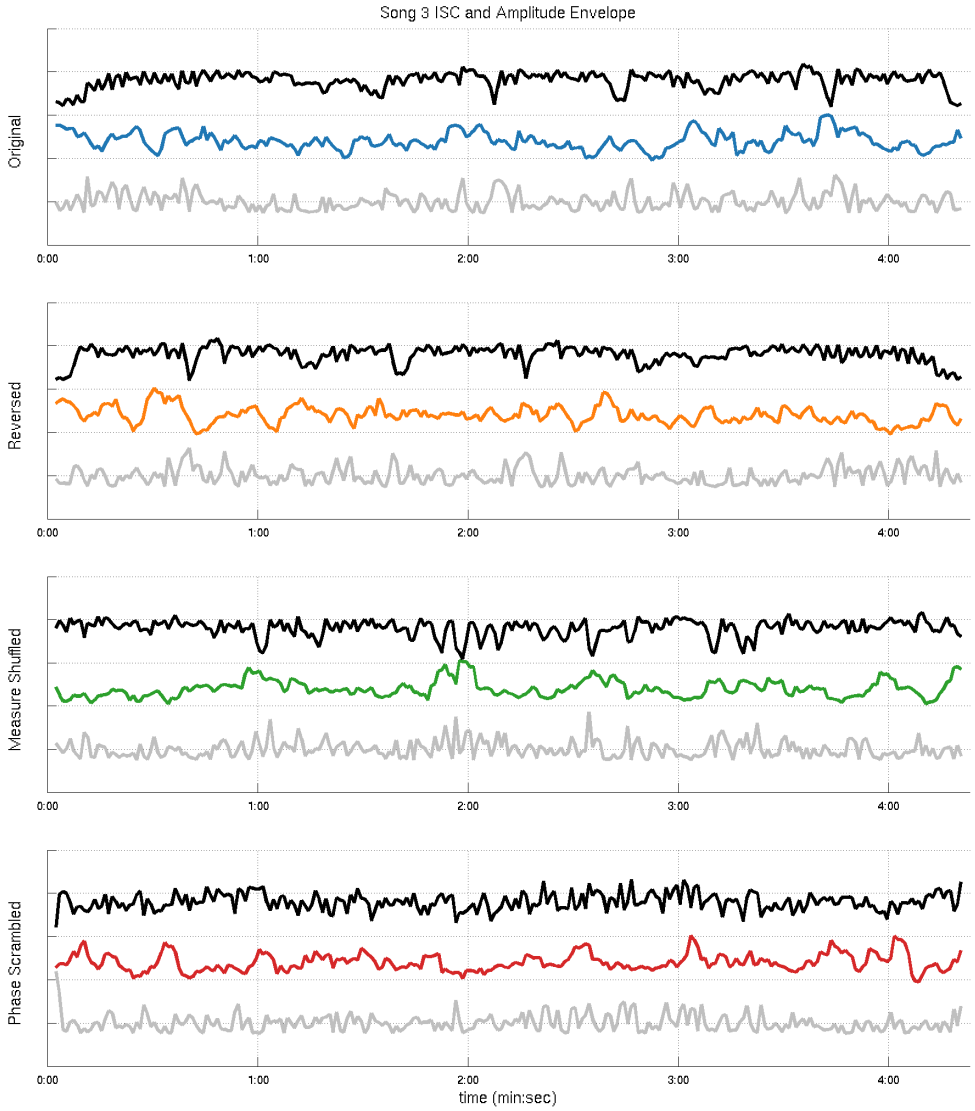

Figure A.15: RC1 ISCs of Song 3 first-listen responses (color), plotted scale-free with stimulus amplitude envelopes (black) and rectified difference envelopes (gray). The original (blue) and reversed (orange) versions of this song produce statistically significantly correlations between the ISC time series and amplitude envelope (Table 3.3).

Bibliography

D. A. Abrams, S. Ryali, T. Chen, P. Chordia, A. Khouzam, D. J. Levitin, and V. Menon. Inter-subject synchronization of brain responses during natural music listening. *The European Journal of Neuroscience*, 37(9):1458—1469, 2013. doi: 10.1111/ejn.12173.

V. K. Agawu. Structural 'highpoints' in Schumann's 'Dichterliebe'. *Music Analysis*, 3(2): 159–180, 1984.

V. Alluri and P. Toiviainen. Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception: An Interdisciplinary Journal*, 27(3):223–242, 2010.

V. Alluri, P. Toiviainen, I. P. Jääskeläinen, E. Glerean, M. Sams, and E. Brattico. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, 59(4):3677–3689, 2012. doi: http://dx.doi.org/10.1016/j.neuroimage.2011.11.019.

V. Alluri, P. Toiviainen, T. E. Lund, M. Wallentin, P. Vuust, A. K. Nandi, T. Ristaniemi, and E. Brattico. From Vivaldi to Beatles and back: Predicting lateralized brain responses to music. *NeuroImage*, 83(0):627–636, 2013. doi: http://dx.doi.org/10.1016/j.neuroimage.2013.06.064.

D. Ariely and G. S. Berns. Neuromarketing: The hope and hype of neuroimaging in business. *Nature Reviews Neuroscience*, 11(4):284–292, 2010.

J. A. Barraza, V. Alexander, L. E. Beavin, E. T. Terris, and P. J. Zak. The heart of the story: Peripheral physiology during narrative exposure predicts charitable giving. *Biological Psychology*, 105:138–143, 2015. doi: http://dx.doi.org/10.1016/j.biopsycho.2015.01.008.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

A. Ben-Yakov, C. J. Honey, Y. Lerner, and U. Hasson. Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. *NeuroImage*, 63(1):501–506, 2012. doi: 10.1016/j.neuroimage.2012.07.008.

Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, 2001.

B. Blankertz, G. Curio, and K. R. Müller. Classifying single trial EEG: Towards brain computer interfacing. In *Advances in Neural Information Processing Systems*, pages 157–164, 2002.

B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 25(1): 41–56, 2008. doi: 10.1109/MSP.2008.4408441.

B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. R. Müller. Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage*, 56(2):814–825, 2011. doi: http://dx.doi.org/10.1016/j.neuroimage.2010.06.048.

B. K. Bracken, V. Alexander, P. J. Zak, V. Romero, and J. A. Barraza. Physiological synchronization is associated with narrative emotionality and subsequent behavioral response. In *Foundations of Augmented Cognition. Advancing Human Performance and Decision-Making through Adaptive Systems: 8th International Conference, AC 2014*, pages 3–13. Springer International Publishing, 2014. doi: 10.1007/978-3-319-07527-3_1.

D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10(4):433–436, 1997.

R. Busselle and H. Bilandzic. Measuring narrative engagement. *Media Psychology*, 12(4): 321–347, 2009.

T. Chin and N. S. Rickard. The Music USE (MUSE) questionnaire: An instrument to measure engagement in music. *Music Perception: An Interdisciplinary Journal*, 29(4): 429–446, 2012.

M. X. Cohen. *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press, Cambridge, MA, 2014.

F. Cong, V. Alluri, A. K. Nandi, P. Toiviainen, R. Fa, B. Abu-Jamous, L. Gong, B. G. W. Craenen, H. Poikonen, M. Huotilainen, and T. Ristaniemi. Linking brain responses to

naturalistic music through analysis of ongoing EEG and stimulus features. *IEEE Transactions on Multimedia*, 15(5):1060–1069, 2013. doi: 10.1109/TMM.2013.2253452.

G. Cui, S. Gopalan, T. Yamamoto, J. Berger, P. G. Maxim, and P. J. Keall. Commissioning and quality assurance for a respiratory training system based on audiovisual biofeedback. *Journal of Applied Clinical Medical Physics / American College of Medical Physics*, 11 (4):3262, 2010.

S. J. Cunningham, D. Bainbridge, and D. McKay. Finding new music: A diary study of everyday encounters with novel songs. In *Proceedings of the 8th International Conference on Music Information Retrieval*, pages 83–88, 2007.

J. Davies. *Riveted: The Science of Why Jokes Make Us Laugh, Movies Make Us Cry, and Religion Makes Us Feel One with the Universe*. Palgrave Macmillan, New York, 2014.

M. H. DeGroot and M. J. Schervish. *Probability and Statistics*. Addison Wesley, Boston, third edition, 2002.

A. Delorme and S. Makeig. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004. doi: http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

S. Dikker, L. J. Silbert, U. Hasson, and J. D. Zevin. On the same wavelength: Predictable language enhances speaker-listener brain-to-brain synchrony in posterior superior temporal gyrus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 34(18):6267–6272, 2014.

J. P. Dmochowski, P. Sajda, J. Dias, and L. C. Parra. Correlated components of ongoing EEG point to emotionally laden attention—a possible marker of engagement? *Frontiers in Human Neuroscience*, 6:112, 2012. doi: 10.3389/fnhum.2012.00112.

J. P. Dmochowski, M. A. Bezdek, B. P. Abelson, J. S. Johnson, E. H. Schumacher, and L. C. Parra. Audience preferences are predicted by temporal reliability of neural processing. *Nature communications*, 5:4567, 2014. doi: 10.1038/ncomms5567.

J. P. Dmochowski, A. S. Greaves, and A. M. Norcia. Maximally reliable spatial filtering of steady state visual evoked potentials. *NeuroImage*, 109:63–72, 2015. doi: 10.1016/j.neuroimage.2014.12.078.

H. Egermann, M. T. Pearce, G. A. Wiggins, and S. McAdams. Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience*, 13(3):533–553, 2013. doi: 10.3758/s13415-013-0161-y.

D. P. W. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.

J. E. Escalas. Imagine yourself in the product: Mental simulation, narrative transportation, and persuasion. *Journal of Advertising*, 33(2):37–48, 2004.

J. E. Escalas. Self-referencing and persuasion: Narrative transportation versus analytical elaboration. *Journal of Consumer Research*, 33(4):421–429, 2007.

E. B. Falk, E. T. Berkman, and M. D. Lieberman. From neural responses to population behavior: Neural focus group predicts population-level media effects. *Psychological Science*, 23(5):439–445, 2012. doi: 10.1177/0956797611434964.

M. M. Farbood, D. J. Heeger, G. Marcus, U. Hasson, and Y. Lerner. The neural processing of hierarchical structure in music and speech at different timescales. *Frontiers in Neuroscience*, 9:157, 2015. doi: 10.3389/fnins.2015.00157.

R. A. Fisher. The design of experiments. Technical report, New York, 1971.

P. Gomez and B. Danuser. Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, 53(2):91–103, 2004. doi: http://dx.doi.org/10.1016/j.ijpsycho.2004.02.002.

P. Gomez and B. Danuser. Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, 7(2):377–387, 2007. doi: 10.1037/1528-3542.7.2.377.

M. C. Green and T. C. Brock. The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5):701, 2000.

M. C. Green, T. C. Brock, and G. F. Kaufman. Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, 14(4):311–327, 2004.

D. Gregory. Using computers to measure continuous music responses. *Psychomusicology*, 8 (2):127–134, 1989.

D. Gregory. Research note: The continuous response digital interface: An analysis of reliability measures. *Psychomusicology*, 14:197, 1995.

O. Grewe, F. Nagel, R. Kopiez, and E. Altenmüller. How does music arouse "chills"? *Annals of the New York Academy of Sciences*, 1060(1):446–449, 2005. doi: 10.1196/annals.1360.041.

O. Grewe, F. Nagel, R. Kopiez, and E. Altenmüller. Emotions over time: Synchronicity and development of subjective, physiological, and facial affective reactions to music. *Emotion*, 7(4):774–788, 2007a. doi: 10.1037/1528-3542.7.4.774.

O. Grewe, F. Nagel, R. Kopiez, and E. Altenmüller. Listening to music as a re-creative process: Physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Perception*, 24(3):297–314, 2007b.

O. Grewe, R. Kopiez, and E. Altenmüller. The chill parameter: Goose bumps and shivers as promising measures in emotion research. *Music Perception: An Interdisciplinary Journal*, 27(1):61–74, 2009. doi: 10.1525/mp.2009.27.1.61.

O. Grewe, B. Katzur, R. Kopiez, and E. Altenmüller. Chills in different sensory domains: Frisson elicited by acoustical, visual, tactile and gustatory stimuli. *Psychology of Music*, 2010. doi: 10.1177/0305735610362950.

D. M. Groppe, S. Makeig, and M. Kutas. Identifying reliable independent components via split-half comparisons. *NeuroImage*, 45(4):1199–1211, 2009. doi: 10.1016/j.neuroimage.2008.12.038.

F. Haas, S. Distenfeld, and K. Axen. Effects of perceived musical rhythm on respiratory pattern. *Journal of Applied Physiology*, 61(3):1185–1191, 1986.

D. J. Hargreaves. The effects of repetition on liking for music. *Journal of Research in Music Education*, 32(1):35–47, 1984.

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

U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, and R. Malach. Intersubject synchronization of cortical activity during natural vision. *Science*, 303(5664):1634–1640, 2004. doi: 10.1126/science.1089506.

U. Hasson, O. Furman, D. Clark, Y. Dudai, and L. Davachi. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron*, 57(3): 452–462, 2008a. doi: 10.1016/j.neuron.2007.12.009.

U. Hasson, O. Landesman, B. Knappmeyer, I. Vallines, N. Rubin, and D. J. Heeger. Neurocinematics: The neuroscience of film. *Projections*, 2(1):1–26, 2008b. doi: doi: 10.3167/proj.2008.020102.

U. Hasson, E. Yang, I. Vallines, D. J. Heeger, and N. Rubin. A hierarchy of temporal receptive windows in human cortex. *The Journal of Neuroscience*, 28(10):2539–2550, 2008c. doi: 10.1523/JNEUROSCI.5487-07.2008.

U. Hasson, G. Avidan, H. Gelbard, I. Vallines, M. Harel, N. Minshew, and M. Behrmann. Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Research: Official Journal of the International Society for Autism Research*, 2(4):220–231, 2009.

J. Haueisen and T. R. Knösche. Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, 13(6):786–792, 2001. doi: 10.1162/08989290152541449.

S. Haufe, S. Dähne, and V. V. Nikulin. Dimensionality reduction for the analysis of brain oscillations. *NeuroImage*, 101:583–597, 2014. doi: http://dx.doi.org/10.1016/j.neuroimage.2014.06.073.

A. Herbec, J.-P. Kauppi, C. Jola, J. Tohka, and F. E. Pollick. Differences in fMRI intersubject correlation while viewing unedited and edited videos of dance performance. *Cortex*, 71:341–348, 2015.

C. J. Honey, C. R. Thompson, Y. Lerner, and U. Hasson. Not lost in translation: Neural responses shared across languages. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(44):15277–15283, 2012.

H. Honing. Lure(d) into listening: The potential of cognition-based music information retrieval. *Empirical Musicology Review*, 2010.

P. Janata. ERP measures assay the degree of expectancy violation of harmonic contexts in

music. *Journal of Cognitive Neuroscience*, 7(2):153–164, 1995. doi: 10.1162/jocn.1995.7.2.153.

P. Janata. The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, (bhp008), 2009.

P. Janata, S. T. Tomic, and S. K. Rakowski. Characterisation of music-evoked autobiographical memories. *Memory*, 15(8):845–860, 2007. doi: 10.1080/09658210701734593.

C. Jola, P. McAleer, M.-H. Grosbras, S. A. Love, G. Morison, and F. E. Pollick. Uni- and multisensory brain areas are synchronised across spectators when watching unedited dance recordings. *i-Perception*, 4(4):265–284, 2013.

M. L. A. Jongsma, P. Desain, and H. Honing. Rhythmic context influences the auditory evoked potentials of musicians and nonmusicians. *Biological Psychology*, 66(2):129–152, 2004. doi: http://dx.doi.org/10.1016/j.biopsycho.2003.10.002.

T.-P. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, and T. J. Sejnowski. Extended ICA removes artifacts from electroencephalographic recordings. *Advances in Neural Information Processing Systems*, pages 894–900, 1998.

B. Kaneshiro and J. P. Dmochowski. Neuroimaging methods for music information retrieval: Current findings and future prospects. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*, pages 538–544, 2015.

B. Kaneshiro, J. Berger, M. Perreau Guimaraes, and P. Suppes. An exploration of tonal expectation using single-trial EEG classification. In *Proceedings of the 12th International Conference on Music Perception and Cognition*, pages 509–515, 2012.

B. Kaneshiro, J. P. Dmochowski, A. M. Norcia, and J. Berger. Toward an objective measure of listener engagement with natural music using inter-subject EEG correlation. In *Proceedings of the 13th International Conference on Music Perception and Cognition*, 2014.

B. Kaneshiro, D. T. Nguyen, J. P. Dmochowski, A. M. Norcia, and J. Berger. Naturalistic music EEG dataset—Hindi (NMED-H). In *Stanford Digital Repository*, 2016a. URL `http://purl.stanford.edu/sd922db3535`.

B. Kaneshiro, D. T. Nguyen, J. P. Dmochowski, A. M. Norcia, and J. Berger. Neurophysiological and behavioral measures of musical engagement. In *Proceedings of the 14th International Conference on Music Perception and Cognition*, 2016b.

S. Khalfa, I. Peretz, J.-P. Blondin, and M. Robert. Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, 328(2):145–149, 2002. doi: http://dx.doi.org/10.1016/S0304-3940(02)00462-7.

J. Kim and E. André. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2067–2083, 2008.

S. Koelsch. Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, 46(1):179–190, 2009. doi: 10.1111/j.1469-8986.2008.00752.x.

S. Koelsch, S. Jentschke, D. Sammler, and D. Mietchen. Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology*, 44(3):476–490, 2007. doi: 10.1111/j.1469-8986.2007.00517.x.

S. Koelsch, S. Kilches, N. Steinbeis, and S. Schelinski. Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLoS ONE*, 3(7): e2631, 2008. doi: 10.1371/journal.pone.0002631.

Z. J. Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalography and Clinical Neurophysiology*, 79(6): 440–447, 1991. doi: http://dx.doi.org/10.1016/0013-4694(91)90163-X.

C. L. Krumhansl. A perceptual analysis of Mozart's Piano Sonata K 282: Segmentation, tension, and musical ideas. *Music Perception: An Interdisciplinary Journal*, 13(3):401–432, 1996. doi: 10.2307/40286177.

C. L. Krumhansl. An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, 51(4):336–353, 1997.

E. Labbé, N. Schmidt, J. Babin, and M. Pharr. Coping with stress: The effectiveness of different types of music. *Applied Psychophysiology and Biofeedback*, 32(3–4):163–168, 2007. doi: 10.1007/s10484-007-9043-9.

A. Laplante and J. S. Downie. The utilitarian and hedonic outcomes of music information-seeking in everyday life. *Library & Information Science Research*, 33(3):202–210, 2011. doi: http://dx.doi.org/10.1016/j.lisr.2010.11.002.

O. Lartillot and P. Toiviainen. A Matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects*, pages 237–244, 2007.

E. L. Lehmann. *Nonparametrics: Statistical Methods Based on Ranks*. Springer, revised edition, 2006.

S. Leino, E. Brattico, M. Tervaniemi, and P. Vuust. Representation of harmony rules in the human brain: Further evidence from event-related potentials. *Brain Research*, 1142: 169–177, 2007. doi: 10.1016/j.brainres.2007.01.049.

D. J. Levitin and V. Menon. Musical structure is processed in "language" areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, 20(4): 2142–2152, 2003. doi: 10.1016/j.neuroimage.2003.08.016.

M. Levy. Improving perceptual tempo estimation with crowd-sourced annotations. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pages 317–322, 2011. doi: 10.1109/EMBC.2014.6945093.

Y. P. Lin, J. R. Duann, W. Feng, J. H. Chen, and T. P. Jung. Revealing spatio-spectral electroencephalographic dynamics of musical mode and tempo perception by independent component analysis. *Journal of NeuroEngineering and Rehabilitation*, 11(1):18, 2014. doi: 10.1186/1743-0003-11-18.

A. J. Lonsdale and A. C. North. Why do we listen to music? A uses and gratifications analysis. *British Journal of Psychology*, 102(1):108–134, 2011. doi: 10.1348/000712610X506831.

P. Loui, H. C. Li, A. Hohmann, and G. Schlaug. Enhanced cortical connectivity in absolute pitch musicians: A model for local hyperconnectivity. *Journal of Cognitive Neuroscience*, 23(4):1015–1026, 2011.

L.-O. Lundqvist, F. Carlsson, P. Hilmersson, and P. Juslin. Emotional responses to music: Experience, expression, and physiology. *Psychology of Music*, 2009. doi: 10.1177/0305735607086048.

C. K. Madsen. Emotion versus tension in Haydn's Symphony No. 104 as measured by the two-dimensional continuous response digital interface. *Journal of Research in Music Education*, 46(4):546–554, 1998.

C. K. Madsen and J. M. Geringer. Differential patterns of music listening: Focus of attention of musicians versus nonmusicians. *Bulletin of the Council for Research in Music Education*, (105):45–57, 1990.

C. K. Madsen, R. V. Brittin, and D. A. Capperella-Sheldon. An empirical method for measuring the aesthetic experience to music. *Journal of Research in Music Education*, 41(1):57–69, 1993. doi: 10.2307/3345480.

S. McAdams, B. W. Vines, S. Vieillard, B. K. Smith, and R. Reynolds. Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception: An Interdisciplinary Journal*, 22(2):297–350, 2004. doi: 10.1525/mp.2004.22.2.297.

J. H. McDonald. *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, third edition, 2014.

V. Menon and D. J. Levitin. The rewards of music listening: Response and physiological connectivity of the mesolimbic system. *NeuroImage*, 28(1):175–184, 2005. doi: 10.1016/j.neuroimage.2005.05.053.

D. Moelants and M. F. McKinney. Tempo perception and musical content: What makes a piece fast, slow, or temporally ambiguous? In *Proceedings of the 8th International Conference on Music Perception and Cognition*, pages 558–562, 2004.

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

K. N. Olsen, R. T. Dean, and C. J. Stevens. A continuous measure of musical engagement contributes to prediction of perceived arousal and valence. *Psychomusicology: Music, Mind, and Brain*, 24(2):147, 2014.

C. Pantev, R. Oostenveld, A. Engelien, B. Ross, L. E. Roberts, and M. Hoke. Increased auditory cortical representation in musicians. *Nature*, 392(6678):811–814, 1998.

L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda. Recipes for the linear analysis of EEG. *NeuroImage*, 28(2):326–341, 2005. doi: 10.1016/j.neuroimage.2005.05.032.

M. L. Phares. Analysis of musical appreciation by means of the psychogalvanic reflex technique. *Journal of Experimental Psychology*, 17(1):119–140, 1934.

T. W. Picton. The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, 9(4):456–479, 1992.

C. Potes, A. Gunduz, P. Brunner, and G. Schalk. Dynamics of electrocorticographic (ECoG) activity in human temporal and frontal cortical areas during music listening. *NeuroImage*, 61(4):841–848, 2012. doi: http://dx.doi.org/10.1016/j.neuroimage.2012.04.022.

C. Potes, P. Brunner, A. Gunduz, R. T. Knight, and G. Schalk. Spatial and temporal relationships of electrocorticographic alpha and gamma activity during auditory processing. *NeuroImage*, 97:188–195, 2014. doi: http://dx.doi.org/10.1016/j.neuroimage.2014.04.045.

D. Prichard and J. Theiler. Generating surrogate data for time series with several simultaneously measured variables. *Physical Review Letters*, 73(7):951–954, 1994.

M. Regev, C. J. Honey, E. Simony, and U. Hasson. Selective and invariant neural responses to spoken and written narratives. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(40):15978–15988, 2013.

P. J. Rentfrow. The role of music in everyday life: Current directions in the social psychology of music. *Social and Personality Psychology Compass*, 6(5):402–416, 2012. doi: 10.1111/j.1751-9004.2012.00434.x.

N. S. Rickard. Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music*, 32(4):371–388, 2004. doi: 10.1177/0305735604046096.

F. A. Russo, N. N. Vempala, and G. M. Sandstrom. Predicting musically induced emotions from physiological inputs: Linear and neural network models. *Frontiers in Psychology*, 4:468, 2013. doi: 10.3389/fpsyg.2013.00468.

V. N. Salimpoor, M. Benovoy, G. Longo, J. R. Cooperstock, and R. J. Zatorre. The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS ONE*, 4(10):1–14, 2009. doi: 10.1371/journal.pone.0007487.

D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch. Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*, 44(2):293–304, 2007. doi: 10.1111/j.1469-8986.2007.00497.x.

R. S. Schaefer, J. Farquhar, Y. Blokland, M. Sadakata, and P. Desain. Name that tune: Decoding music from the listening brain. *NeuroImage*, 56(2):843–849, 2011. doi: http://dx.doi.org/10.1016/j.neuroimage.2010.05.084.

R. S. Schaefer, P. Desain, and J. Farquhar. Shared processing of perception and imagery of music in decomposed EEG. *NeuroImage*, 70:317–326, 2013. doi: http://dx.doi.org/10.1016/j.neuroimage.2012.12.064.

T. Schäfer, P. Sedlmeier, C. Städtler, and D. Huron. The psychological functions of music listening. *Frontiers in Psychology*, 4:511, 2013.

R. Schmälzle, F. E. K. Häcker, C. J. Honey, and U. Hasson. Engaged listeners: shared neural processing of powerful political speeches. *Social Cognitive and Affective Neuroscience*, 10 (8):1137–1143, 2015. doi: 10.1093/scan/nsu168.

E. Schubert. Modeling perceived emotion with continuous musical features. *Music Perception: An Interdisciplinary Journal*, 21(4):561–585, 2004. doi: 10.1525/mp.2004.21.4.561.

E. Schubert. Reliability issues regarding the beginning, middle and end of continuous emotion ratings to music. *Psychology of Music*, 41(3):350–371, 2013.

E. Schubert and W. Dunsmuir. Regression modelling continuous data in music psychology. *Music, Mind, and Science*, pages 298–352, 1999.

E. Schubert, K. Vincs, and C. J. Stevens. Identifying regions of good agreement among responders in engagement with a piece of live dance. *Empirical Studies of the Arts*, 31 (1):1–20, 2013.

I. H. Shin, J. Cha, G. W. Cheon, C. Lee, S. Y. Lee, H. J. Yoon, and H. C. Kim. Automatic stress-relieving music recommendation system based on photoplethysmography-derived heart rate variability analysis. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6402–6405, 2014. doi: 10.1109/EMBC.2014.6945093.

E. Skoe and N. Kraus. Auditory brain stem response to complex sounds: A tutorial. *Ear and Hearing*, 31(3):302–324, 2010. doi: 10.1097/aud.0b013e3181cdb272.

J. A. Sloboda, S. A. O'Neill, and A. Ivaldi. Functions of music in everyday life: An exploratory study using the experience sampling method. *Musicae Scientiae*, 5(1):9–32, 2001. doi: 10.1177/102986490100500102.

J. O. Smith. *Spectral Audio Signal Processing*. W3K Publishing, `http://books.w3k.org`, 2011. URL `https://ccrma.stanford.edu/~jos/sasp/DTFT_Real_Signals.html`. Accessed 30 May, 2016.

J. Solomon. Deconstructing the definitive recording: Elgar's Cello Concerto and the influence of Jacqueline du Pré. *Unpublished manuscript*, 2009. URL `http://people.csail.mit.edu/jsolomon/assets/dupre.pdf`.

D. Sridharan, D. J. Levitin, C. H. Chafe, J. Berger, and V. Menon. Neural dynamics of event segmentation in music: Converging evidence for dissociable ventral and dorsal networks. *Neuron*, 55(3):521–532, 2007. doi: http://dx.doi.org/10.1016/j.neuron.2007.07.003.

N. Steinbeis, S. Koelsch, and J. A. Sloboda. The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, 18(8):1380–1393, 2006. doi: 10.1162/jocn.2006.18.8.1380.

S. Stober, D. J. Cameron, and J. A. Grahn. Classifying EEG recordings of rhythm perception. In *Proceedings of the 15th International Society for Music Information Retrieval Conference*, pages 649–654, 2014.

S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn. Towards music imagery information retrieval: Introducing the OpenMIIR dataset of EEG recordings from music perception and imagination. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*, 2015.

I. Sturm, B. Blankertz, C. Potes, G. Schalk, and G. Curio. 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. doi: 10.3389/fnhum.2014.00798.

I. Sturm, S. Dähne, B. Blankertz, and G. Curio. Multi-variate EEG analysis as a novel tool to examine brain responses to naturalistic music stimuli. *PLoS ONE*, 10(10):e0141281, 2015. doi: 10.1371/journal.pone.0141281.

P. Toiviainen, V. Alluri, E. Brattico, M. Wallentin, and P. Vuust. Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data. *NeuroImage*, 88(0):170–180, 2014. doi: http://dx.doi.org/10.1016/j.neuroimage.2013.11.017.

M. S. Treder, H. Purwins, D. Miklody, I. Sturm, and B. Blankertz. Decoding auditory attention to instruments in polyphonic music using single-trial EEG classification. *Journal of Neural Engineering*, 11(2):026009, 2014. doi: 10.1088/1741-2560/11/2/026009.

W. Trost, S. Frühholz, T. Cochrane, Y. Cojan, and P. Vuilleumier. Temporal dynamics of musical emotions examined through intersubject synchrony of brain activity. *Social Cognitive and Affective Neuroscience*, 10(12):1705–1721, 2015.

C.-G. Tsai, R.-S. Chen, and T.-S. Tsai. The arousing and cathartic effects of popular heartbreak songs as revealed in the physiological responses of listeners. *Musicae Scientiae*, 2014. doi: 10.1177/1029864914542671.

D. M. Tucker. Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, 87(3):154–163, 1993. doi: http://dx.doi.org/10.1016/0013-4694(93)90121-B.

G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002. doi: 10.1109/TSA.2002.800560.

I. van den Bosch, V. Salimpoor, and R. J. Zatorre. Familiarity mediates the relationship between emotional arousal and pleasure during music listening. *Frontiers in Human Neuroscience*, 7(534), 2013. doi: 10.3389/fnhum.2013.00534.

R. J. Vlek, R. S. Schaefer, C. C. A. M. Gielen, J. D. R. Farquhar, and P. Desain. Sequenced subjective accents for brain-computer interfaces. *Journal of Neural Engineering*, 8(3):036002, 2011a. doi: 10.1088/1741-2560/8/3/036002.

R. J. Vlek, R. S. Schaefer, C. C. A. M. Gielen, J. D. R. Farquhar, and P. Desain. Shared mechanisms in perception and imagery of auditory accents. *Clinical Neurophysiology*, 122(8):1526–1532, 2011b. doi: http://dx.doi.org/10.1016/j.clinph.2011.01.042.

T. P. Zanto, J. S. Snyder, and E. W. Large. Neural correlates of rhythmic expectancy. *Advances in Cognitive Psychology*, 2(2–3):221–231, 2006.

G. H. Zimny and E. W. Weidenfeller. Effects of music upon GSR and heart-rate. *The American Journal of Psychology*, 76(2):311–314, 1963.