The Cognitive Organization of Rhythmic Sounds: Metric Influence on Temporal Order Acuity

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Brandon Tyler Paul

Graduate Program in Speech and Hearing Science

The Ohio State University

2012

Master's Examination Committee:

Dr. Lawerence L. Feth, Advisor

Dr. Per B. Sederberg
Copyrighted by

Brandon Tyler Paul

2012
Abstract

Meter, a hierarchical mental organization used to group temporal patterns of sound, presents a unique opportunity to study the dynamic cognitive structuring of auditory perception. Often discussed in alternating cycles of phenomenally strong and weak beats, meter appears to guide our attention to expected points in time. Here, we examine the possibility that improved auditory perception occurs at moments when events are most strongly anticipated, and also examine the effect of long-term training and experience that may covary with metric expectations.

Eleven recruited subjects (six musicians and five non-musicians) participated in a two part study. A preliminary study asked listeners to discriminate timing distances between brief pairs of click pulses. The study used an adaptive procedure to determine a stimulus level distance where performance was standardized across participants. Generally, performance rates for all subjects were consistent.

Each participant’s timing distance was used in the main study, where listeners heard equally-spaced sequences click pulses—one of which was temporally reversed in order. While imposing different metric organizations (i.e., hearing clicks groups of two or three),
participants were to identify the metric position in which the click pulse was reversed temporally. During this procedure, the electroencephalogram was recorded from participants’ scalps.

Behavioral results indicate that musicians did not perform better than non-musicians at the task, a surprising finding suggesting that long-term training does not affect top-down influence on temporal acuity. Another surprising finding is that the context of click pulses, which can be presented in normal or reversed order, had effects on judgment. Consistent with predictions, conditions with metric groups of three yielded a worse performance overall, suggesting ternary organization of sounds is a more difficult metric hierarchy to maintain. Finally, beat-based differences arose only in comparing weak beats of one metric condition to all other beats; although there was a main effect of reporting metric positions that temporal reversals occurred, support for the notion of beat-based differences that underlie varied levels of attentional focus is spurious.

In analyzing neural signals, preliminary analysis suggests that mentally-maintained meter that is imposed upon sequences with identical stimuli may be observable in the electroencephalogram. However, differences in musician neural responses appear to be highly individualized, and perhaps not observable in non-musicians. However, a comparison of strong and weak beats in neural waveforms for all participants does not show significant differences; thus, evidence for mental representation for meter cannot be reported.
Findings from the study are implicated in understanding the neural mechanics of perceiving and organizing large acoustic structures such as speech and music.
Acknowledgments

This document owes many people thanks, and ultimately reflects the true nature of the collaborative effort it takes to tackle a scientific problem. In looking back upon endless hours of typing, coding, analyzing, and troubleshooting to execute this research project, I realize I was only able to meet success by leaning on the shoulders of advisors, collaborators, assistants, and friends. I dedicate this manuscript to them.

I’d first and foremost like to extend my deepest appreciation and gratitude for my advisor, Dr. Lawrence Feth, and co-advisor and referee, Dr. Per Sederberg. Drs. Feth and Sederberg have provided invaluable guidance and feedback across tireless hours to help me complete the current study. I can only hope to pay forward the effort they have given me when I am in their position of mentorship.

For their assistance in the data collection process, I’d like to acknowledge the graduate students and research assistants in the OSU Computational Memory Laboratory, headed by Dr. Sederberg. I’d specifically like to thank postdoctoral researcher Dr. Troy Smith for his many hours of effort and help on technical aspects of the project; from corraling research assistants, coding and programming, up-keeping the EEG system, and data analysis. Along
with the Computational Memory Lab, I’d like to thank OSU Psychoacoustics Laboratory members Niall Klyn, Matthew Davis, and Yonghee Oh for their help.

I’d also like to thank members of the Cognitive and Systematic Musicology Laboratory for their help and feedback in the research design and conceptual approaches. Yuri Broze, Dan Shanahan, Katelyn Horn, Josh Albrecht, Gary Yim, and Joe Plazak have all been supportive of me in the highs and lows of completing this document, and their years of peership would not be traded for the world. Leading this group, and deserving all the more thanks, is Dr. David Huron. For several years Dr. Huron has been an advisor, collaborator, recommender, and mentor, exuding outstanding scholarship and experience to help me keep both big-picture ideas and methodological nuances in line to tell compelling stories about the relationships of listeners’ perceptions, cognitions, and emotions.

I’d like to finally thank all my friends and family, including my parents, grandparent, and brother for their support through my seven years at The Ohio State University. I’d especially like to highlight the contributions of my friend and roommate Kyle McNicholas for his help in understanding programming and coding, as well as a general tolerance for my insanity (evidenced by my continuing urge to roll around on his bedroom floor whining about whatever current plight was most poignant) as we both completed Masters theses in Spring 2012. To Kyle: good luck on your endeavors in Austin, and thank you for one of the craziest and best years of my life.
Vita

2010............................................................................B. Mus., The Ohio State University
2010 to present .......................................................Graduate Administrative Associate, The

Ohio State University

Publications


Fields of Study

Major Field: Speech and Hearing Science
Table of Contents

Abstract ................................................................................................................................. ii

Acknowledgments .................................................................................................................. v

Vita ........................................................................................................................................ vii

Table of Contents .................................................................................................................. ix

List of Figures ...................................................................................................................... xiii

Chapter 1: Introduction and Literature Review .................................................................. 1

   Conceptual Definitions: Rhythm, Tempo, Pulse, and Meter ............................................. 3

   Short Review of Behavioral Studies Concerning Meter .................................................... 8

   Dynamic Attending .......................................................................................................... 11

   Rhythm and Meter Perception and the Role of the Motor System .................................. 17

   The Correlates of Event-Related Potentials and Event-Related Fields to Meter ............ 20

   Introduction to Event-Related Potentials and Rhythm ..................................................... 20

   Differentiating Rhythm and Meter Processing using DC Shifts and Early-Latency
   Components ...................................................................................................................... 22
Implications and Conclusion........................................................................................................................ 95

Bibliography ....................................................................................................................................................... 96

Appendix A: Ollen Musical Sophistication Index ............................................................................................. 107

Appendix B: Proportion correct of behavioral responses for all experimental conditions.
........................................................................................................................................................................... 110
List of Figures

Figure 1. Schematic of example rhythm. ................................................................. 4
Figure 2. Musical analogues to Figure 1 in both binary and ternary meter. .............. 5
Figure 3. Combination of an internal, self-sustaining oscillation that directs an energy pulse (taken from Large & Jones, 1999, Figure 4, p. 126). ........................................... 12
Figure 4. Attentional pulses that correspond to a temporal structure of tones. (taken and modified from Jones et al., 2002, Figure 1, p. 314). .................................................. 13
Figure 5. Scatterplot of OMSI scores for all participants. ........................................ 63
Figure 6. Click pairs modeled after Ronken (1970). ............................................... 65
Figure 7. Stimuli used in the main experiment. ...................................................... 69
Figure 8. Overall behavioral performance of musicians and non-musicians............... 76
Figure 9. Proportion of correct responses for binary and ternary meter.................... 77
Figure 10. Proportion of correct responses for context conditions......................... 78
Figure 11. Proportion of correct responses on strong and weak beats in binary, and proportion.............................................................................................................. 79
Figure 12. ERP waveform for Subject 1 in the musician group examining strong and weak beats ........................................................................................................ 81
Figure 13. ERP waveform for Subject 2 in the musician group examining strong and weak beats.

Figure 14. ERP waveform for Subject 3 in the non-musician group examining strong and weak beats.

Figure 15. ERP waveform for Subject 4 in the non-musician group examining strong and weak beats.
Chapter 1: Introduction and Literature Review

When a person listens to speech or music, how exactly does their brain function to organize the time-based relationships of sounds that they are hearing? From note to note, beat to beat, word to word, phrase to phrase, it appears that the brain organizes these patterns of sound to understand—and even predict—temporal relationships of perceived acoustic events.

Although a considerable corpus of scientific literature in the last three decades has endeavored to understand how the brain is able to represent and track auditory events in time, the precise mechanisms are still unknown. Loosely, it is thought that distributed neural activity operating as dynamic, entrained oscillators help to focus attention to specific time points in a hierarchical manner. When perceived acoustic events coincide with predicted moments, a person’s perception of these events appears to change. It is thought that the predicted events are perceived with more salience, and thus encoded more strongly than when not predicted. From this basic notion, the current study seeks to extend an understanding of time-based predictions by examining higher-level cognitive processes that guide our perception of unfolding streams of sounds, and seeks to answer the following two questions: Is our temporal acuity improved when we can focus varying amounts of attention to specific time points through temporal predictions? If so, what neural activity underpins this perceptual modulation?
A project examining the effects of top-down hierarchical organization and long-term training on perception attempts to develop the current understanding of structuring temporally-patterned sounds. The project employs both behavioral and electrophysiological methods to correlate listener responses and brain activity when listening to rhythmic sequences.

Specifically, the phenomenon of meter, an attention-driven cognitive process that helps to structure sounds hierarchically, is used to help listeners direct their attention to specific time points that contain small acoustic differences. Musical training serves as an ideal model to contrast domain-specific experience in this modality, and is built into the study as a further correlate.

Broadly, the implications of this study aim to develop the understanding of junctures between auditory cognition and perception. As Iversen, Repp, & Patel (2009) note, the phenomenon of meter can be considered an ideal model for examining the dynamic organization of perception, as meter appears to utilize processes of attention, memory, and sensorimotor coupling (pg. 59). These mechanisms appear to be important for listeners’ interactions with speech and music.

The plan of the current study is summarized in a short overview: First, conceptual definitions and a comprehensive literature review reveal a story describing the mechanisms by which it is thought that we organize sequences of sounds, through both behavior and electrophysiology. This story, as will be seen, has several gaps that raise questions regarding
the perceptual changes related to endogenously-maintained metric organization and long-term training. To attempt to fill in these gaps, a detailed research methodology is motivated that compares behavioral responses to electroencephalographic recordings. After completion of the experiment, analysis of behavioral results follows, along with preliminary neural data accompanying behavioral findings. Finally, a discussion of results is presented alongside implications within speech and music.

Conceptual Definitions: Rhythm, Tempo, Pulse, and Meter

When discussing the ordering of events that occur in time, it is important to be concerned with the conceptual definitions of rhythm, tempo, pulse (or beat), and meter. These terms are used to describe the respective pattern, rate, regularity, and structure of events, and refer to either properties of external time structures or internal representations of these structures. Definitions of terms used here are adapted from McAuley (2009) and Clarke (1999). Rhythm is a broad term that refers to the overall serial pattern of events that occur in time. Rhythms contain events that have durations themselves and a duration that spaces each event. Depending on the discipline, rhythms can be used to describe time relationships for several phenomena. For example, circadian or marine tidal rhythms are examples of rhythms in biological organisms (Sharma, 2003; Fingerman, 1957) and rhythms can characterize the orbit of planets around a star (Williams, 1981).

Although rhythm is an encompassing term, the time scale of concern here is with sounds that
are sequenced in short durations that range from milliseconds to dozens of seconds. Such patterns of sounds exist as spoken language and music. As an example, refer to Figure 1 and its musical counterpart Figure 2. A sequence of short sounds is depicted as black boxes in a recurrent pattern. The rhythm itself is highlighted by a red box, and occurs as a cyclic pattern of three sounds (although the first grouping contains two onsets) with three short arbitrary durations.

Figure 1. Schematic of example rhythm.
Figure 2. Musical analogues to Figure 1 in both binary and ternary meter.

_Tempo_ refers to the rate at which rhythmic events occur. The rates of events can be discussed relatively, such as _fast_ or _slow_, or can be discussed as absolute values in milliseconds or seconds that occur between events. In Figure 1, tempo is determined by an inter-onset interval (IOI) that refers to the amount of time that elapses between each event in the sequence. The IOI is discussed as having a duration of 600 milliseconds (ms) between each
instance of the three-sound event, and also 1200 ms between the start and end of a new rhythmic cycle. In the musical example of Figure 2, tempo is discussed using the relative marking of *fast*, and is also denoted by a metronomic marking of 100 beats per minute (BPM). BPM refers to the duration of beats or pulses that are to occur in a span of one minute.

A *pulse*, synonymous to *beat*, refers to an underlying referent regularity within a rhythmic sequence. Phenomenally, a pulse is described as an accented periodic marker from a sequence of sounds to which one may be able to nod their head to clap their hands. Arising endogenously, a pulse is usually discussed as a perceived element of a rhythmic sequence, as pulses can coincide with sound onsets or silences. Perceived pulses often arise from listening to music and can also be extracted from types of speech, such as in poetry’s use of feet and meter. The pulse in Figure 1 is denoted by a sequence of green hash marks that point to the underlying regularity of the sound pattern, and in Figure 2 using the same method. Note that the pulse occurs when sounds do and do not arise. As pulses follow the temporal structure of sounds, they may also be characterized by their tempo.

Pulses can also occur on multiple time scales. A faster series of pulses underlies the example sound pattern in Figure 1, noted by nested hash marks. A slower series of pulses envelopes multiple fast pulses. *Meter* describes the hierarchy of pulses or beats that occur on multiple time scales that coincide. From the largest instance of pulse, smaller pulses are divisions and subdivisions of the nominal pulse. This hierarchy is outlined by a blue box. When instances
of pulses align in the hierarchy, a structure of phenomenally-experienced *strong* and *weak* beats emerges. Strong beats, associated with a higher number of pulse alignments, exhibit a greater *pulse salience*. Strong and weak beats are also accompanied by perceived loudness differences, with stronger beats experienced to be louder than weak beats. This phenomenon, known as *subjective accenting*, occurs even when the acoustical properties of events on each beat are identical (Parnutt, 1994). When listeners encounter the first audible event, they generally construe this instance as a strong beat unless otherwise instructed (Povel & Okkerman, 1981).

Often, the relationships of hierarchical levels are discussed as ratios. In the provided example, the level of the highest order pulse structure occurs at a 1:4 ratio to the nominal pulse, and a 2:1 ratio to the lowest hierarchical level. In the musical correlate, meter is denoted by a time signature of two over four. The two refers to the number of pulses that occur per measure. A measure refers loosely to the period of the meter, or the number of nominal pulses it takes to complete one cycle. The four refers to the *type of durated note that is to receive the pulse*, directing a listener or sight-reader to the level that is the nominal pulse. Meter is also thought to be multistable. Not only does it exist in hierarchies, but it can also exist as two different organizations (Repp, 2007). Pointing to Figure 2, this rhythmic sequence can be experienced by organizing the onsets in groups of three, and there are several other possibilities.
Although several studies have been conducted on perception, production, and cognition of rhythms, tempos, and underlying pulses (see Clarke, 1999; London, 2004; Repp 2005; Honing, in press, for reviews), the current focus is on meter. Generally, meter is considered an interpretation of a rhythmic sequence. It is a top-down organization of temporally-based information for the purposes of guiding action and generating predictions. Meter can arise from acoustic features of rhythmic sequences such as grouping, physical accents (differing amplitudes), or the coincidence of multiple events (Lerdahl & Jackendoff, 1983; Longuet-Higgins & Lee, 1984). Meter can also arise from imposed or imagined organizations, such as willfully hearing Example 2 in multistable cyclic structures of two or three. This creates a distinction of exogenously-induced meter versus meter induced endogenously, and this differentiation has been the principal focus underlying theories of meter for several decades.

Accordingly, music-theoretic accounts of meter highlight the interaction of contextually and physically based cues from rhythmic groupings alongside an abstraction of regularity; or more simply, the respective juncture of perceiving physical accents and mental schema that segment them in time (Palmer & Krumhansl, 1990). Some theories place emphasis on the role of the former, suggesting meter is an emergent process driven by regularities within sound sequences (e.g., Meyer, 1956), and other theories emphasize an imposition of regular time points that help to mark and group accented structures (e.g., Benjamin, 1984). It is clear
from music-theoretic work that a principle concern was differentiating bottom-up and top-down processes that contribute to meter, indirectly attempting to distinguish perceptual and cognitive processes. Although such a distinction is not dichotomous, these early accounts of meter generated interest in testing the predictions and assumptions of theoretical work using scientific approaches.

Empirical research on meter perception and induction support both aspects of bottom-up and top-down processes, and several examples of studies highlight the interaction. Palmer and Krumhansl (1990) demonstrated that metrical interpretations affect perception of temporal patterns, with recognition memory for events on metrically strong beats being improved compared to less salient beat positions. Large, Fink, and Kelso (2002) showed that when participants tap along to complex, unpredictable rhythmic sequences that invite multiple hierarchies of periodicity, participants reliably synchronize to structures that aren’t necessarily reflected in stimulus properties. In other words, participants’ intention of employing metric structures is similar and corroborates performance with stimuli that are simple and predictable. This suggests that an endogenous process, separate from learning based on temporal patterning of rhythms, is employed by listeners. Further sensorimotor synchronization research has demonstrated improved tapping stability when participants tap along with metrically-strong beats (Patel, Iverson, Chen, & Repp, 2005), and no difference in tapping stability was found when participants tapped to sequences with physical cues to induce meter versus imagined meter (Repp, Iverson, & Patel, 2009).
Perceiving physical properties of rhythmic structures influence metric interpretations as well, such as event duration, amplitude, frequency changes, and spectral changes (e.g., Windsor, 1993; Zimba & Robin, 1998). Drake, Penel, and Bigand (2000) demonstrated that expressive fluctuations in musical examples affect metric tapping periodicities within listeners, with subjects willfully jumping from hierarchical levels in order to guide perception of variable temporal sequences. Hannon, Snyder, Eerola, and Krumhansl (2004) provide evidence that larger musical structures, such as melodic contour and melodic repetition, create reliable cues in influencing meter. Thus, it appears that not only do basic acoustic features influence meter perception, but structural combinations of these cues arising as melody can also inform listeners’ metric organization.

Although it appears that listeners are able to use both exogenous cues and endogenous interpretations to impose meter, what mechanisms allow us to track, predict, and interact with rhythmic structures that contain hierarchical beat periodicities? More concisely, how do we perceive temporally-patterned events that are fundamental to establishing meter? Models that attempt to describe rhythmic tracking break down into interval and entrainment models. Interval models propose that time intervals are encoded as memory traces built from statistical learning (Drake & Botte, 1993), or from encoding time intervals that are then matched to an internally-represented grid (Povel & Essens, 1985). However, these models assume a fixed interval sequence of events and are not robust to temporal fluctuations that occur in time-based structures, such as expressive timing in music performance or pauses found in speech. Considering the necessity to be flexible, multistable, hierarchical, and partly
endogenous, a likely candidate for such mechanisms are entrainment models. Entrainment models feature internal oscillatory processes that react to changes in an external temporal structure. In other words, entrainment models are characterized by external time structures that “drive” internal rhythms. In turn, the internal rhythms help establish an anticipatory process that describe when an event is likely to occur. The following review focuses on such an entrainment model.

*Dynamic Attending*

One of the prevailing theories of perceiving temporally-spaced events is the theory of dynamic attending. In their pivotal paper, Large and Jones (1999) addressed two questions regarding how we track events that unfold in time: What properties of temporally-extended events help us to anticipate on focus on time points? Also, when temporal structure changes, how does our attention to specific points in time change to match new structures? Temporal structures are the time courses of events that the brain can process, such as moving images or rhythmically-presented sounds. The authors theorized that rhythms of attention help listeners to interact with structured temporal information, built from behavioral work by Jones (1976) and Jones and Boltz (1989). These rhythms are organized “pulses” of attentional focus, or energy that respond to both expected and unexpected temporal information that the brain processes.

These endogenous rhythms attempt to synchronize to external time structures. When these
rhythms do synchronize, they are said to *entrain*. Entrainment can also be described as an internal timekeeper that matches to external time events. Figure 3 depicts a schematic of internal rhythmic processes. Here, a sustained neural oscillation combines with an attentional pulse, and generates time points that are predicted to occur. Trains of these oscillatory pulses create attending rhythms (taken from Large & Jones, 1999, Figure 4, p. 126). Figure 4 (Jones et al., 2002) corresponds the schematic of internal oscillations to external time structures. Since the external time structure itself drives the attending from the bottom-up, the authors consider dynamic attending to be a low-level process.

![Figure 3. Combination of an internal, self-sustaining oscillation that directs an energy pulse](image)

(taken from Large & Jones, 1999, Figure 4, p. 126).
The attentional pulses that comprise internal rhythms are predicted to vary in their width based on properties of external rhythm structures. For example, if a rhythmic structure is isochronous and regular, the attentional focus is very narrow: Expected onsets of events in the structure are highly predictable and thus internal rhythms allocate attentional energy to a very small time window. Conversely, if a rhythmic sequence is highly variable with event-rate modulations and individual event time shifts, very wide attentional pulses spread allocated attention to larger ranges of time. As a result, the dynamic attending model
proposed by Large and Jones predicts increased sensitivity to times when attention is narrowly focused, and violations of expectations during these times will be more discriminable than when attention is less focused.

Moreover, these internal oscillations can arise as a complex hierarchy that can be sensitive to complex time structures found in speech and music. Several nested oscillations in a hierarchy can then generate small- and large-time scale expectancies that can guide anticipatory focus. An assumption about multiple oscillations relies on the establishment of a referent oscillation. This intermediate oscillation is thought to reflect the pulse extracted from a temporally-structured event (e.g., see the nominal pulse in Figure 1). Faster or slower oscillations in harmonic relationships (e.g., 2:1, 3:1) can generate expectancies on multiple time scales. In other words, oscillations that evenly subdivide larger oscillations “tune” to faster rates, and operate concurrently with larger oscillations that may encompass multiple periods of a referent oscillator per one oscillation. It is possible then to focally attend to different levels of related oscillations, and generate expectations on short, intermediate, and long term levels. Due to the hierarchical structuring of such anticipatory oscillations, dynamic attending is a plausible theory behind the emergence of meter.

To summarize, the dynamic attending theory utilizes the three entities of external rhythms, internal rhythms, and their interaction as entrainment, as a basis for tracking events that unfold in time. Internal rhythms are pulses of attentional energy that are determined by the regularity and predictability of external sequences. These internal rhythms interact to form
hierarchies of expectations on different time scales, and are governed by the complexities of external temporal structures. The dynamic attending process is thought to be driven from the stimuli in a bottom-up fashion, and then influenced by top-down mechanisms as willful shifts in attentional focus.

To test this theory, Large and Jones (1999) presented three experiments that incorporate timing perturbations in tone sequences to assess listeners’ responses to changes in temporal structures. An initial experiment found that isochronous sequences with small deviations resulted in high discrimination accuracy, but a paradoxical finding was that accuracy increased when deviant onset magnitudes where high. This finding contrasts the notion of sensitivity being highest in a time window with the highest concentration attentional energy—those pulses that are generated by very regular sequences and small timing deviations. The authors attempt to explain this finding by noting that two-alternative forced choice (2AFC) paradigms used in this experiment reward correct discrimination of the most different change in stimulus. Large and Jones argue that support for dynamic attending in this study comes from the significant interaction of isochronous rhythms with late, small magnitudes of deviant offset tones. To reconcile, a second study demonstrated that timing judgments were most accurate in contexts that provided expected onsets of tones, opposed to conditions where a tone was unexpected due to a late or early presentation. One final experiment supported the idea that two internal oscillations generate varying expectancies on multiple time scales (i.e., meter), generalizing the result to stimuli that resemble complex structures found in music and speech.
Several subsequent studies lend further evidence to the theory of dynamic attending. By following the design of experiment two in Large and Jones (1999), Barnes and Jones (2000) completed seven follow-up experiments. The authors utilized a time interval discrimination task while manipulating subdivisions of sequences, as well as expectancy of intervals preceding a judgment interval. Overall, Barnes and Jones found that a sequence of timing intervals helps to “tune” internal oscillations to particular interval durations, and this tuning helps to determine timing information on which comparisons can be made. Jones, Moynihan, MacKenzie, and Puente (2002) asked listeners to instead make pitch comparison judgments instead of comparisons of timing intervals, and found results consistent with timing judgments. Further work on dynamic attending shows that accuracy of time judgments and rhythm production also correlate with age and musical training (Drake, Jones, & Baruch, 2002).

In sum, dynamic attending theory posits that internal rhythms entrain to external stimuli help to generate expectations in time that direct attentional energy to given time points. The regularity of external time structures allows internal oscillatory rhythms to focus attention on a given time point, and irregularity tends to widen attentional focus. The attentional pulses are purported to increase sensitivity to stimulus properties used in both time and pitch perception tasks. Oscillatory processes generating attentional pulses can be structured in coinciding hierarchies, possibly accounting for the phenomenon of meter.
Rhythm and Meter Perception and the Role of the Motor System

Although the principle concern at present is in perceiving rhythm and meter, rhythm production has been a reliable method to elucidate underpinnings of meter and entrainment (Bispham, 2006; see Repp, 2005 for a review). However, using overt rhythm production as a proxy to examine perceptual underpinnings has poignant drawbacks. Errant stimulus-production phase relationships and error correction from feedback arising from tapping sounds may obscure true perceptual mechanisms (Merker et al., 2009). As a result, several studies have attempted to remove motor responding in order to more clearly outline perception-production boundaries. For example, the reviewed work of Mari Jones and colleagues employ perceptual tasks that support mechanisms of entrainment and attending while sidestepping a reliance on overt motor responses. In light of perception and production dissociation, recent evidence suggests that activity in motor regions of the brain—even when no overt movement is implicated in experimental tasks—are implicated in processing temporally extended sequences (Phillips-Silver, Aktipis, & Bryant, 2010). This suggests that clear distinctions in perceptual and production systems are not feasible, and that our ability to track and predict events in time is possibly locked to our ability to plan and execute movement.

Neuroimaging studies provide perhaps the strongest evidence for the influence of motor activation in the brain on rhythm perception. Grahn and Brett (2007) utilized functional magnetic resonance imaging (fMRI) to determine neuroanatomical correlates of rhythm
perception while subjects only listened to simple, complex, and irregular rhythmic sequences. They found that bilateral activation of supplemental motor areas (SMAs) along with the basal ganglia are correlated to presenting simple rhythmic sequences—those sequences where predictability is the highest. Moreover, musical training was correlated with activation of the premotor cortex (PMC) and cerebellum, and these regions were not significantly active in participants without musical training. A follow-up study by Grahn and Rowe (2009) suggested that rhythmic sequences with a clear pulse were correlated with stronger PMC-SMA-putamen connectivity, along with the auditory cortex. Meter-inducing accents also modulated connectivity between the auditory cortex and PMC, with this effect being higher in subjects with musical training. Fujioka, Trainor, Large, and Ross (2012) recorded magnetoencephalographic (MEG) signals while subjects listened to temporally-regular tone sequences. They found that dynamic cortical connectivity between auditory cortices and several motor regions is associated with time-locked beta-band oscillatory modulations, suggesting neural synchrony in these regions is implicated in predictive attending.

Parkinson’s disease has also been used to study the role of the motor system in rhythm production. Grahn and Brett (2009) played temporally regular and irregular rhythmic sequences to a sample of individuals with Parkinson’s disease, as well as a control group. Results from rhythmic reproduction tasks indicate that impairments of motor planning found in Parkinson’s disease due to dopaminergic dysfunction in the striatum also correlate negatively with reliable underlying pulse production to rhythmic sequences. Grahn and Brett’s findings suggest a role of the motor system in pulse perception of rhythmic
sequences. Similarly, meter identification in patients with *spina bifida meningomyelocele* (SBM) is impaired in contrast to healthy control subjects (Dennis et al., 2009). SBM is characterized by irregular volumetric structures in the cerebellum that are implicated in executing motor functions, supporting implications from Grahn and Brett (2009).

From the reviewed studies on mechanisms by which we track rhythmic sequences, it is suggested that the phenomenon of meter may be a convenient method to examine interactions between high and low level processing. Perceived differences in regularity of acoustic cues both inform and are organized by established mental hierarchies. Stemming from the idea that listeners track rhythmic sequences through dynamic, entrained oscillations that focus attention on predicted time points, hierarchies of oscillations in harmonic relationships perhaps allow us to perceive and produce rhythmic patterns. Meter appears to draw upon cognitive faculties of memory, attention, and movement in order to establish and maintain these hierarchies, and ultimately guide our perception of events in time through concerted organization. A pivotal question from these behavioral studies might be, are the results supported by experiments examining time courses of neural activity?
The Correlates of Event-Related Potentials and Event-Related Fields to Meter

Introduction to Event-Related Potentials and Rhythm

One possible way to examine how the brain organizes rhythmic sequences is through event-related potentials (ERPs) extracted from the scalp-recorded electroencephalogram (EEG) (Ford & Hillyard, 1981; Zatorre, Belin, & Penhune, 2002). When brain structures give rise to activity during specific computational processes, voltage changes that result from the activity can be recorded using an array of electrodes. The voltage changes that arise from the performed operations are referred to as ERPs (Luck, 2005). Although the latency and amplitudes of specific ERPs can change along with different conditions, underlying components are thought to represent the same cognitive or sensory process.

The benefit of recording EEG signals from the scalp rests in the technique’s high temporal resolution. Neural activity can be examined with precision down to the millisecond, making EEG recording of the most advantageous neuroimaging techniques for correlating behavioral events with the time course of dynamic neural activity. In contrast, the technique’s poor spatial resolution impairs localizing exactly what neuroanatomical region is generating these signals. Although some algorithms attempt to estimate the source of neural activity in conjunction with appropriate electrode placement, localizing a source of EEG signals is best when correlated fMRI (Michel et al., 2004).
Several studies have used ERPs to understand perceptual and cognitive processing of rhythm, and demonstrate reliable responses to both steady-state and dynamic rhythmic arrays. For example, omissions of expected onsets in rhythmic patterns have been shown to elicit early and late ERP responses (van Zuijen et al., 2004; Brochard, Abecasis, Potter, Ragot, & Drake, 2003), and long- and middle-latency responses to changes in rhythm appear to be rate-dependent (Snyder & Large, 2004). Changes in tempo can also elicit late-latency differences within subcomponents of a expectancy-related response (Jongsma, Meeuwissen, Vos, & Maes, 2007), and organizational strategies when listening to rhythms can differentiate evoked responses to anticipated onsets (Snyder, Pasinski, & McAuley, 2011).

Aside from rhythmic patterns and tempos, ERPs can also be used to examine neural correlates of meter. An important first step is to differentiate processing that occurs for both meter and for rhythm. Although several studies have shown behavioral differences in perceiving meter and rhythm changes (e.g., Grube & Griffiths, 2009), neural correlates of these differences are needed to reveal distinct temporal dynamics and estimate neuroanatomical regions responsible for both perceiving meter and rhythm. Such a distinction is important to ensure that mechanisms representing serial patterning of event durations (rhythms) are not the same as those mechanisms that are responsible for representing temporal hierarchies of underlying periodicities within the rhythms (meter).

In line with this notion, Will and Makeig (2011) posit that it is necessary to distinguish neural
activity that may arise due to endogenous, higher processes responsible for grouping or organizing rhythmic sequences and those fixed neural responses that arise due to basic processing stimuli within a sequence. In other words, does neural activity in response to rhythmic sequences occur due to passive processing? Or are there active mechanisms which may be responsible for entrainment, expectations, and imposed structure? Indeed, a study conducted by Will (unpublished, reported in Will & Makeig, 2011) showed differences in early auditory sensory components between conditions of continuous stimulation and transient, intermittent groups of tone onsets. Particularly, N1 components were significantly reduced in continuous rhythmic stimulation than in intermittent conditions, as well as an enhancement of N2, P1, and P2 components. Moreover, there appeared to be stronger background activity in neural activation during continuous rhythmic stimulation conditions, although this was not tested for significance. Results from this study suggest that active, endogenous processes might underlie evoked responses to stimuli organized periodically, consistent with the phenomenon of entrainment and predictions of dynamic attending.

_Differentiating Rhythm and Meter Processing using DC Shifts and Early-Latency Components_

One of the earliest studies examining the neural correlates of rhythm and meter perception was conducted by Kuck, Grossbach, Bangert, and Altenmüller (2003). Using only a sample of trained musicians, a same-different task required participants to compare rhythmic or metric sequences that were either regular or irregular by varying different groupings of tone
onsets. In comparing meter conditions, physical accents in the sequences cued listeners to specific strong-weak beat structures. During EEG analysis, direct current (DC) shifts in 12-second time epochs around the sequences were compared for changes in meter and rhythm conditions. The authors found that DC shifts in the EEG signal arose as a biphasic negative-going plateau for all conditions. Overall, results indicated that sustained cortical activation in frontotemporal brain regions occurred when participants compared metric structures. Rhythmic structures gave rise to more centroparietal activation differences in DC shift when compared to conditions manipulating meter. The study provided incipient evidence for differential processing of rhythmic structures and metric changes due to distinct topographical activation for the conditions, and may suggest that two different mechanisms account for the perception of meter and rhythm.

Further work investigating early components of meter and rhythm processing were examined by Geiser, Ziegler, Jancke, and Meyer (2009). In this study, rhythmic sequences were constructed that changed in both meter and rhythm akin to Kuck et al. (2003). In three conditions, participants were either to detect changes in rhythm or meter, or detect tones of different frequency. The tone detection task was used to draw attention away from rhythm or meter changes. Moreover, the study compared musically-trained participants against individuals without training. Changes of both rhythm and meter elicited negative deflections in the ERP waveforms around 150 ms post-stimulus compared to non-deviant rhythms\(^1\).

\(^1\) Although this component may resemble mismatch negativity (MMN), the authors posit that their methodology does not resemble those used in traditional MMN studies; rather, their interpretation of the component is conservatively named the N150 here.
When detecting rhythmic changes, the negative deflection exhibited significantly larger amplitudes than meter changes. Interestingly, negative ERP deflections for meter changes only occurred during the task of detecting meter. In other words, the task of detecting different sounding tones in the sequence also elicited negative deflections when rhythmic deviants were presented, but meter changes during the tone-deviant task did not give rise to negative deflections in this time window for changes of meter. This suggests that active attending to meter is a requirement for detecting metric changes and perhaps is related to an underlying attention-related mechanism that is different from processing rhythms. No differences in ERPs were found between musicians and non-musicians at this latency, suggesting the role of training in detecting meter and rhythm changes did not arise in EEG signals. This finding was incongruent with behavioral performance on the task, as detection of meter and rhythm changes was significantly better in musically-trained subjects. It is possible that differences may arise in components of later latencies. However, those components were not examined here.

Aside from differentiation of meter and rhythm, processing differences between metrically-strong and weak positions in rhythmic sequences have been explored at early latencies. Particularly, mismatch negativity (MMN) components have been used extensively in auditory research to determine what features of stimuli can be detected (e.g., Näätänen & Alho, 1997). When violations of regularities from sound stimuli are presented, a negative-going component arising around 200 ms post-stimulus is thought to reflect a pre-attentive perceptual mechanism that responds to violations of expectations. Although MMN responses
have been found in absence of attention (e.g., when attending to another stimulus or task aside from the target violation), attention does appear to modulate MMN responses (Sussman, 2007).

Ladinig, Honing, Haden, and Winkler (2009) used MMN responses to differentiate strong and weak metric positions within stimuli of drumset patterns presented to musically untrained participants. Subjects were asked to listen to types of drumset patterns that had omissions of drumset “voices” (e.g., hi-hat, snare drum, and bass drum sounds) that occurred either on strong or weak metric positions and report instances in which they heard these omissions. A comparison condition had subjects listen to the sequences passively.

Behaviorally, listeners performed better and faster at reporting deviations on metrically-strong positions than weak positions. This result was corroborated by neural responses when comparing deviant stimuli to non-deviant stimuli. Presentation of both strong-beat and weak-beat deviants elicited a strong negative going component at 150-200 ms in frontocentral regions, described as MMN. Peak MMN latencies for deviants occurring on strong beat positions were much shorter than for weak beats, but MMN amplitudes were not significantly different between beat positions (Ladinig, Honing, Haden, & Winkler, 2011; erratum to Ladinig et al., 2009). MMN responses arising during passive listening were also not significantly different from those during active attending.

The authors suggest that meter may arise as a lower-level phenomenon due to the lack of an
effect. Aside from null results being difficult to interpret, an alternative interpretation might be that such attention-related differences arise more distinctly at later latencies, as they are strongly associated with endogenous, attention-related mechanisms that can exhibit top-down control over rhythmically patterned events (thoroughly reviewed below). Moreover, as MMN is thought to be associated with expectational violations based on memory for salient time points, (Winkler, Reinikainen, & Näätänen, 1993), an alternative interpretation of the results is that metric processes rely on memory for temporal representations of rhythmic sequences that work in concert with active attention-related mechanisms, and cognitive operations that encode and predict metric hierarchies may be reflected in several ERP components. This is consistent with the idea that internal representation of pulse, meter, and timing incorporates widespread dynamic activation across multiple neuroanatomical modules, consistent with functional neuroimaging studies (Grahn & Brett, 2007; Zatorre, Chen & Penhune, 2007; Fujioka et al., 2012).

Late-Latency Responses and Subjective Accenting

Stemming from the notion that late-latency ERPs are strongly associated with endogenous, attention-related mechanisms, recent work on subjective accenting has revealed electrophysiological differences between beat positions in metrically-organized sequences. As a reminder, subjective accenting is the phenomenal loudness differences perceived in physically-identical stimuli. Several studies have endeavored to describe top-down processing mechanisms that may be implicated in metric organization of strong and weak
beat structures. These studies often require the detection of a deviant stimulus in a sequence of comparison stimuli—often referred to as *oddball* paradigms—have been strongly associated with eliciting P300 responses from the cortex (Donchin & Coles, 1988). The P300 has been linked to stimulus context updating and shown to arise when listeners’ expectations are violated in rhythmic and musical sequences; variance in latency and amplitude of this component are attributed to task difficulty as well as allocation of attention to the task (Granot & Donchin, 2002; Luck, 2005 pp. 42-43).

In Brochard, et al. (2003), groups of musicians and non-musicians listened to a sequence of tones that were spaced isochronously by 600 ms. The number of tones in each sequence varied from 13 to 16 onsets. On tones 8 through 11, a deviant stimulus—a tone of equal frequency but decreased in amplitude by 4 dB—was randomly presented. One or two of the tones in positions 8 through 11 was a deviant. Here, strong beats are associated with odd-numbered onsets and weak beats are associated with even-numbered onsets. Participants’ task was to identify the tones that were decreased in amplitude in each sequence. No indication was given on how to perceptually organize (employ meter) during the experiment—it was assumed participants would interpret the sequences in a binary structure.

Results from EEG recording indicate that a larger P300 component was associated with subjectively accented deviants (strong beats) in comparison to unaccented deviants (weak beats), estimated to arise in the parietal region. The authors suggest that the differences in the components are attributed to higher-level cognitive processing as opposed to pre-attentive
components, although this was not tested directly. This top-down influence on component amplitude, consistent with previous evidence suggesting that late-latency responses may alter early stages of processing although delayed in the ERP waveforms (as in Olson, Chun, & Allison, 2001). P300 components also appeared to occur earlier in musicians than in non-musicians, and with significantly stronger amplitude. The authors propose that enhanced temporal processing—through extensive training—yields more efficient processing of the rhythmic sequences.

Building from the latter study, Abecasis, Brochard, Granot, and Drake (2005) used a similar methodology to Brochard et al. (2003) but employed accents during the isochronous sequence that were physically different from the unaccented beats. The participants were comprised of a sample of only non-musicians. Two conditions, binary and ternary groupings, were used to examine the differences between types of meter. The types of meter were determined by utilizing tones of longer duration for strong beats. Again, larger P300 amplitudes in the parietal region were observed for accented positions in both binary and ternary conditions in contrast to unaccented positions. The authors also observe that the P300 amplitude in the ternary condition was significantly smaller in the binary condition. As a conclusion, Abecasis and colleagues argue that a binary structure is a “default” perceptual organization mechanism for Western-enculturated listeners, evidenced by differing P300 components, and also supplied further evidence for beat-based neural processing differences.

Using the same procedure as Brochard et al. (2003), Potter et al. (2009) sampled from a
population of professional musicians. A specific early-onset ERP, Processing Negativity (PN), was hypothesized to be detectable for subjectively accented beats. PN is thought to be an ERP related to selective attention in processing of a targeted stimulus (Näätänen, 1982). Due to a referencing issue in recording the EEG signals in Abecasis et al. (2005), PN was unable to be found. After replicating the study with a musician sample, Potter and colleagues found evidence for PN during this task. Results indicated that for deviant tones in accented (strong beat) positions, a PN component was observable in contrast to deviant tones in unaccented (weak beat) positions in the left temporal region. Corroborating earlier studies, a N2/P300 complex associated with the processing of deviant tones was observable, with deviant tones in the accented position having significantly larger mean amplitude than deviant tones in the unaccented position. The authors conclude that attenuation via subjective accenting to isochronous tones is associated with a left-lateralized PN component on subjectively accented beats, and larger N2/P300 amplitudes parietally were related to subjectively accented beats over unaccented beats.

Schaefer, Vlek, and Desain (2011) examined differences in endogenously-imposed meter and exogenously-induced meter in isochronous sequences using binary, ternary, and quaternary metric organizations in participants. These listeners were not defined on criteria of musical training. Behaviorally, participants were to identify if a presented probe accent was on a weak or a strong beat. After analysis, N1/P2 complexes in the 100-300 ms latency post-stimulus were significantly larger in accented positions than in unaccented positions, with no significant difference between endogenous or exogenous meter. Differences were also found
in different types of unaccented positions, with the first unaccented position being significantly different from subsequent unaccented positions. Similarly, components in the latency window 300-450 ms post-stimulus showed significant differences in accented positions than unaccented positions, with no differences again between subjectively-maintained meter and objectively-induced meter.

To summarize, subjective accenting studies using EEG have shown three important findings. First, it appears that there are no differences in ERPs between endogenously-produced meter and meter that is induced from physical characteristics of stimuli. Second, differences in perceptual or cognitive processes are associated with strong and weak beats within metric structure, although the specific processes may be unclear. Finally, binary meter might be either easier to maintain in listeners, or it is a “default” organization when imposing metric structure since it is thought to be the simplest form of imposing a hierarchy. Although group differences in samples of musicians and non-musicians were considered, supporting evidence for the distinction is unclear.

**ERP and ERF Evidence for Metric Hierarchies**

Although subjective accenting has provided evidence for strong and weak beat structures to arise in physically-identical sequences of rhythms, components that arise due to expectation violations cannot supplant neural activity that may directly represent temporal hierarchies. In other words, ERP components such as MMN and P300 only offer an indirect method to
understanding the imposition of meter. Although few studies have been conducted on direct correlates of pulse or beat representation, evidence from two studies suggests that internally-maintained metric organization is observable in the EEG.

Nozaradan, Peretz, Missal, and Mouraux (2011) sought to examine whether entrainment arising as steady-state evoked potentials (EPs) represent pulse and meter perception. The authors recruited subjects with mixed musical training to listen to rhythmically-modulated tones. Stimuli were comprised of long duration sine tones that were modulated at a rate of 2.4 Hz to create a beat structure, with a beat corresponding to the peak of the modulation. A further 11 Hz amplitude modulation over the sequence, creating irregularities in the beat structure, intended to suppress an involuntary binary organization as reported by Abecasis et al., (2005). Importantly, the 11 Hz modulation ensured that pulse representation was endogenous by requiring participants to maintain a binary organization to counteract the irregular modulation. The subjects’ task was to attend to each sequence in three ways. First, a binary condition asked subjects to imagine the amplitude modulated peaks in a periodic strong-weak beat structure. A second condition had participants imagine a ternary meter resulting in a strong-weak-weak beat structure. Finally, a control condition incorporated a 4 ms gap randomly inserted into the amplitude-modulated tones that was a part of a detection task for participants.

The authors’ results indicate that significant increases in EEG amplitude correspond to the pulse frequency of the amplitude-modulated stimuli, both occurring at a rate of 2.4 Hz.
During binary meter conditions, additional amplitude increases occurring at 1.2 Hz, as well as additional .8 Hz and 1.6 Hz increases for ternary conditions, suggest that imagined meter resulted in an additional periodic signal that was locked to stimulus amplitude modulations. More simply, subdivisions of neural activity occurring in simple ratios from the endogenous pulse coincided with metric type. These subharmonic periodicities may reflect a top-down metric biasing of the stimulus processing, providing further support for entrainment models of rhythm perception that propose nested, subharmonic oscillations. Although neural activity was distributed across both hemispheres, evoked potentials were elicited maximally at frontocentral and temporal regions. Nozaradan and colleagues demonstrate that imagined meter directly affects EP amplitudes when listening to sequences that are not purely periodic. However, the experimenters were unable to find significant beat based differences. The results are further clouded by recruiting subjects with mixed musical experience, although it is unclear whether group differences may directly result in ERP amplitude differences.

MEG recordings also have supplied evidence for processing differences between metric types that are organized through endogenous (imagined) metric interpretations. Components extracted from MEG are referred to as event-related fields (ERFs), as they represent fields of magnetic activity created by electric dipoles resulting from neural activity; this is contrasted with event-related potentials (ERPs) which measure voltages. Using a sample of musicians with an average of 21 years of formal training, Fujioka, Zendel, and Ross (2010) had subjects tap along in either a binary or ternary meter to a preceding set of clicks. A presented tone signaled participants to stop tapping and to keep organizing a further set of clicks using the
meter from the tapping sequence. Another tone signaled subjects to again tap along to the sequence as a manipulation check.

Comparing strong and weak beats in both binary and ternary metric structures, all conditions gave rise to significant ERF responses around 60 ms post-stimulus in the right and left superior temporal gyrus. Upon examining the differences between binary and ternary meters, the authors found much later responses (~250 ms) in the right hippocampal region for ternary conditions than in binary (~80 ms), and an enhanced signal at the left striatum for binary peaking around the same latency. Binary meter conditions also yielded larger ERF responses than ternary in the right auditory cortices between 80 and 200 ms post-stimulus. Interestingly, strong and weak beat differences were only found in ternary meter conditions. Specifically in ternary meter, larger peaks around 80 ms post-stimulus were shown for strong beats in the right hemisphere, as well as the left hippocampus. The right hippocampal regions showed multiple ERF peaks at 150 ms.

Overall, Fujioka, Zendel, and Ross (2010) provide further evidence for processing differences in binary and ternary meter, with ternary meter giving rise to ERFs at a later latency in the right hippocampal region and the left striatum and weaker amplitudes in the right primary and secondary auditory cortex. Although beat-based differences did not arise in binary conditions, strong and weak beat differences in amplitude and latency in the right hemisphere and hippocampus were noted for ternary meter conditions. The results are consistent with the idea that endogenously-maintained meter utilizes several interacting.
distributed networks for organizing rhythmic structures, which call upon motor, pre-motor, auditory, and memory systems. The role of the hippocampus in such metric representations was not previously shown in studies, and may reflect memory for interval timing that facilitates anticipatory attending. The processing differences not found in binary metric structures between strong and weak beats are compatible with results from Abecasis et al. (2005), where binary metric structures may be viewed as a “default” low-order metric type that requires less attentional energy to maintain.

In sum, ERP and ERF studies show beat-based differences in processing are thought to reflect endogenous, temporal organizations of rhythmic sequences referred to as meter. Differentiated from rhythmic processing, meter can be evoked due to regularities in physical properties of stimuli or through internally-maintained impositions that yield similar neural correlates. Although the precise mechanism of meter remains vague, reliable differences in processing metric structures lend credence to the use of neuromagnetic and neuroelectric signals as a tool to uncover the sensory and cognitive processes involved in rhythmic attending.
Oscillatory Activity Underlying Meter

Introduction to Neural Oscillations and Meter

Recorded ERP responses involve time windows that extend several hundred milliseconds after a stimulus is presented to a subject. When considering rhythmic processing, rapid presentation of stimuli provides the opportunity for late-latency responses to overlap with the onset of a successive stimuli within sequences. As an example, some components such as P300 can vary in latency, depending on the conditions of the experiment (Luck, 2005). IOIs in the range of 300-500 milliseconds have been used in rhythmic studies, potentially overlapping with P300 components. Thus, precise temporal processing of such signals presents a problem when looking at evoked potentials as averaged waveforms (as noted by Zanto, Large, Fuchs, & Kelso, 2005; Iverson, Repp, & Patel, 2009; Will, 2011). Recently, oscillatory activity analyzed from EEG and MEG has provided a window to explore the process by which the brain attends and reacts to rhythmic sequences. Examined through time-frequency analyses, oscillatory activity preserves the ability to look at temporally-focused responses that do not overlap as in cognitively-related late-latency responses of evoked potentials, as the focus is on time windows immediately surrounding an expected onset.

Aside from latency-related benefits of time-frequency analyses to sidestep ERP issues in rhythm perception studies, oscillatory activity is also correlated with several perceptual and cognitive processes. Events, objects, motor plans, and expectations are thought to be
represented in part by coordinated and self-organized dynamic states of neural activity. Cortical regions process and exchange information in parallel, sustaining spatiotemporal patterns of activity that can arise as highly-precise, synchronized oscillations (Uhlhaas et al., 2009). Initially noted in recordings from the cat visual cortex, neural oscillations are thought to be responsible for several context-sensitive cognitive and perceptual functions, such as memory, attention, and sensory processing (Gray et al., 1989, Kreiter & Singer, 1996; Engel et al., 1991). Synchronized oscillations provide the brain with an efficient way to coordinate widespread neural firing as precise zero- or near-zero phase locking relationships during oscillatory patterning occur over long distances (König et al., 1993). These oscillations are often discussed in terms of frequency bands of neuronal firing. High frequencies, such as beta and gamma, comprise bands that correspond to a firing rate of 12-100 Hz. Low frequency activity, often below 12 Hz, comprise delta and theta bands (Uhlhaas et al., 2009).

An important distinction in oscillatory responses concerns evoked versus induced activity (David, Kilner, & Friston, 2006). Depending on how oscillatory activity is averaged, evoked and induced responses can be extracted from neuroelectric and neuromagnetic signals. Evoked activity is revealed when doing a time-frequency analysis of the signal after activity is averaged over trials. Induced activity is estimated when time-frequency analysis is performed for all trials and subsequent power is then averaged. By subtracting the estimated evoked power from this average, the induced activity remains. In relationship to stimuli, evoked activity is phase-locked to a stimulus. Induced activity does not exhibit such phase locking, and peaks of power only time lock to a stimulus (Tallon-Baudry & Bertrand, 1999).
Roelfsema et al. (1997) demonstrated that widespread coordination of neural firing is essential for coupling distributed functions, such as sensorimotor processing and anticipation. Results from this study suggest self-generated, high-frequency activity in gamma- and beta bands to be associated with attentional focus and may link sensory and executive brain systems that are responsible for expectation generation. It follows logically that gamma- and beta band activity are candidates for attending to rhythmic sequences. Indeed, high-frequency oscillations in the brain appear to occur in dynamic bursts that may correspond to onsets of sounds in a rhythmic sequence (Snyder & Large, 2005; Fujioka, Trainor, Large, and Ross, 2009; Fujioka et al., 2012). It is thought that these bursts of oscillatory activity may underlie the perception of meter and pulse, while also coordinating the communication of auditory and motor regions that are implicated in rhythm production (Snyder & Large, 2009).

Joel Snyder and Ed Large (2009) propose a resonance model—compatible with dynamic attending—that account for listener responses to rhythmic stimuli. Broadly, the pulse resonance model posits several predictions: spontaneous neural oscillations give rise to endogenous periodicities. These endogenous periodicities synchronize to external stimuli through entrainment. The primary periodicity represents the pulse frequency, and meter is represented through higher- and lower-order oscillations in conjunction with the primary periodicity.

As can be seen from the following reviewed studies, neuroelectric and neuromagnetic brain
responses and their behavioral correlates support the resonance model. Key to both this model and dynamic attending, brain responses appear to track and predict the onset of an expected event when placed in a sequence that invites rhythmic participation. The metrical organization of these sequences may represent a merger of top-down and bottom-up processing that draw upon cognitive mechanisms of attention and memory, as well as sensorimotor activation in the absence of overt movement. Temporal perturbations in rhythmic sequences disrupt the brain’s predictive responses only briefly, with brain responses again adapting to rate changes as predicted by the resonance model. Importantly, endogenously-generated metrical interpretation of sequences seems to generate neural responses similar to instances where meter is induced with physical changes of stimuli. As studies unfold, research on the neural mechanisms of predictive timing and entrainment shift from a focus of oscillatory gamma band responses to beta band responses. This shift highlights the importance of a putative audio-motor systems coupling, and suggests that motor activity is an important underlying mechanism of tracking and attending to events in time. As a reminder, previous work has examined latencies of responses in middle and late time windows (e.g., 150-200 ms post-stimulus). The locus of activity for the following studies is placed just before, on, or just after an expected stimulus—those regions that may correspond to peaks of attentional pulses proposed in theories of dynamic attending.
One of the earliest studies examining the oscillatory correlates of rhythmic processing was conducted by Snyder and Large (2005). In this study, the authors examined evoked gamma activity and induced gamma activity (~20-60 Hz range) through EEG recording while participants of mixed musical experience listened to rhythms organized in strong and weak beats. The authors found that all subjects exhibited induced and evoked gamma responses time locked to onsets of tones in rhythmic sequences, but variation in power and latency was found between subjects. Evoked gamma responses were found near 50 ms post-stimulus, and some participants had multiple activations in this region. One participant exhibited evoked gamma responses prior to stimulus onset. Induced activity preceded tone onset for two subjects and other subjects had maximal peaks immediately after stimulus onset. Two subjects appeared to subdivide the interval, exhibiting four distinct peaks for two beats. Generally, evoked gamma activity occurs after induced activity, while induced activity generally occurs before stimulus onset. Evidence of correlations between meter and brain activity were mixed. Notably, evoked gamma activity was stronger when tone intensity was stronger suggesting stimulus sensitivity followed strong and weak beats; however, induced gamma activity showed less intensity dependence. Evoked gamma band responses decreased in power following tone omissions, and induced responses arose before the time where tone onsets should occur (suggesting it occurs with anticipatory activation). The same was found for omitted/loud and omitted/soft conditions.
It is important to note that Snyder and Large (2005) do not control for precise imposed metric organization of the rhythmic sequences. They contrasted an exogenously-generated binary rhythm versus an ambiguous rhythm without physical differences. Participants may have utilized several strategies in order to track this temporal array, including subdivision of the rhythmic sequence or endogenously-generated organization of either sequence. The issue of control may contribute to differences in latency of gamma responses before and after expected onsets, as well as the occurrence of multiple gamma bursts in one of the subjects (i.e., a correlate of subdivision). In light of these caveats, this study showed anticipatory responses from the gamma band for rhythmic sequences that are also modulated by stimulus properties. From these two findings, the authors suggest that such anticipatory processes assist in the perception and memory of temporally-structured events, implicated in enhanced perceptual discrimination and encoding as suggested by Potter et al. (2009).

Addressing another prediction set forth by resonance models and dynamic attending, Zanto et al. (2005) investigated gamma band activity in relation to auditory sequences that are perturbed. In this study, perturbed sequences are trains of stimuli that shift in period at a target stimulus, where upon the inter-onset interval is decreased or increased relative to context IOI. For example, if preceding stimuli are separated by 500 ms, a deviant stimulus might occur a shift in period by +/- 125ms, yielding an IOI of 375 ms or 625 ms. Subsequent stimuli in the sequence will return to original IOI of 500 milliseconds. This yields three conditions: on time, where all IOI are 500; early, where a deviant IOI will be +125ms ahead of the expected onset; and late, where IOI deviates to 625 ms from the preceding stimuli.
Zanto and colleagues examined the gamma band responses to these sequences for participants of mixed musical experience. Evoked and induced gamma band activity was examined from scalp-recorded EEG and again was found to be time-locked to stimulus onset. Induced gamma activity was found to precede evoked activity, consistent with Snyder and Large (2005). During perturbations, tones occurring earlier yielded longer latencies for both evoked and induced gamma activity, while induced activity for late tones occurred prior to tone onset and earlier than against tones occurring on time. After perturbation, induced activity returned to baseline after both early and late tone perturbations. For the early condition, peak power for all gamma activity was greatest at perturbed tone onset. The authors note that this relationship between gamma responses is asymmetrical: Evoked peaks occurred where tones were expected; for induced activity, the peaks occurred where expected in the late tone perturbation conditions. Put concisely, the authors note that “early tones seem to come as a surprise, whereas late tones are actively missed” (pg. 545). Similar to Snyder and Large (2005), metric interpretation was not controlled during this study. Although brain responses did not elicit as large of latency inconsistencies, modulation of these responses via metric interpretation cannot be ruled out when considering gamma response asymmetries.

Noting such inconsistencies in metric interpretation, Iverson, Repp, and Patel (2009) were interested in differences between endogenously- and exogenously-generated meter. Specifically, instances in rhythmic sequences where pulses aligned with silences were used to examine what the authors refer to as an “imagined beat effect,” where response differences
result from a top-down metrical interpretation of a rhythmic sequence. In the first task, participants listened to an ambiguous rhythmic sequence that could be metrically interpreted in two ways by participants “feeling” strong beats in one of two possible positions within the sequence. In a second task, amplitude differences between tones in the sequences provided an exogenously-induced meter. Overall, when looking at beta and gamma band activity, evoked beta amplitude showed increases at time points where a beat was imagined. Such differences did not arise for gamma-band activity. The result was consistent for conditions where meter was both produced by physical accents from amplitude changes, and for conditions with imagined meter over identical rhythmic sequences.

Interestingly, Iverson, Repp, and Patel (2009) re-interpret the findings of Snyder and Large (2005) positioning that the gamma responses of predictive timing are indeed beta responses. The bandwidth considered gamma by Iverson and colleagues occupies the range of 30-50 Hz, whereas the beta band is considered to be 15-30 Hz. This range is contrasted with Snyder and Large (2005), as they consider the gamma range to cover 20 to 60 Hz. Importantly, gamma-related changes during attending to rhythmic sequences—depending on bandwidth interpretation—may suggest changes in perceptual discrimination of stimulus properties due to anticipation whereas beta activity may be related to predictive timing. It is critical to note that there is not a universally accepted range of bands that correspond to activity. Thus, the line between gamma and beta band activity obscures what mechanisms are contributing to processing and interacting with temporal structures. However, the findings of Iverson, Repp, and Patel may have contributed to the shift of focus to the beta band. As will be reviewed,
this frequency range is related to sensorimotor regions of the brain (Murthy & Fetz, 1992; Sanes & Donoghue, 1993). Thus, beta activity may represent coupling of auditory and motor systems and be implicated in a person’s ability to synchronize motor movements with auditory activity, such as in tapping along to rhythms or dancing.

Fujioka, et al. (2009) investigated beat processing correlates through MEG at the beta (~20 Hz) and gamma (~40 Hz) bands by using isochronous tones with probabilistic omissions. The subjects listened passively while watching a silent movie. Similar to Snyder and Large (2005), every other tone was presented 6 dB below the previous tone to induce a binary meter. Half of the blocks omitted a soft tone; half the blocks omitted a loud tone. Omissions occurred at a probability of 30%. Participants listened passively as they watched a silent movie.

There were no significant differences found in a grand average of neural activity between loud and soft tone presentations. This may imply that induced meter (listened to passively) was not observable for this task. The authors, however, did find stimulus related changes in both bands for all conditions—loud/soft and loud/soft omission, suggesting stimulus characteristics can modulate gamma and beta power in processing rhythmic sequences that are consistent with Snyder and Large (2005). Gamma activity peaked at 80 ms post stimulus and during omissions at 100 ms post-stimulus. Amplitudes of activity were larger after omitted and non-omitted tones than at times before stimuli occurred (185 ms pre-stimulus). The authors suggest that gamma activity reflects an endogenous process since the gamma
activity did occur during omissions albeit with a longer latency. This has implications in stimulus and pulse encoding and anticipatory processes. The distribution of gamma activity is suggested to be spread across many regions aside from the auditory cortex (e.g., prefrontal and anterior cingulate cortex), which are not analyzed here. Beta power had a decrease after each stimulus that reached minimum 200 ms post-stimulus. This decrease did not happen after omissions. This was significantly different from 20 ms post-stimulus and this significantly interacted with latency: Power decreased significantly after non-omitted tones but not for omissions. Fujioka and collaborators suggest that motor preparatory processes signaled by timing cues are aided by activity reflected in the beta band.

In a follow up study, Fujioka and colleagues (2012) set out to examine beta band oscillatory activity and predictive timing. To assess functional connectivity between auditory and motor regions, phase coherence and amplitude of beta activity was examined in 12 participants. The study does not specify if any subject has had formal or sustained musical training. While recording brain activity from MEG, subjects in the study listened to sequences of tones in conditions of regularity or irregularity.

Overall, results suggest that beta oscillation power in the range of 20-22 Hz followed the periodicity of the presented sequences. The bursts of beta responses appeared to be modulated at the periodicity rate of the rhythmic stimuli, and are discussed as increases and decreases of such activity in a ramped manner surrounding the onset of tones in the sequences. When contrasting the rise of beta activity before predicted onset with the fall of
activity after the onset in auditory cortices, significant differences were noted between irregular and regular rhythms. Beta ramping prior to stimulus onset was found to be much steeper in irregular conditions. However, no significant difference was found between irregular and regular rhythms in the decreases of activity that occur post-stimulus. These results suggesting regular and predictable rhythmic sequences invite endogenous predictive timing mechanisms. Aside from activity in the auditory cortices, a correlation was also found in beta modulations originating from motor-related cortical and subcortical regions. Phase coherence in auditory and motor regions was also modulated consistently within auditory and motor regions; although the phase coherence was correlated, it remains unclear if this coherence represents functional connectivity between the regions. Critically, results from Fujioka and colleagues (2012) present the first evidence of such beta modulations found in the cerebellum, supplementary motor areas, and inferior frontal gyrus that align with fMRI studies (Grahn & Brett, 2007, Zatorre, Chen, & Penhune, 2007).

Taken together, the summarized studies on oscillatory neural activity and rhythmic sequence processing suggest that several coordinated bands of activity underlie rhythmic perception and predictive timing in the brain. Moreover, it appears that examining oscillatory activity serves to be a reliable method to uncover predictive timing mechanisms, as well as perceptual discrimination and encoding processes. Evoked and induced gamma band activity may reflect anticipatory processes that precede the onset of expected auditory events and are sensitive to perturbations in sequences (Snyder & Large, 2005; Zanto et al., 2005; Fujioka et al., 2009). Beta oscillations, periodically modulated at periodicities of temporal structures, also predict
expected auditory events and may represent neural communication between auditory and motor areas and aid in predictive timing, even without overt movement of the body (Iverson, Repp, & Patel, 2009, Fujioka et al., 2009, Fujioka et al., 2012).

A critical facet of the studies should be highlighted that may account for some differences. First, the role of metrical interpretation of sequences, representing top-down processes that can govern predictive timing, remain vague. Iverson, Repp, and Patel (2009) demonstrated differences in the beta band for imagined beats during two types of metrical interpretation that were both exogenously and endogenously informed. For other studies, exogenously-induced binary interpretations of rhythms were predominantly used, and no control was used to account for an endogenously-generated binary interpretation of sequences. Often, the sequences were listened to passively while subjects silently watched a film or were instructed to remain still. Results from Nozaradan et al. (2011) suggest that interpretations of ambiguous sequences led to distinct changes in brain activity for mentally-maintained meter; without controlling for these effects, functional significance of oscillatory activity remains obscured. Potter et al. (2009) and Schaefer, Vlek, and Desain (2011) demonstrated such beat based differences in maintained metrical interpretation, showing differential processing between strong and weak beats at late latencies. Although there is no evidence that processing differences are reflected in time windows surrounding stimulus onset, it cannot be ruled out that metric interpretation as strong and weak beats contribute to high-frequency activity modulations during rhythmic attending.
Although the current review has primarily focused on bursts of neural activity in high frequencies (e.g., gamma and beta bands), low-frequency activity in theta and delta bands also have been explored in relation to rhythmically presented stimuli. In these cases, low-frequency activity has a benefit of possibility occurring at the same periodicity as rhythmic stimuli. For example, a rhythm with an IOI of 600 ms occurs a frequency of 1.67 Hz. Delta waves, occurring at frequencies between 0-4 Hz may period- and phase-lock to temporal structures. Evidence of such activity was reported by Lakatos et al. (2008). The authors examined low-frequency entrainment of oscillations to visual stimuli in macaque monkeys and found delta wave entrainment to presentations of rhythmically-displayed lights. If such activity occurs in non-human primates to a different sensory modality, it is plausible that low-frequency oscillatory activity may also be involved in auditory rhythms in humans.

Consistent with this notion, Will and Berg (2007) presented listeners with continuous streams of rhythmic stimuli comprised of synthesized drum sounds that were spaced isochronously for two minutes while subjects listened passively. The two-minute interval was followed by a period of pink noise or silence. Primarily, Will and Berg focused their analysis on phase alignment correlations across trials in order to assess coherence. Overall, phase coherence was significantly different between the rhythm and silence or noise conditions. Whereas beta and gamma (13-44 Hz) synchronization to rhythmic stimuli generated an increase in phase coherence in these bands over the rhythmic periods, low frequency activity from 1-12 Hz in
the delta to alpha range exhibited larger coherence, especially around 3 Hz. The authors argue that this phase coherence represents the organization of neural activity during periodic stimulation, related to endogenous processes as they are not stimulus-locked. Overall, the results suggest that low-frequency neural activity also contribute to predictive timing evidenced in stimulus-locked spectral power within higher frequency bands and may be related to the entrainment of internal oscillations to periodic stimuli.

Musician and Non-musician Differences

A common theme underlying neuroelectric and neuromagnetic correlates of rhythm and meter perception points to the differences in responses that arise between individuals with musical training or experience and those without musical training. Due to the necessitation of integrating multiple cognitive, sensory, motor, and executive systems, it has been argued that musicians serve as ideal models of neuroplastic changes due to years of experience, exposure, and focused training (Münte, Altenmüller, & Jäncke, 2002). Thus, they offer unique opportunities to be compared to individuals without such training to reveal much about neural dynamics and functional connectivity, even over the course of a life span. However, it is often the case that the criteria separating the two groups in behavioral and neuropsychological studies are inconsistent, undistinguished, or ignored. When such distinctions are absent or are not made clear, results from studies that could rely on such training could become clouded or misinterpreted. It goes without further explanation that musician and non-musician group differences present several validity-related problems when
trying to assess both behavior and neural responses when faced with auditory tasks.

In neuroimaging studies, structural differences within the brain have consistently differentiated individuals with musical training from those who do not. Gaser and Schlaug (2003) used fMRI to show increases in cortical volume in auditory and motor regions in musicians compared to non-musicians. Chen, Penhune, and Zatorre (2008) showed increased activity in prefrontal regions in musicians during rhythmic learning tasks. It also appears that coding of speech and music signals is enhanced in subcortical structures in musicians over non-musicians (Strait, Kraus, Skoe, & Ashley, 2009; Musacchia, Sams, Skoe, & Kraus, 2007).

ERP studies have also contributed to the musician and non-musician comparison. For example, Koelsch, Schröger, and Tervaniemi (1999) demonstrated that professional musicians exhibited MMN components to mis-tuned frequencies in musical chords. Prolonged attention-related ERPs in were found in musicians as compared to non-musicians (Münte et al., 2003). Amplitude and latency differences in MMN responses arose between professional musicians and non-musicians in violations of tone sequences and melodies (Lopez et al., 2003).

In contrast, several studies have mixed or vague musician and non-musician differences. For example, no difference was found between musicians and non-musicians in MMN responses to temporally-regular sequences, but a difference was found in sequences that did not exhibit
temporal regularity but had regularity in the quantity of tones presented in a sequence (van Zuijen et al., 2005). Similarly, no difference was found in latency or amplitude in omission-evoked potentials (thought to be related to P300 responses) between rhythmically-trained musicians (e.g., professional drummers and bass guitarists) and non-musicians in tones omitted from isochronous rhythmic sequences; variations in OEP latency did significantly differ, with rhythmically-trained musicians exhibiting less variance in jitter than non-musicians (Jongsma, Quiroga, & van Rijn, 2004).

Short Review of Musician and Non-Musician Differences Arising in Neuroelectric and Neuromagnetic Studies of Meter

In both the reviewed ERP/ERF and oscillatory studies concerning meter, the use of musicians and non-musicians is inconsistent. Three studies utilized a minimum number of years of musical training, but did not specify type of training or whether it was continuous and had coincided with the time of the respective study (Kuck et al., 2003; Potter et al., 2009; Fujioka et al., 2010). In contrast, three studies used samples of non-musicians or those with less than one year of informal training (Abecasis et al., 2005; Ladinig et al., 2009; Fujioka et al., 2009). Three experiments did not specify musical training, (Will & Berg, 2007; Iverson, Repp, & Patel, 2009; Fujioka et al., 2012), while four studies reported a mixed sample ranging from no formal training to university-level music faculty (Snyder & Large, 2005; Zanto et al., 2005; Nozaradan et al., 2011; Schaefer, Vle, & Desain, 2011). Brochard et al., 2003 compared musicians and non-musicians, with musicians requiring at least five years
of formal training. One study—Geisler et al. (2009)—incorporated a perceptual task (Gordon, 1989) that distinguished recruited musicians from non-musicians and correlated with self-reports of musician status. Whereas the mix of musical and non-musical experience might suggest a generalizability of results concerning the existence of mechanisms responsible for meter, the effect of musical training on specific tasks and within neural correlates of meter remains untested directly.

It appears that even strategies for organizing rhythmic sequences differ between trained musicians and untrained listeners. A study by Jongsma, Desain, Honing, and van Rijn (2003) examined P300 responses to violations in rhythmically isochronous sequences in musicians and non-musicians (no criteria given). Musicians were found to exhibit larger P300 responses to a higher-order metric violation than non-musicians, who in contrast exhibited a larger P300 for sequential violations. These results suggest that musicians may prefer to organize rhythmic sequences due to metric hierarchies, and sequential processing may be more favored by non-musicians. Thus, musicians may be more prone to using top-down strategies learned during extensive training in order to organize rhythmic patterns than do non-musicians.

Instruments Used for Differentiating Musicians and Non-musicians

It is clear from the reviewed studies on neural correlates of meter that musician and non-musician differences have not been made distinct. It is important then, to use a valid
instrument—whether through self-report or perceptual testing—to determine an individual’s musical experience and training. Examples of perceptual tests in the literature include the Montreal Battery of Evaluation of Amusia (Peretz, Champod, & Hyde, 2003) and the Musical Ear Test (Wallentin et al., 2010), that ask listeners to make same-different comparisons of musical features. Self-report measures, aside from basic questions asking about length of formal musical training, include Preisler (1993) and Cuddy and Lyons (1981). These classification measures generally incorporate questions regarding formal musical knowledge and aspects of musical activities (e.g., Cuddy & Lyons, 1981; Preisler, 1993). An obvious approach is to create an instrument that combines both methods; indeed, Müllensiefen, Gingras, Stewart, and Musil (2011) are currently developing a tool that is designed to incorporate both self-report and perceptual measures. ERP tests designed to show musician differences have also been proposed, such as a MMN paradigm that examines listeners’ responses to violations in musical sequences (Vuust et al., 2011).

One promising instrument that examines a confluence of musical experience, involvement, exposure, and training is the Ollen Musical Sophistication Index (OMSI, Ollen, 2006). The OMSI is a 10-item self-report questionnaire that attempts to characterize musical training and experience based on 9 validated indicators. Joy Ollen was motivated to develop the tool after a large sample of reviewed studies demonstrated that number of years of formal training or was not a reliable indicator of musical ability, with inconsistencies in relating perceptual performance to years of musical training found in several studies (e.g. Ollen’s examples of Krumhansl 1996; Panion 1989; and Butler & Brown 1984). It appeared that various forms of
musical involvement including concert attendance, compositions written and performed, earned degrees in music, and hours of regular practice, all contribute to a notion of musical sophistication.

Sophistication—intended to be synonymous to musical ability—is utilized to encompass musical skill in addition to involvement, ultimately connoting “connoisseurship” (pg. 5). As a consequence, Ollen created the OMSI after distilling 9 indicators contributing to musical sophistication from an original 38 categorical indicators from a large survey of musicians. The OMSI’s indicators focus on the variables of years of continued musical activity, years of private music lessons, years of daily music practice, current length of daily practice, enrollment in music classes at a collegiate level, years of classes at a collegiate level, amount of experience composing music, number of concerts attended in the last year, and a self-assessment of musicianship level.

To develop the tool, Ollen first extracted 38 categories of variables from a large sample of music studies that examined musician and non-musician differences. Second, an Internet questionnaire asked a sample of musical experts to provide indicators of musical ability. These were compared to the 38 original categories, further parsing down the variables into 17 categories with 29 indicators. These 29 indicators were placed into a questionnaire given to a sample of 633 adults in a variety of amateur and professional musical ensembles. Conductors and directors from these ensembles then ranked each of the 633 participants on a scale, and the data were analyzed using logistic regression. The resulting nine-indicator model
accounted for nearly 80% of the 633-person sample, which became the OMSI. A follow-up study cross-validated the original results, generalizing to a larger culturally-diverse sample (Ollen, 2009).

Due to the consideration of multiple variables that may contribute to explaining musical abilities—and more importantly the generalizability of the tool due to the sample sizes in the original and validation study—the OMSI presents itself as a strong candidate to discern musicians and non-musicians, especially in the case of examining meter. If it is the case that meter employs multiple cognitive functions in order to track and predict temporally-organized sounds, musical sophistication as measured by the OMSI might accurately account for extensive training in both a domain-general and domain-specific perceptual and cognitive functions (i.e., larger focus of attention and improved memory for auditory tasks, as well as enhanced auditory perception). Moreover, an added benefit of the instrument is that it does not dichotomize a sample: Meter-related studies often have a cutoff criterion (e.g., at least x years of formal training) that serve as arbitrary distinctions based on convenience. Here, the OMSI portrays musical sophistication as a continuum of experience and ability, allowing the correlation of this variable to both behavioral and neural responses to rhythmic sequences.

Motivation of Study

From the reviewed studies, it appears that attention focused to points in time can enhance auditory perception. Temporal expectations established by predictable event rates increase
accuracy in discriminating pitch and timing, and these expectations may be built into
hierarchies reflecting a series of nested nonlinear dynamic oscillators that are harmonically
related. There is evidence that this type of activity is reflected as synchronized neural activity
arising from internally-maintained meter employed by participants in several experiments.
Meter used by these participants is purported to draw upon several cognitive mechanisms,
including attention, memory, and motor planning to enhance encoding of auditory events and
predict their occurrence. Although mixed, results from some studies (e.g., Brochard et al.,
2003) support the notion that long-term training can strengthen metric representations and
thus facilitate the organization of perceptual events.

However, it remains unclear what differences may arise in auditory perception under
different types of endogenously imposed meter, and whether enhancements of auditory
perception of meter is explained by long-term training. Specifically, modulations of attention
that are purported to be the result of pulse level alignment might give rise to differences in
auditory perception. These differences may arise between perceived strong and weak beats,
with enhanced processing occurring at strong beats over weak beats. It is possible to design a
study to examine these relationships, and provide supportive evidence to a developing story
on the way to a comprehensive understanding of meter. Such a study would take of
advantage of an opportunity to further understand higher-level cognitive operations that may
guide our perceptual experience in the auditory modality, implicating itself in the perception
and cognition of speech and music.
A study is presented here that is in accordance with this opportunity. Listeners are asked to listen to simple rhythmic sequences of brief stimuli, and are asked to mentally-impose two different types of meter upon the sequences. To test whether hierarchical organization can enhance auditory perception, small deviations in individual stimulus properties are required to be reported in relation to the type of meter used by listeners. Moreover, each listener’s musical training is measured, and this measure is correlated to both behavioral responses and recorded neural activity. Primarily, the current project hopes to corroborate earlier studies by finding the neural activity that may be representing meter, and using this information, predict when listeners will make errors. Differences are expected in perception to arise that are related to metrically-strong and metrically-weak positions, with the former related to enhanced perception. If supporting results are found for these predictions, it is suggested that meter establishes hierarchies of attentional energy at specific time points that correspond to strong and weak beat structures, increasing sensitivity to auditory stimuli that occur when expected.

Most studies looking at meter-related effects on auditory perception have concerned perceived pitch differences and IOI timing judgments in IOIs in sequences of tones, or manipulate event rate to judge how sensitivity to timing changes underlies dynamic attending (e.g., Large & Jones, 1999; Barnes & Jones, 2000). Here, a novel perceptual ability was used: temporal order acuity. Instead of varying a temporal structure itself, properties of each individual stimulus are varied in presentation to examine sensitivity when attending dynamically and imposing an internally-maintained simple metric hierarchy.
Temporal order acuity concerns the ability of the auditory system to detect the delay between two brief sounds. Importantly, such studies do not necessarily try to determine when the threshold that one sound appears as two sounds, but being able to discriminate pairs of such sounds by varying differences in intensity and time between the brief sounds while maintaining a constant energy spectrum. Thus, such stimuli often are perceived as one sound with phenomenal differences, arising due to a process known as temporal integration (Green, Birdsall, & Tanner, 1957).

Depending on the characteristics of the stimuli and the procedure, varying results are found for thresholds of discrimination of such sounds, generally on the order from .07 ms to 5 ms with extensive training (Green, 1970; Henning & Gaskell, 1981). Generally, it appears that around 12 to 15 ms is when two of such brief sounds used in temporal auditory acuity studies is perceived as two sounds (Joliot, Ribary, & Llinás, 1994), and durations of each stimuli have been varied from around 250 microseconds to 200 ms (Ronken, 1970; Zwislocki, 1960). Although there is some disagreement about whether temporal integration occurs at a peripheral level before neural transduction in the auditory system or whether such a process is more central (noted by Krumbholz & Wiegrebe, 1998), such stimuli are proposed here to be adequate for examining top-down effects on peripheral auditory processing.
Further examples of work on temporal order acuity show a range of stimulus types and approaches that give rise to nuances in discrimination (as provided by Green, 1971). Patterson and Green (1970) utilized transient stimuli that included phase differences in frequency bands from a wideband pulse, and listeners were able to detect changes down to the range of 1 to 2 ms. Duifhuis (1971) provides some evidence that temporal order acuity varies with frequency range, with lower-frequency signals being more difficult to discriminate than higher frequencies. Attempting to generalize to speech processing, Leshowitz (1971) used 50 ms synthesized stimuli mimicking fricatives, demonstrating that listeners demonstrated acuity to such stimuli down to 5 ms.

Don Ronken’s (1970) study on temporal order acuity provides the basis for the stimuli used in the present experiment. Ronken constructed stimuli comprised of two wideband rectangular pulses that are 250 microseconds in duration. In his studies, Ronken varied the distance between each wideband pulse as well as the amplitude difference between them. Procedurally, listeners in his study made same-different distinctions of stimuli in a two-interval adaptive procedure. Generally, results show that participants were able to discriminate timing differences down to 2 ms. Ronken also found that timing separation distances were easier to discriminate than amplitude differences.
Experimental procedure, hypotheses, and operationalizations

In order to examine detectability of click pairs modeled after Ronken (1970) in rhythmic sequences, a two-part experiment was conducted where isochronous sequences of Ronken clicks were assembled and presented to participants. One of the click pairs was reversed in order, creating a deviant stimulus. Using different types of meter—groupings of two (binary) or three (ternary)—participants reported metric positions on which deviant click pairs occurred within sequences. During this task, participants’ neural data was recorded using EEG.

Overall, the following hypotheses were made:

H1: In both groups, neural representations of meter will reliably predict errors in reporting metric positions containing deviant stimuli.

H2: Musician performance on the reporting metric positions of deviants will be significantly better than the non-musician group.

H3: Binary meter would yield a better performance of detecting deviants than ternary meter.

H4: In both groups, detection of deviant stimuli on metrically strong positions (beat 1) will be
better than on weak beats (beat 2 in binary and beat 2 and 3 in ternary).

Operationally, meter is defined as cyclic organization of sounds in groups of two (binary) or three (ternary), creating two meter conditions. Beats were operationally defined as positions within the cyclic organization. Strong beats were the first position of both binary and ternary conditions, and weak beats were positions two for both conditions and position three for ternary conditions. Long-term experience was operationally determined by a self-report measure of musicianship, and this measure is used to generate two comparable groups of musicians and non-musicians. Target stimuli were defined as a time-reversed click pairs embedded within a context of Tocks, and vice versa. In order to test the possibility that presentation order of the clicks influences results, the context was varied by incorporating frontwardly-presented clicks (Ticks, following terms used by Green, 1971) and backwardly-presented clicks (Tocks).
Chapter 2: Methods

Participants

Eleven participants were recruited to complete the experiment and consisted of students from The Ohio State University. Six males and five females, age range 19-31 ($M = 23.91$, $SD = 3.83$) reported normal hearing and indicated right hand dominance. Before the experiment began, all participants were briefed on the procedure, risks, and benefits of the study in accordance with the requirements of the Institutional Review Board at The Ohio State University. All subjects signed consent forms indicating their understanding of the experiment, and were informed that they would be able to leave the study at any time immediately and without question. No participants chose to leave the study. Upon successful completion of the study, all subjects received twenty U.S. dollars.

Subjects were also assessed on their musical ability and expertise in order to describe two groups within the sample by completing the Ollen Musical Sophistication Index (OMSI) (Ollen, 2006; Ollen, 2009). As a reminder, the OMSI is a 10 item self-report questionnaire that attempts to characterize an individual’s musical ability and experience that assigns a sophistication score—ranging from 0 to 1000—based upon validated indicators of musical
experience. The Ollen Musical Sophistication has been reproduced in Appendix A.

Participant results from the OMSI scores were examined on a scatterplot, found in Figure 5. The average OMSI score ranged from 99 to 989, with a mean of 568.18 ($SD = 386.05$). For the purposes of separating the two groups for the main experiment with preliminary analysis, a self-indication of musicianship built into the OMSI was linked to each score. Two clear groups from the self-report question emerged at ends of the 1000 point scale. A two-tailed $t$ test was conducted on the OMSI results for the two groups. Results indicate that distributions of musicians were significantly different from the non-musician distribution ($t(9) = 9.16, p < .05$). These two groups were then used to generate the experimental groups of musicians and non-musicians used in the present study. Analyzing scores for the new groups, the average OMSI for musicians is 887.50 ($SD = 139.00$), and for non-musicians is 185.00 ($SD = 109.35$).
Figure 5. Scatterplot of OMSI scores for all participants.

Screening Experiment

Materials

All stimuli for the screening experiment were presented through Sennheiser HD 280 Pro headphones running from a Dell Precision R5400 computer in a presentation GUI running in MATLAB (Oh, 2009). Participants were seated comfortably in front of a 15” Proview computer monitor within an Industrial Acoustics Corporation Sound Attenuated Booth.

Earphones were calibrated following a method outlined by Russotti, Santoro, & Haskell
The earphones were initially calibrated using a Larson-Davis 824 Sound Level Meter. The sound level meter was calibrated with a CAL 250 class 1 calibrated sound source emitting a 250 Hz sine tone at 114 dB. SPL readings were shown to be congruent in the sound level meter. The Sennheiser earphones were placed upon a KEMAR manikin containing ear simulators with ¼” microphones. Peak-to-peak equivalent sound pressure level measurements were used to calibrate the Ronken click pairs used in the screening experiment (Haughton, 2002). Peak-to-peak equivalent SPL compares the pressure waveform of a transient click to that of a long duration sinusoidal waveform. During this process, the tallest click in the Ronken click pair was matched in amplitude to a 1000 Hz sine tone through a Tektronix 2214 digital storage oscilloscope. Here, SPL measurements were made from the 1000 Hz tone through the sound level meter. Following Haughton (2002), click pairs presented during the experiment were at 106.8 dB peSPL\(^2\) (peak-equivalent SPL). A TENMA 72-1023 Multimeter indicated both left and right insert phones to be receiving an input of .105 V RMS.

**Stimuli**

Stimuli used for the screening experiment were generated in MATLAB and followed the stimuli used by Ronken (1970). The stimuli used were two 250 microsecond wideband pulses separated by a variable time interval \(t\), with one click presented at half the amplitude \(A\) of the

\[\text{peSPL} = 10 \log_{10} \left( \frac{p^2}{2 \times 10^{-6}} \right) \]

\[\text{SPL} = 10 \log_{10} \left( \frac{p^2}{2 \times 10^{-6}} \right) \]

\[\text{peSPL}^2 = 10 \log_{10} \left( \frac{p^2}{2 \times 10^{-6}} \right)^2 \]

Although this sound pressure level appears to be presented at dangerous levels, the transience of the brief clicks is not phenomenally perceived as loud as 106 dB. Thus, listening levels were comfortable for all participants.
other (6 dB difference). Phenomenally, the frontwards presentation is often described as a “Tick” whereas the reversed presentation is thought to sound like a “Tock.”

![Diagram of click pairs with Amplitude and Time axes showing A = 1, A = 1/2, and time t for both Tick and Tock.]

Figure 6. Click pairs modeled after Ronken (1970).

**Procedure and Analysis**

In order to assess subjects’ perceptual discrimination abilities of Ronken clicks, a two alternative forced-choice paradigm was employed using a 1-up 3-down adaptive procedure (Levitt, 1971), where subjects’ task was to determine if presented click pairs—Ticks and Tocks—were heard to be the same or different. The procedure used three trial blocks with no more than 100 trials to determine a performance rate of 79%. The initial time difference between each click was set at 10 ms, and adapted in increments of 2 ms. The inter-stimulus
interval for all trial blocks was set at 600 ms. Participants were given feedback after each trial, and were offered breaks in between trial blocks if desired. Timing separations for each trial block were then averaged to determine a final click separation interval. Each subject's average click separation interval was used for the EEG portion of the experiment.

Following Levitt (1971), performance rate is determined by a transformed up-down method. During this procedure, the properties of presented stimuli are dependent upon the performance of each subject. Stimuli change in increments, known as steps. Steps can either move up or down. Upward steps increase the magnitude of stimulus change, whereas downward steps decrease the magnitude. During this specific procedure, a 1-up 3-down procedure indicates that incorrect responses in a same-different task increase the magnitude, or moving a step up by one. Conversely, if the magnitude of stimulus change is to decrease, participants must make three correct responses. To get to a 79% rate of performance, the chances of a step up must equal the probability of a down response. Thus, if the probability \( P(X) \) of stimulus level X must equal .5, and three correct responses are required to decrease stimulus magnitude, then \( [P(X)]^3 \) would yield .794.
Main Experiment

*Materials*

All stimuli during the EEG experiment were presented binaurally through EARTONE 3A transducers to EAR3-14A insert ear buds placed comfortably in participants’ left and right ears. Calibration of insert phones followed the procedure used in the screening experiment. Here, the insert phones were affixed to a 2 cc coupler on a Larson davis 2575 1” pressure microphone, and were shown to produce 98.9dB peSPL (.048 V RMS through the multimeter). Overall, this produces an 8 dB difference from the screening stimuli. All stimuli for the main experiment were presented through a 15” Apple MacBook Pro and M-Audio Amplifier producing tones and clicks through RtAudio library (Scavone, 2001) built into PyEPL presentation software (Gellar et al., 2007).

EEG data collection occurred in a 6.5 ft by 8.5ft electrically-shielded, sound attenuated room. Stimulus presentation and behavioral data collection employed a 15” Apple MacBook Pro with a 2.66 GHz Intel Core i7 processor. A wireless (Bluetooth) keyboard was used by participants to navigate and respond to the experimental software. Sync pulses were sent via USB from the MacBook Pro into a LabJack automation device which was connected via DB cable to the amplifier.

A 128-electrode Brain Products actiCAP cap housing an array of 96 electrodes was
connected to a Brain Products actiCHamp amplifier powered by a Brain Products Powerpack battery. Recorded EEG signals with sync pulses were transmitted via DB cable to a Dell Optiplex 980 desktop computer utilizing a 2.5 GHz Intel Core i5 vPro processor with used for recording. The recording computer utilized PyCorder software to record unfiltered EEG data, sampled at 1000 Hz.

**Stimuli**

Stimuli were comprised of click pairs used in the screening experiment, but were instead generated using NumPy arrays (NumPy, 2005) within a Python environment and presented through PyEPL. The timing distance between each click in the pair was determined from the performance rate of the screening experiment matched to the participant. Figure 7 describes each trial stimulus. Within the experiment, isochronous sequences of Ronken click pairs were presented to listeners. Stimuli used in the main experiment were 22 click pairs separated by 600 ms and presented either in a forward or backward fashion. Each trial was preceded by priming tones depending on condition. A series of either 6 priming beeps with alternating frequencies at 720 Hz (green boxes, strong beats) and 440 Hz (purple boxes, weak beats) was presented for binary meter, or 9 priming beeps with two cycles of 720 Hz tones followed by two 440 Hz tones for ternary. The first blue box highlights the “standard” sequence of positions 1-12, in which there were no deviant stimuli presented. A “target” region randomly presented a deviant stimulus (reversed-order click pair) on one position on onsets 13-18 that are boxed in red. A final 4 “continuation” clicks, boxed in blue, and completed each
sequence. Aside from the priming sequence, stimuli were identical for binary and ternary condition. Participants were unaware that the deviant stimulus was presented in a limited range; they were given no indication to how early or late a deviant stimulus might occur.

**Stimulus Sequences Used in Main Experiment**

![Stimulus Sequences Diagram]

**Figure 7.** Stimuli used in the main experiment.

**Procedure**

Seated comfortably in front of a computer monitor with a keyboard placed in their lap, participants were presented the experiment running in PyEPL featuring a black screen with white text. At times when sounds were being presented, participants were asked to fixate on a white cross displayed in the middle of the computer monitor to reduce ocular and muscular artifacts during EEG recording. As a reminder, subjects were presented with examples of
alternating clicks to demonstrate the type of stimuli used in the experiment. After, the following instructions were shown to each subject:

Thank you for participating in our study on meter! Your task is to listen to sequences of clicks and to detect a click that sounds different from the other ones. While listening to the sequences, you will be asked to group sequences in either groups of TWO or THREE. Types of groups will happen in trial blocks: Some blocks will have all TWO groupings, and others will have all THREE groupings.

You will be instructed what type of grouping to use for each trial block. To initiate a trial block, you will be prompted to press any key on the keyboard. First, you will be presented with a fixation dot in the center of the screen. Please focus on the fixation dot for all trials.

Each trial begins with a sequence consisting of either 6 or 9 beeps. This will help you remember how to group the clicks in the sequence. The louder beep signifies the 1st beat of the grouping. After the priming sequence finishes, a short span of time will pass. The click sequence will then start playing after the priming sequence. Group the clicks that you hear in the same way the priming sequence grouped them: in either groups of TWO or THREE.

Listen carefully for the different click and remember what position in the groupings the deviant stimulus occurred. Wait until the sequence completes.

If the different click happened on the first position of the group, press 1 on the keyboard keypad. If the different click happened on the second position, press 2 on the keyboard keypad. If on the third position, press 3.

After you respond to what position the different click occurred, the next trial will start again with the priming sequence.

As described, the participants were to listen to the sequences and organize the onsets into a temporal structure in order to impose a metric hierarchy. Two types of meter were used: binary and ternary (See Figure 7). Binary meter involved organizing the sequences into periodic alternating strong and weak beats, described as 1 (strong) and 2 (weak) to the participant. Ternary meter consisted of a strong beat followed by two weak beats, defined to each subject as beats 1, 2, and 3 respectively. The subjects’ task was to listen to each
sequence and organize it using a specific instructed meter, and listen for a click pair that sounded different from the context clicks. In a forced-choice paradigm, subjects were to report the metric position—1 or 2 in binary or 1, 2, or 3 in ternary—in which they detected a different sounding click pair. Participants were given five seconds after each trial in order to respond. If they did not respond within this time range, another trial was presented.

Each session consisted of three practice blocks and ten trial blocks. The practice blocks were intended to expose each participant to the stimuli sequences and familiarize them with responding and listening, and were randomly determined binary or ternary conditions. After the practice blocks, participants were able to initiate the start of each of the eight trial blocks. Each trial block consisted of twelve consecutive sequences where listeners used one type of meter, either binary or ternary, to organize the sequences. Five binary blocks were randomly presented alongside five ternary blocks. Before each trial block began, participants were instructed of the type of meter to use.

As a further reminder, 100 ms sine tones with 10 ms rise and fall times were used to prime each subject before each trial. The sine tones outlined strong and weak beat binary or ternary structure, and used 720 Hz tones as strong beats and 440 Hz tones as weak beats. Six priming beeps, three alternating strong-weak beat pairs were used for binary primes; nine priming beeps with three alternating strong-weak-weak groups were used as ternary primes. After each priming sequence, a jitter of silent space was placed before click sequences began to ensure that the start of click pairs were heard to be strong beats (Povel & Okkerman, 1981;
Thomassen, 1982) and not continuous of each prime.

Overall, each experiment consisted of 120 sequences broken down into two types of metric organization and two types of contexts. After each trial block, participants were given the opportunity to take a short break. The experiment lasted roughly 60 minutes depending on break durations of participants. In combination with the screening experiment, and EEG system setup, participants devoted roughly 180 minutes.

In sum, participants were asked to organize equally-spaced click sequences in periodic groups of either two or three. Each sequence contained 22 clicks, and participants listened for a deviant stimulus somewhere in the sequence. Participants then reported the metric position in which a deviant stimulus occurred. To report the deviant stimulus, participants responded using the keyboard keypad by pressing 1, 2, or 3 corresponding to each beat in which a deviant click might be presented.

*Preprocessing and Analysis – Preliminary ERP Data*

As the ActiCap system does not have a dedicated channel for online referencing, all neural data was re-referenced offline to linked mastoid electrodes. Individual neural data analysis was performed using Python Time Series Analysis (PTSA). A high-pass filter with a .25 Hz cut-off was implemented to remove any low frequency artifacts. All data was baseline
corrected from -100 to 0 ms. An additional 30 Hz low-pass filter was used to smooth the ERP waveform.

*Full ERP Analysis of Strong and Weak Beats*

In order to examine strong and weak beat differences for all subjects at all electrode sites during the first 12 onsets of clicks, EEG signals were epoched from -250 ms to 600 ms around strong and weak beats. Signals in this range were baseline corrected from -250 ms to 0 ms around events of interest and downsampled to 100 Hz.

An independent samples *t* test examined distributions of voltage differences around strong and weak beats for all EEG channels and converting the p-value to a z-score. This generated a matrix of z-scores for each channel. For each individual participant, a non-parametric bootstrap procedure shuffled strong and weak beat events and compared them again using independent samples *t* tests for a number of 500 shuffles. Z-score matrices were then summed for all participants for both matrices of 500 shuffles and un-shuffled actual data. Threshold-free cluster enhancement was applied, increasing the peak of each cluster which produced a value that was saved in a cluster-enhanced matrix. The actual EEG data was compared to this cluster-enhanced matrix. A 95th percentile threshold was set to compare values in which actual data needed to exceed.
Chapter 3: Results

Screening Experiment Results

Results from the screening experiment indicate that no thresholds dipped below 11 ms. The average threshold was 11.5 ms (SD = 0.2) suggesting that results were consistent across participants. A Pearson product moment correlation indicated no significant correlation between OMSI scores and timing separation (r = -.5, n = 11, p > .05).

Main Experiment Results

Behavioral Results

Findings from the main experiment indicate that overall performance was at chance for correctly identifying the metric position of a deviant stimulus. A 2 x 3 factorial ANOVA examined between-subjects differences of musicians and non-musicians; within-subjects variables included strong and weak beats nested in variables of binary and ternary meter. Context variables described deviant ticks embedded within a tock sequence or vice versa. Proportion of correct responses for all conditions can be found in Appendix B. As depicted in
Figure 8, significant differences did not arise between musicians and non-musicians ($F(1,19) = 3.072, p > .05$). A main effect of meter indicated that performance in the binary conditions was better than in the ternary condition ($F(1,19) = 9.8, p < .05$), as demonstrated by Figure 9. Accompanied by Figure 10, a significant main effect of context indicated that performance was better for both groups when a deviant Tock was embedded within a context of Ticks ($F(1,19) = 5.39, p < .05$). A main effect of beat was found ($F(2,19) = 8.34, p = .105$), and a post hoc Tukey HSD test showed significant differences of performance on beat three to beats one and two across both metric conditions. However, the post hoc test did not indicate differences between beats one and two. Beat-related differences are depicted for both binary and ternary meter in Figure 11. A significant three-way interaction arose between ternary meter on beat three in the non-musician group, ($F(1,19) = 4.17, p < .05$). No other significant simple effects or interactions were found.
Figure 8. Overall behavioral performance of musicians and non-musicians.
Figure 9. Proportion of correct responses for binary and ternary meter.
Figure 10. Proportion of correct responses for context conditions.
Figure 11. Proportion of correct responses on strong and weak beats in binary, and proportion of correct responses for strong and both weak beats in ternary.

In sum, group differences did not arise between musicians and non-musicians, and performance in binary meter conditions was higher than in ternary meter conditions. Context differences arose, with Tick contexts yielding a better performance than Tock contexts. A significant main effect of strong and weak beats overall, and a significant difference was found in a post hoc test for beat three in ternary versus beats one and two collapsed across binary and ternary conditions. No significant simple effects or interactions were found.
EEG Results

Preliminary ERP waveforms for four participants (two musicians and two non-musicians) are presented below in Figures 12-15. The waveform plots imagined strong beats as red traces and weak beats as blue traces for the continuation sequence that precedes the region in which a deviant stimulus would occur. The comparison should be noted at time points 0.0 to 0.6 along the abscissa. Voltage is on the ordinate, with positive voltages pointing upward. This region represents the area where endogenously-maintained meter for both binary and ternary is imposed upon each trial sequence.
Figure 12. ERP waveform for Subject 1 in the musician group examining strong and weak beats.
Figure 13. ERP waveform for Subject 2 in the musician group examining strong and weak beats.
Figure 14. ERP waveform for Subject 3 in the non-musician group examining strong and weak beats.
Results overall show mixed differences in processing strong and weak beats. For musician Subject 1, a difference in strong and weak beat processing is evident in the P300 complex, around the time point .3 s. Larger amplitudes in this component are found for strong beats over weak beats. An opposite effect is noted for the second musician, Subject 2. Larger amplitudes were found for weak beats over strong beats around .15 s post-stimulus. Overall
for the non-musicians, it appears differences in strong and weak beat processing are evident, but in opposing directions. For non-musicians, some evidence for enhanced strong beat processing arises. In Subject 3, an amplitude difference in strong beats arises first a few milliseconds after stimulus onset, which may be systematically related to differences in amplitude for strong beats around .17 s post-stimulus. For Subject 4, there appears to be no differences in strong and weak beat processing due to the amount of variance; although average traces may lean toward enhanced strong beat processing, an effect cannot be observed. Note that differences for all waveforms are not statistical differences as results are still preliminary.

In examining results from comparing the cluster-enhanced matrices of z-score transformed shuffled voltage distributions for all subjects at all electrode sites to the collected EEG data for conditions, electrode and time values did not exceed a threshold of the 95th percentile, indicating that there were no significant differences found between strong and weak beats.
Chapter 4: Discussion and Implications

Together, results show that musicians and non-musicians did not perform differently. Binary meter yielded better performance in this task than did ternary meter, and beat-based differences arose between beats three in ternary meter versus beats one and two combined in binary meter. The context of deviant stimuli within sequences of Ticks and Tocks produced a surprising effect, with Tocks embedded within a series of Ticks appearing to be better detected than its opposing context. Preliminary neural analysis demonstrate beat-based differences at one electrode site that arise mostly in musicians, but individual differences show opposite processing effects of strong and weak beats. However, a follow-up analysis compared strong and weak beat differences across all participants for all electrode sites and showed no significant differences between these two conditions.

Discussion of Screening Experiment Results

Results from the temporal order acuity task used in screening experiment indicated that performance across all participants was consistent. Average timing separation did not escape 11 ms, suggesting that all participants performed similarly. However, performance hovered around the ceiling of permissible stimulus magnitude. It is possible that this ceiling effect is
not yielding a valid 79% performance. Note that timing separation in the screening experiment was not to exceed 12 ms, as this was the point it is thought that click pairs start to disintegrate into two sounds (Joliot, Ribary, & Llinás, 1994).

Discussion of Main Experiment – Behavioral Results

Although overall performance rates were above chance, proportion correct for all conditions in reporting deviant stimuli were low. One possibility is that the salience of the stimuli was not strong enough to reliably yield a high detection rate in sequences. Resolving Ronken click pairs appears to be a difficult task overall, due to the near-ceiling timing differences of exhibited by all participants. In order to make deviant stimuli more salient, a pitch discrimination task (as used by Jones et al., 2002) or a click omission paradigm (a modification of experiments used by Potter et al., 2009) may be more suitable for examining clear meter-related differences that may modulate auditory perception.

However, the motivation behind the study was to examine changes in auditory perception that might be facilitated by structured hierarchies of temporal predictions. A difficult task, such as those used here involving temporal order acuity, may still be best suited as it requires high amounts of attentional energy to resolve small differences in temporal structure of brief sounds. An alternative approach to increase overall performance may be to employ a training paradigm that is traditionally used in psychophysical experiments to examine acuity (see Green, 1971). This would give listeners more exposure and experience in resolving small
timing judgments over repeated sessions, which would then be followed by the current experimental paradigm. Such training might make the stimuli more distinguishable in a rhythmic sequence, and ultimately yield distinct differences when stimuli are presented on strong and weak beats for both musician and non-musician groups.

A finding inconsistent to the second hypothesis was that musicians did not outperform non-musicians at this task. Although negative results are difficult to interpret, it can be suggested that domain-specific training as long-term musical experience does not provide advantages in using attention focused at specific points in time to resolve small timing differences. Moreover, OMSI scores did not correlate with proportion correct during the main experiment \( (r = -.07, p > .05, \text{two-tailed}) \) nor did the scores correlate with timing separations used in the screening experiment \( (r = -.5, p > .05, \text{two-tailed}) \). Taken with results from the main experiment, it appears that musician and non-musician differences did not factor significantly in temporal acuity in the current study.

Consistent with hypothesis three, meter-related differences arose between binary and ternary conditions. Abecasis et al., (2005), Potter et al., (2009), Fujioka et al., (2010), and Nozaradan et al., (2011) all found differences in deviant-related processing between binary and ternary meters. Subjective accenting studies with ERPs suggested that binary meter was a default metric organization, and may be easier to maintain endogenously when organizing rhythmic sequences. Consistent with these findings, binary meter yielded a better performance than ternary meter. One possible account for these differences is that ternary meter requires more
effort to maintain, detracting from resources required to detect small differences in timing during this task.

Another interpretation of this finding might be an artifact of the probability of a deviant click occurring within the sequence in both binary and ternary meters. As each trial required a forced response, it is almost certain that an amount of guessing influenced the performance rates that were acquired. In binary meter, the probability of correctly guessing the correct metric position is .5 for both strong and weak beats. In ternary meter, the probability of a correct guess is .333. Thus, the lower probability of correctly guessing the metric position in ternary conditions might give rise to the differences in binary and ternary meter overall. It is plausible then to reinterpret the results using scores that are corrected for guessing. Such a procedure would entail subtracting a value of one from the number of wrong responses divided by the number of possible guessing options, then taking this value and subtracting it from the number of correct responses (Frary, 1988). Corrections for guessing would attempt to remove error introduced by the forced choice paradigm. However, a problematic assumption made during corrections for guessing is that incorrect responses are the same as guessing. If neural representations of a metric type can predict what beat position a deviant stimulus occurred, then the incorrect response assumption by guess corrections can undermine the neural predictions. Without further neural analysis, however, implementation of a correction for guessing might reveal a more accurate set of behavioral responses.

A main effect of beat indicates that beat-based differences arose across both metric
conditions. To examine specific differences, a post hoc test revealed significant differences between beats 3 and beats 1 and 2 across all conditions, but this difference collapsed over binary and ternary meter. Statistically, results are consistent with the fourth outlined hypothesis. Compatible with probabilistic issues discussed above in meter-related differences, the lowest probability metric position that may have occurred for listeners in the task was deviant stimuli presented on beat three of ternary meter. Confirming this notion, deviant stimuli on beat three arose during only 16% of all trials. This presentation difference may account for lower performance rates for this sub-condition. Moreover, the use of partial eta squared to examine effect size of this finding was found to be .013, meaning that the ratio of effect size to the effect size in conjunction with error variance is very small; in other terms, only a little over 1% of the overall variance is explained by beat-based differences.

Taken as a whole, support for hypothesis four is spurious at best. Previous behavioral research suggested that the alignment of pulse levels extracted from stimulus regularities directs attention-driven hierarchical neural oscillations to specific points in time. Aligned pulses narrow attentional focus to events that encompass a predicted instance, and would potentially enhance perception of sounds that occur when they are anticipated. For this specific task, it was hypothesized that enhancements of perception would be observable in a listener’s temporal acuity. However, results do not support the predictions derived from Snyder and Large’s (2009) resonance model or Large and Jones (1999) theory of dynamic attending.
Although not to be taken with confidence, a trend in beat-based differences in ternary meter suggest that stronger beats yield better performance in deviant click detection than weak beats in non-musicians. This is not consistent in the binary condition. This trend follows findings in Fujioka, Zendel, and Ross (2010), where beat based-differences did not arise in amplitude and latency in binary meter, but did arise in ternary meter when ERFs were calculated. Further collected results continuing on this trend may lend further support for beat-based differences.

In examining Tick and Tock contexts, a significant difference was found between performance in locating a deviant Tock embedded within a series of Ticks, and a deviant Tick embedded with a series of Tocks. An asymmetrical relationship was found, with the former condition yielding a higher proportion of correct responses. This surprising effect may be explained by an asymmetry of temporal masking (Resnick & Feth, 1975). The asymmetry arises due to differences that occur between the two clicks from forward and backward masking. If a smaller pulse precedes a larger pulse (Tocks), the backward masking effect exhibited by the larger pulse is less pronounced than the reverse, where forward masking occurs due to the the larger pulse precedes the smaller pulse (Ticks). The study by Resnick and Feth confirmed the purported asymmetry in Ronken-style click pairs, with the authors varying the intensity of individual pulses as well as their timing separation. It is plausible that the effect is observable in the present study, with a less-masked signal (Tock) being easier to detect.
Discussion of Main Experiment – Neural Results

Preliminary analysis of ERP waveforms reflecting beat types in the standard sequence (the 12 onsets preceding the target sequence which contain deviant stimuli) provide mixed results. In musician Subjects 1 and 2 at a centrally-located parietal electrode, a reverse effect was found for the differential processing of strong and weak beats. Amplitudes of strong beats seemed to be enhanced during endogenously-maintained strong beats for Subject 1, whereas amplitudes for weak beats were enhanced for Subject 2. Although preliminary, these observations invite considerations of individual differences that may underlie imagined strong and weak beats. Such considerations may carry over to observations of ERP waveforms in the non-musician Subjects 3 and 4. Subject 3 exhibited some amplitude differences for strong beats, whereas this difference was not noted for Subject 4. Based on these results, no solid conclusions can be inferred.

A final analysis did not find support for hypothesis one, which is related to the prediction of responses in reporting deviant stimulus locations based upon strong and weak beat representations evidenced in preceding rhythmic sequences. In this analysis, strong and weak beat conditions were examined across all electrode sites in EEG recordings for all subjects. No significant differences were found in ERP waveforms when examining conditions of strong and weak beats across both binary and ternary meter during the sequences preceding the target region containing the deviant stimulus. Since no differences were found for strong and weak beat processing, mental representation of meter was not able to be found. Thus, it is
unable to make a prediction of listeners’ response from this representation in this present analysis.

This finding aligns with strong and weak beat differences that were not found in Nozaran et al. (2011), where only mental representation of meter was found in ERP amplitude differences that occur at a subharmonic divisions of the metric conditions of binary and ternary. Moreover, an MEG study by Fujioka et al. (2010) found strong and weak beat differences localized to specific brain divisions only in ternary meter. This suggests that a further analysis separating binary and ternary meter might elicit differences in in metric positions, and collapsing across both metric types obscured evidence for a mental representation of meter.

ERP studies on subjective accented only found strong and weak beat differences in ERP amplitude and latency during presentations of deviant stimuli (e.g., omission of a tone) of P300 components (Brochard et al., 2003; Abecasis et al., 2005; Potter et al., 2009; Schaefer, Vlek, & Desain, 2011). Thus, it may be possible to corroborate these findings by examining responses during deviant stimuli in all conditions. Further analysis of strong and weak beat processing during deviant click presentations may yield similar results, but is still inconsistent with hypothesis one.
Future Directions

An obvious first step is to continue with neural analysis. Here, ERP waveforms were the only analysis used. As discussed, separation of binary and ternary metric conditions might lend evidence for representations of meter. Collapsing strong and weak beats into one analysis may not be appropriate for looking at representation of meter, and it is possible to examine ERP waveforms across one period of each type of meter. For example, a time window of 1800 ms would encompass one period of ternary meter, and a period of binary meter would elapse over 1200 ms. Differentiation of these two metric types might support the plausibility of mental representation of metric type. A second step is to examine ERP responses to strong and weak beats that occur during the presentation of deviant stimuli. Although overall performance rates are low, it might be possible to compare correct responses in reporting metric positions by using ERP waveforms that elicit clear deviant detection responses, such as pronounced P300 components.

As discussed, using ERP analysis to examine neural correlates of meter is only one way to approach signals recorded from the EEG. Time frequency analysis, in which power and phase spectra can be examined for activity in different frequency bands of neural activity, might be more apt for investigating metric representations. Specifically, may be possible to find relationships of spectral power and phase coherence to meter or beat conditions. One might expect that predictive timing ramps revealed by Fujioka et al. (2012) to arise in the current analysis; gamma activity that reflects anticipatory processes to encode stimulus
timing properties is also a candidate for metric representation (e.g., Snyder & Large, 2005). Thus, ERP analysis may only reveal part of the story involved in neural activity underpinning time predictions established by metric hierarchies.

Implications and Conclusion

In light of the findings, prospective results are nevertheless implicated in understanding mechanisms that might represent time-based relationships of speech and music. Broadly, speech and music contain periodicities that extend through time, and by organizing these regularities using top-down mechanisms, our perceptual experience of sounds is guided. By continuing to collect both behavioral and neural data, a better conceptual understanding of predictive mechanisms may emerge. Specifically, mental representations of meter that arise as differentially processed strong and weak beats--maintained solely endogenously--are thought to reflect those higher level mechanisms that merge with bottom-up processing. Thus, further analysis aims to contribute to understanding the cognitive processes underlying temporal predictions and attention, and extend the knowledge that inform models of dynamic attending and neural resonance that are built from the properties of regularities in complex sound structures.
Bibliography


Dennis, M. et al. (2009). Strong-Meter and Weak-eter Rhythm Identification in Spina Bifida


Green, D. M., Birdsall, T. B., & Tanner, W. P. (1957). Signal detection as a function of


189-192.


Psychological Research, 66, 3-17.


Münte, T.F., Altenmüller, E., & Jäncke, L. (2002). The musician’s brain as a model of


Oh, Y. (2009). Stimulus presentation software coded in MATLAB [computer software]. Columbus, OH.


early, temporally late stages of visual processing. *Brain, 124*, 1417-1425.


Appendix A: Ollen Musical Sophistication Index

1. How old are you today?
   _____ age in years

2. At what age did you begin sustained musical activity? “Sustained musical activity” might include regular music lessons or daily musical practice that lasted for at least three consecutive years. If you have never been musically active for a sustained time period, answer with zero.
   _____ age at start of sustained musical activity

3. How many years of private music lessons have you received?
   If you have received lessons on more than one instrument, including voice, give the number of years for the one instrument/voice you've studied longest.
   If you have never received private lessons, answer with zero.
   _____ years of private lessons

4. For how many years have you engaged in regular, daily practice of a musical instrument or singing? “Daily” can be defined as 5 to 7 days per week. A “year” can be defined as 10 to 12 months. If you have never practiced regularly, or have practiced regularly for fewer than 10 months, answer with zero.
   _____ years of regular practice

5. Which category comes nearest to the amount of time you currently spend practicing an instrument (or voice)? Count individual practice time only; not group rehearsals.
   __ I rarely or never practice singing or playing an instrument
   __ About 1 hour per month
   __ About 1 hour per week
   __ About 15 minutes per day
__ About 1 hour per day
__ More than 2 hours per day

6. Have you ever enrolled in any music courses offered at college (or university)?

__ No (Skip to #8)
__ Yes

7. (If Yes) How much college-level coursework in music have you completed? If more than one category applies, select your most recently completed level.

__ None
__ 1 or 2 NON-major courses (e.g., music appreciation, playing or singing in an ensemble)
__ 3 or more courses for NON-majors
__ An introductory or preparatory music program for Bachelor’s level work
__ 1 year of full-time coursework in a Bachelor of Music degree program (or equivalent)
__ 2 years of full-time coursework in a Bachelor of Music degree program (or equivalent)
__ 3 or more years of full-time coursework in a Bachelor of Music degree program (or equivalent)
__ Completion of a Bachelor of Music degree program (or equivalent)
__ One or more graduate-level music courses or degrees

8. Which option best describes your experience at composing music?

__ Have never composed any music
__ Have composed bits and pieces, but have never completed a piece of music
__ Have composed one or more complete pieces, but none have been performed
__ Have composed pieces as assignments or projects for one or more music classes; one or more of my pieces have been performed and/or recorded within the context of my educational environment
__ Have composed pieces that have been performed for a local audience
__ Have composed pieces that have been performed for a regional or national audience (e.g., nationally known performer or ensemble, major concert venue, broadly distributed recording)

9. To the best of your memory, how many live concerts (of any style, with free or paid admission) have you attended as an audience member in the past 12 months? Please do not include regular religious services in your count, but you may include special musical productions or events.
10. Which title best describes you?

__ None
__ 1 - 4
__ 5 - 8
__ 9 - 12
__ 13 or more

109
Appendix B: Proportion correct of behavioral responses for all experimental conditions.

<table>
<thead>
<tr>
<th>Event</th>
<th>Condition</th>
<th>Group</th>
<th>Proportion Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Conditions</td>
<td></td>
<td></td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.54</td>
</tr>
<tr>
<td>Meter</td>
<td>Binary</td>
<td></td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Ternary</td>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.48</td>
</tr>
<tr>
<td>Beat</td>
<td>Strong</td>
<td></td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Weak</td>
<td></td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Binary Strong</td>
<td></td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Binary Weak</td>
<td></td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>Ternary Strong</td>
<td></td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Ternary Weak 2</td>
<td></td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Ternary Weak 3</td>
<td></td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>Context</td>
<td>Tick</td>
<td></td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Took</td>
<td></td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Musicians</td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Non-musicians</td>
<td></td>
<td>0.5</td>
</tr>
</tbody>
</table>