

5-1-2012

Modality-Specific and Modality-General Encoding of Auditory and Visual Rhythms

Amanda Pasinski
University of Nevada, Las Vegas

Follow this and additional works at: <https://digitalscholarship.unlv.edu/thesesdissertations>



Part of the [Cognition and Perception Commons](#)

Repository Citation

Pasinski, Amanda, "Modality-Specific and Modality-General Encoding of Auditory and Visual Rhythms" (2012). *UNLV Theses, Dissertations, Professional Papers, and Capstones*. 1608.
<http://dx.doi.org/10.34917/4332589>

This Thesis is protected by copyright and/or related rights. It has been brought to you by Digital Scholarship@UNLV with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Thesis has been accepted for inclusion in UNLV Theses, Dissertations, Professional Papers, and Capstones by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.

MODALITY-SPECIFIC AND MODALITY-GENERAL ENCODING
OF AUDITORY AND VISUAL RHYTHMS

by

Amanda Claire Pasinski

Bachelor of Arts
University of Nevada, Las Vegas
2007

A thesis document submitted in partial fulfillment
of the requirements for the

Master of Arts in Psychology

Department of Psychology

College of Liberal Arts

The Graduate College

University of Nevada, Las Vegas

May 2012



THE GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Amanda Claire Pasinski

entitled

Modality-Specific and Modality-General Encoding of Auditory and Visual Rhythms

be accepted in partial fulfillment of the requirements for the degree of

Master of Arts in Psychology

Department of Psychology

Joel Snyder, Committee Chair

Erin Hannon, Committee Member

Jefferson Kinney, Committee Member

Eugenie Burkett, Graduate College Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

May 2012

ABSTRACT

Modality-Specific and Modality-General Encoding of Auditory and Visual Rhythms

by

Amanda Claire Pasinski

Dr. Joel Snyder, Examination Committee Chair
Associate Professor
University of Nevada, Las Vegas

The perception of timing information plays a large role in our everyday activities, yet we still do not accurately understand the mechanisms underlying these perceptions. Both modality-general and modality-specific mechanisms have been suggested to account for perceptual timing. The use of a new auditory tempo perception paradigm can be used to examine various brain responses - measured via electroencephalography (EEG) - thought to index timing perception. This study applied this paradigm to both auditory and visual rhythms, and compared event-related potentials (ERPs) to task performance. Auditory and visual contingent negative variation (CNV) components showed two distinct voltage patterns across the scalp: The auditory CNV appears to show contributions from temporal areas, while the visual CNV appears to show contributions from occipital areas. There were larger CNV amplitudes in the auditory modality than in the visual, suggesting the CNV indexes modality-specific processing. A late, memory-dependent positive-voltage component did not show these modality-related topographical or amplitude differences, and instead reflects modality-general processing. This suggests timing information is encoded intrinsically at a sensory level, and this information is then routed to a cognitive, decision-making area for further processing.

TABLE OF CONTENTS

ABSTRACT	iii
LIST OF FIGURES	v
CHAPTER 1 INTRODUCTION AND REVIEW OF RELATED LITERATURE	1
Purpose of the Study	1
Intrinsic Versus Dedicated Timers.....	2
Evidence for Dedicated Timers.....	3
Evidence for Intrinsic Timing.....	8
Intermediate Views.....	13
Auditory and Visual Differences	15
Beat Versus Interval Strategies and the Current Paradigm	19
Sensory-Evoked Potentials	22
The Contingent Negative Variation as an Index of Timing	26
Novelty and Context-Updating P3	30
Possible Sources and Generators for Event-Related Potentials.....	33
Motivations for the Current Study	35
CHAPTER 3 MATERIALS AND METHODS.....	37
Participants	37
Materials and Procedure	37
Electrophysiological Recording.....	39
Data Analysis	40
CHAPTER 4 RESULTS	44
Behavioral Data.....	44
Electrophysiological Data	48
CHAPTER 5 DISCUSSION	57
Discussion of Results	57
APPENDIX 1 IRB APPROVALS	62
REFERENCES	64
VITA	72

LIST OF FIGURES

Figure 1 Five-Tone and Four-Tone Patterns.....	38
Figure 2 Proportion of “Speeding Up” Responses by Final Interval.....	46
Figure 3 Proportion of “Speeding Up” Responses for modality and type interactions....	47
Figure 4 ERP Traces and Topographies for Auditory and Visual CNVs.....	52
Figure 5 Mean amplitude across the 68 electrodes for CNV.....	53
Figure 6 ERP traces and topographies for Auditory and Visual P3s	55
Figure 7 Mean amplitude across the 68 electrodes for P3.....	56

CHAPTER 1

INTRODUCTION AND REVIEW OF RELATED LITURATURE

Time perception occurs while listening to the ‘tick-tock’ of a clock, the meter of music, the spacing in speech, and even while watching the back and forth play of a tennis match. Perception of timing information plays a large role in our everyday activities, yet we still do not understand what mechanisms underlie these perceptions. A large corpus of research has been conducted on the nature of timing information and whether modality-specific or modality-general processes can account for our perception of rhythm.

A beat is the perception of a regular, periodic series of events. If imagining a metronome, the ‘clack’ that sounds each time the arm hits the side is an onset, or event; the time between onsets as the arm travels to the other side is the period; and the temporal expectancies for each successive onset form the beat. The beat can be perceived in line with the rate of physical event onsets, or can be perceived to a multiple or fraction of the rate of onsets. Deviations from an external rhythm result in phase and period corrections, where phase refers to the relationship between successive events in time and period refers to the intervals between these events. Therefore, phase correction acts on the asynchronies or differences in events, while period correction acts on the differences between successive intervals (Semjen, Vorberg, & Schulze, 1997). These complex corrections take place seemingly without effort during everyday tasks, and a large debate concerns whether this is due to the contributions of dedicated neural centers or circuits, or if the information necessary is inherent in the firing of neural populations.

Intrinsic Versus Dedicated Timers

Studies examining timing processes generally distinguish supra-second from millisecond range intervals, motor timing tasks from non-motor timing tasks, and sequences of intervals from isolated intervals. Traditionally, one of the major theoretical debates concerns whether timing mechanisms are best thought of as computing time intervals like a stopwatch or synchronizing to temporal patterns like an oscillator, which will be discussed later. More recently, however, a major theoretical focus has been on whether temporal information is inherent in the neural dynamics of modality-specific areas, or if there are dedicated areas or circuits in the brain that control the processing of time. A recent review by Ivry and Schlerf (2008) examined the arguments concerning dedicated and intrinsic timers. In dedicated models, all modalities draw information from a central timing module; the cerebellum (which is important for matching movement with events and pairing events in time to consequences) and basal ganglia (a possible gatekeeper mechanism and an area important to long-interval discrimination) are often implicated as specialized timers. Dedicated timers could also be distributed across the cortex, and the supplementary motor area (SMA) and dorsolateral prefrontal cortex (DLPFC) appear to be likely candidates for distributed timers. Lesion, split-brain patient, functional magnetic resonance imaging (fMRI), and transcranial magnetic stimulation studies have shown these areas and others to be important to timing; however, this may be confounded by the cerebellum's contributions to motor learning, error detection, or sensory integration (Penhune, Zatorre, & Evans, 1998), or because these areas merely represent pathways in timing circuits.

Because these dedicated brain areas may in fact be processing other information, intrinsic models argue that these areas are not necessary for processing timing. Instead, these models suggest that timing is inherent in the neural dynamics of the cerebral cortex, including primary sensory areas. This leads to the idea of modality-specific timing: visual neurons might be responsible for visual timing, and auditory neurons for auditory timing. A possible mechanism for intrinsic timing comes from the state dependent network model (SDN), which argues that temporal processing is inherently encoded in the state of neural networks (Karmarkar & Buonomano, 2007). While intrinsic models are verifiable via computational models, they appear to be limited to very short intervals; these models may require input from the more cognitive-related dedicated memory areas discussed below to process longer intervals.

Evidence for Dedicated Timers

The cerebellum and basal ganglia have been implicated in both time perception and production tasks (see Hazeltine, Helmuth, & Ivry, 1997 for a review). Grahn and Brett (2007) used a temporal reproduction task to examine possible brain areas involved in beat perception, and their fMRI results implicated the basal ganglia and supplementary motor area (SMA). They presented participants with sequences that were metrically simple, complex, and non-metric, and asked participants to replicate each sequence. They tested both musicians and non-musicians, and found that musicians showed greater activity in the pre-SMA/SMA, right premotor cortex, and right and left cerebellum areas than non-musicians for all types of sequences. This was mirrored with greater accuracy in the behavioral responses of the musicians as compared to the non-musicians; the greater

activation in these areas corresponded to better temporal acuity, and therefore must contribute to timing perception and reproduction.

However, finding true neural correlates of timing is made difficult by confounding factors: Is brain activity in these areas due to interval encoding, decision-making, sensory-motor activation, or mnemonic storage of intervals? Harrington and colleagues examined differences in brain activation during both interval encoding and decision-making (Harrington, Boyd, Mayer, Sheltraw, Lee, Huang, & Rao, 2004). Participants were asked to compare two intervals; the standard interval occurred first, and was either 1200 or 1800 ms, and was followed by a variable comparison interval. Image acquisition was synchronized to the onset of the first tones of both the standard and comparison intervals. An additional 6 images per trial were collected, 42% of which were either blank or resting scanned intervals collected at the end of the trial that served as a baseline control. Image acquisition at 2 and 4 seconds post-trial were expected to reveal activation patterns indicative of interval encoding, while 8 and 10 second scanning intervals were expected to reveal processes specific to decision making (based on response time and motor cortex activity due to button presses).

For the encoding phase, activity in the left medial frontal cortex and right pre-SMA regions dipped below baseline, and returned to baseline after the end of the trial. However, activity in the right medial frontal cortex and anterior and posterior cingulate areas, bilateral superior temporal cortex, areas of the parietal cortex, and lobules VI and VII of the cerebellum increased above baseline. For the decision phase, greater activation for difficult decisions was found in the left medial frontal cortex, bilateral prefrontal cortex, left superior temporal cortex, left superior and inferior parietal cortex, left middle

occipital cortex, and the putamen and nucleus accumbens. For the easy discriminations, greater activity was found in the right parahippocampus. This supports the theory that basal ganglia, cerebellum, and cortical areas modulate interval encoding; however, this activation correlated with behavioral measures in only the right caudate, right inferior parietal cortex and precuneus, right parahippocampus and hippocampus, and left cerebellum, suggesting a right hemisphere bias for temporal processing.

However, the search for true dedicated timers goes beyond simple encoding or decision-making. A review by Lewis and Miall (2006) argues that the mechanisms needed for timing vary depending on whether the task uses short or long intervals, requires motor or non-motor timing, or whether the task presents continuous sequences or broken, irregular events. They assert that tasks involving continuous monitoring and motor involvement regarding a predictable sequence (such as rapidly paced finger tapping) recruit primary sensorimotor and premotor areas and are highly automatic. Tasks requiring monitoring over longer periods without isochronous sequences or motor involvement tend to recruit right hemispheric prefrontal and parietal cortices and require more cognitive control. These tasks form a bridge between the narrow range of millisecond timing and the well-understood range of circadian rhythms. This also suggests that smaller dedicated timing areas might fall under the cognitive control of a higher brain area.

The area most commonly attributed to cognitively controlled timing is the dorsolateral prefrontal cortex (DLPFC); it has been reliably associated with working memory in both animal and human research, and lesion, transcranial magnetic stimulation (TMS), neuroimaging, and Parkinson's disease studies have implicated this

area in timing. Lewis and Miall (2006) argue that the basal ganglia could provide a direct circuit to the prefrontal cortex. The basal ganglia could act as a clock or a counter that accumulates and stores a series of neural pulses; the number of pulses stored could then represent an interval. Neurons in the striatum could then communicate this interval information to the DLPFC, which could then act as a memory component of timing (see Church and Broadbent, 1990 for an animal review).

This pulse-accumulator system is not limited to the auditory modality; a study by Jin, Fujii, and Graybiel examined the connections between basal ganglia and cortical circuits in monkeys (2009). They inserted multiple electrodes into the dorsolateral prefrontal cortex (DLPFC) and the caudate nucleus (CN) of macaque monkeys, and presented them with a visual timing task that did not require the learning of precise intervals. Out of the total number of neurons studied, some 70 DLPFC neurons produced phasic responses for both visual and motor modalities, while the CN contained about 16 such neurons. Even in a passive looking condition, prefrontal and striatal neurons formed time-stamped representation of the visual and motor events. While this shows the importance of dedicated areas because similar activity occurred for both visual and motor tasks, it also suggests that temporal encoding may be a natural part of stimulus and task encoding, at least for short time intervals.

The importance of higher cortical areas suggests that attention might play a crucial role in modulating temporal processing. Noguchi and Kakigi used MEG to examine the cortical representations of a temporal visual stimulus (2006). Due to the powerful nature of attention, they predicted that subjective duration of a visual stimulus would be more closely associated with neural activity in higher-level visual areas. Their

stimuli consisted of either a “<” or “>” symbol followed by the same or different symbol (e.g. “< delay <” or “< delay >”). The first presentation of the stimulus represented the standard interval and was sustained for 600 ms. The second presentation of the stimulus represented the comparison and varied within ± 100 ms of the standard, including a duration equal to the standard. They found a slow buildup of neural activity in the frontocentral areas, consistent with previous findings of an accumulator representation in the SMA. This may also represent a contingent negative variation (CNV); the CNV has been implicated in attention-demanding timing tasks and reflects expectancy (Walter, Cooper, Aldridge, McCallum, & Winter, 1964; Pfeuty, Ragot, & Pouthas, 2003). Noguchi and Kakigi also found a bias towards reporting “shorter” in the same condition; this cannot be explained by chronostasis from saccades, but the “greater than” and “less than” symbols may have lead to subjective differences because of their associated meanings. The response bias was predicted by the authors to arise from weaker activity in higher visual areas for the “same” conditions; this weaker activity could lead to less output to the accumulator and therefore shorter subjective duration judgments for “same” than “different” conditions. While the authors tried to explain their results in terms of a clock model, this weaker activation of visual areas suggests at least partial intrinsic activity. If the visual system is less adept at temporal encoding, and this weaker activation can be seen through different duration judgments, this may suggest that temporal information is inherent at modality-specific areas and is then routed to higher brain areas for decision-making. The goal of intrinsic models then is to understand how much of this temporal information is available at higher sensory areas.

Evidence for Intrinsic Timing

Despite the growing evidence showing dedicated areas acting as internal time-keepers, a second model of temporal processing states that these areas are not in fact where the timing process occurs. A study by Buonomano, Bramen, and Khodadadifar (2009) examined the state-dependent network (SDN) model as an explanation of sub-second sensory timing. The SDN model states that temporal processing is inherently encoded in the evolving state of neural networks. During an interval discrimination task, the first event initiates a change in the network, and because of evolving activity during the empty interval, the network will be in a different state when the second event arrives. Differences in that population state (depending on the interval length) can then code for time. This model requires local intrinsic activity, but can be modified by more top-down cortical activity, as may occur with attention.

Limits to short-term plasticity play a crucial role in this model. A study by Karmarkar and Buonomano in 2007 compared this SDN model to the idea of a pacemaker-accumulator model used in Gibbon's Scalar Expectancy Theory (1997). A major argument is that the SDN needs time to reset, requiring something on the order of a few hundred milliseconds. To test if this occurs *in vivo*, they presented standard intervals and asked participants to judge if a comparison was shorter or longer. The comparison could be either a 2 tone (2T) or a 3 tone (3T) interval; the first tone of the 3T comparison acted as a distracter, and due to its random occurrence, participants could not develop strategies to ignore it. Distracters could either be fixed (producing an interval equal to that of the final two tones) or variable, and participants could have different strategies to

deal with the 3T interval depending on whether timing occurred according to an accumulator-pacemaker model or a SDN model.

For the 3T interval, participants could either store the interval between tone 1 and 2, then reset that count to store the time from tone 2 to 3 (a reset strategy). They could also subtract the time at tone 2 from the time at tone 3, obtaining the interval between tones 2 and 3 (a temporal arithmetic model). Both strategies could be implemented in the pacemaker model: The accumulator could be reset, or the accumulator could provide a linear metric to use for the temporal arithmetic. Both strategies predict that performance in the 2T and 3T comparisons should be similar in both the fixed and variable conditions, as the predictability of the distracter should not affect the encoding of the two intervals (between tones 1 and 2 and between 2 and 3). However, in a SDN model, the limits on short-term plasticity do not allow for the implementation of a reset, and because of the nature of neural dynamics, there is no linear metric of time to use for arithmetic. The SDN model predicts that performance on the 2T and 3T fixed trials should be comparable, as the feedback at the end of the trial can be used to form consistent foundations for the building of internal representations. However, performance on the 3T variable trials would be impaired, as the state of the network would not be reproducible across trials.

Using target intervals of 100 ms, performance was indeed impaired significantly on the 3T variable conditions only. However, distracters had no effect when the target interval was increased to 1 second. This implies a different memory-based mechanism that allows participants to independently keep track of the two sequential second-long intervals, and is consistent with a linear metric of time. This suggests that perhaps

intrinsic activity can account for automatic, millisecond range timing, but suprasecond timing requires memory-dependent dedicated mechanisms. This can be verified if modality-specific brain responses can be found with sources in local areas; suprasecond timing however may require a memory component that is more modality-general.

To test the generalizability of the SDN model, Buonomano, Bramen, and Khodadadifar (2009) presented participants with a two-interval forced-choice procedure, in which participants judged the length of a comparison to a standard interval. The standard interval was kept constant at 100 ms, and the comparison interval varied according to an adaptive procedure ($100 \text{ ms} \pm \Delta t$). The mean inter-stimulus interval (ISI) varied for short and long conditions (a mean of 250 ms and 750 ms respectively) and was chosen from a uniform distribution. Behavioral data showed different difference limens (DLs), but similar point of subjective equality (PSE) values for all conditions. The lack of an effect found for PSE suggests this is not caused by time compression or dilation, but by a change in precision of temporal discrimination. They then replicated this study using comparison intervals of 50, 250, 500, 750, and 1000 ms; they found that only the shortest intervals (50 and 250 ms) produced impairment of discrimination with regards to the 100 ms standard. They suggest that this reflects a SDN that resets between 250 and 500 ms. The study also addressed whether interval learning generalized across frequency channels, but found mixed generalization when training participants on short or long ISIs.

Karmarkar and Buonomano (2003) argue that short ISIs impair interval discrimination, but that timing information is not the only variable in interval encoding. Karmarkar and Buonomano's study showed that interval learning generalized to unlearned frequencies and filled durations, but not to unlearned intervals of a different

length; non-temporal features such as frequency interval type appear to have top-down influences. Because interval training did not generalize to untrained intervals, the authors argue that specific time-spans are processed by dedicated circuits; however, because training generalized to novel frequencies, these circuits process intervals across non-temporal features. Generalization occurred across filled durations as well, and the authors suggest that this implicates cortical areas and cerebellar circuits instead of via early subcortical plasticity mechanisms.

In order to examine the contributions and extent of early synaptic circuits (and mechanisms such as short-term plasticity) and how they contribute to the state of the network, Buonomano (2000) performed a computational analysis of simple disynaptic circuits and examined three time-dependent properties that were predicted to be crucial in shaping postsynaptic responses to temporal stimuli. These included paired-pulse facilitation (PPF) of monosynaptic EPSPs, paired-pulse depression (PPD) of fast IPSPs, and slow IPSPs. Using one excitatory and one inhibitory unit, an excitatory unit showed a preference to either the first or second pulse of a pair. In a number of different disynaptic configurations, the first pulse generated a subthreshold EPSP, but the second pulse lead to a suprathreshold EPSP because of the PPF of the EPSP and the PPD of the IPSP. However, when provided a stronger input and a slow inhibitory response to the excitatory unit, the suprathreshold response changed from the second pulse to the first, and the PPF and PPD became lost in the increased strength of the slow IPSP. This means that inhibitory input can modulate the preference for the first or second pulse; EPSP strength then increases to generate a suprathreshold response to the initial event.

When looking at the same disynaptic circuits, Buonomano then applied different weights to the synapses to examine if they could respond selectively to different intervals. Different weightings caused the excitatory units to fire exclusively to different intervals as measured by the firing of the excitatory unit to only the second pulse; increasing the strength of the input to both units increased the length of the preferred inter-pulse interval. While this was reliable in parallel changes to both units, the data is more complex when input strength is different between the units.

While the notion that order and interval discrimination can occur with just two cells is robust, it's unlikely that there are learning rules that allow for this small of a network to function over a large range of intervals. Buonomano (2000) then constructed a large network representing a system of cortical sensory neurons. The network received two input pulses, with each pulse consisting of a burst of three spikes at 300 Hz. This large network then fed to a recognition network of 5 output units and 400 input units, each representing the number of spikes from an excitatory unit in the sensory network. The network was trained on 12 presentations of 5 different intervals: 50, 100, 150, 200, and 250 ms. It was then tested over another series of 12 simulations, using 12 different test intervals (from 25-300 ms at 25 ms increments). By using a supervised learning rule, the network was able to read out the population code of the sensory network. Importantly, the network was able to generalize to the unlearned intervals by firing maximally to the learned intervals, but sub-optimally to intervals close to the learned interval, and not at all to intervals even further away in duration. This suggests that continuous change in short-term plasticity and slow IPSPs are what underlie the ability of a network to discriminate between intervals.

Intermediate Views

Buetti, van Dongen, and Walsh (2010) claim to have an intermediate view between dedicated and intrinsic models. They investigated the neural correlates of temporal expectancy in the visual cortex and other cortical structures. Participants were required to push a button as quickly as possible after a target changed to a “go” signal; keeping track of the elapsed time the target stimulus remained on screen before changing to the go signal would help them anticipate the proper response time. Buetti and colleagues predicted differences in brain activation during the waiting time between the target and go stimulus. Areas correlated with anticipation included visual area V1, as well as SMA, SMG, vermis, and midfrontal cortex. However, activation in some of these areas (most notably the SMA) could represent motor preparation for the button press. Using retinotopic mapping, they concluded that the areas activated in the visual cortex were limited to areas directly stimulated by the target stimulus; retinotopically specific signals could reflect temporal expectancy in V1/V2 and V3. This lends support to intrinsic models, with possible modulation from other cortical areas, suggesting that the cerebellum, SMA, and parietal cortex may be involved in representing elapsed time.

Another study by Buetti, Walsh, Frith, and Rees (2008) suggests that there might be a partial dissociation between temporal reproduction and estimation, and different circuits could be involved in motor and perceptual timing. During fMRI, participants were asked to either reproduce an interval (a completely motor act) or to press a button when a comparison interval had been presented for a time equal to a standard (which is based on perceptual judgments). The control conditions were very similar, with button

presses or releases being timed to the removal of the stimulus (whereas the experimental conditions required the participant to remove the stimulus). These four conditions provided a series of contrasts that allowed the authors to isolate brain activation specific to reproduction or estimation.

Areas activated specifically for reproduction included the right inferior parietal cortex (IPC), the left inferior frontal gyrus, the left middle frontal gyrus, and the left putamen. Activations were also found in the bilateral SMA, right middle frontal gyrus, basal ganglia, and cerebellum. Areas activated specifically for estimation included the left inferior frontal gyrus, left putamen, bilateral globus pallidum, right caudate nucleus, and right cerebellar hemisphere. Areas activated more strongly for reproduction than estimation included the right IPC, left fusiform gyrus, cerebellar vermis, and area V5 (MT). Activation was also found in the right pre-SMA, left premotor cortex, left middle frontal gyrus, left IPC, and bilateral cerebellum. No areas were activated more strongly for the estimation task than the reproduction task. While cerebellar and basal ganglia activity was found for both tasks, there was also cortical sensory-specific activation in V5, suggesting dedicated modulation of intrinsic activity.

This sensory-specific activation was also found in the auditory modality. Buetti, Bahrami, and Walsh (2008) used TMS during auditory duration and frequency discrimination tasks and determined that the superior temporal cortex (which includes primary and secondary auditory cortex) played a strong role in auditory timing (see N'Diaye, Ragot, Garnero, & Pouthas, 2004 and Shih, Kuo, Yeh, Tzeng, & Hsieh 2009, for evidence this area plays a role in visual timing as well). During TMS, the stimulation was applied to either the vertex (as a control) or the right or left STG at comparison

stimulus onset. When TMS was applied over the right STG, participants were less accurate in temporal discrimination and required greater differences between standard and comparison than when TMS was applied to left STG or the vertex. No effect of TMS on frequency discrimination was observed, nor on PSE at any site or for any task. TMS at right STG lowered overall accuracy and reduced sensitivity in performance of auditory timing tasks as compared to left STG or vertex stimulation. This suggests a right lateralized, modality-specific, distributed form of auditory temporal processing.

The strength of the intrinsic models comes from the ability of sensory-specific areas to process time without the need for a dedicated central clock mechanism. Unfortunately, the intrinsic models listed previously cannot account for timing outside of a short millisecond range. Whether dedicated timers must allocate memory to the intrinsic timers or read out the outputs from intrinsic networks as a type of accumulator pulse still remains unclear. It is also unclear how intrinsic timers can account for the growing evidence that subcortical areas play crucial roles in temporal estimation and reproduction: Cerebellar areas could be modality-general, or these areas may be stops in intrinsic pathways. It may be that activation of timing mechanisms is heavily dependent on the task. More work should be done on fully dissociating the effects of short and long intervals, and of motor and non-motor timing.

Auditory and Visual Differences

If timing mechanisms are task- and therefore modality-dependent, then there may be significant differences when using an auditory versus a visual task. The visual system surpasses the auditory system for spatial information (as evidenced by the ventriloquism

effect; see Thurlow & Jack, 1973). However, the auditory system seems to surpass the visual for temporal processing. There also seem to be different limits when it comes to synchronizing with auditory or visual stimuli. A study by Repp (2003) examined the limit of participants' ability to tap to extremely fast sequences, and found that while the synchronization threshold of auditory stimuli occurs at an IOI of about 120 ms, the threshold for visual stimuli is about 450 ms. Even with the added benefit of subdividing these fast sequences, the visual modality cannot compete with the auditory modality for synchronization, and in fact, subdivision and meter may only help auditory rhythm perception but not visual rhythm perception. A study by Patel, Iversen, Chen and Repp (2005) found that visual synchronization required longer intervals when compared to auditory synchronization, and metrical structure and subdivision did not improve participants' tapping to visual rhythms. This apparent disadvantage of the visual system appears to be robust: A study by Repp and Penel (2004) found again that tapping variability was higher in visual trials than in auditory, and auditory distracters increased variance significantly more than visual distracters (which were largely ineffective). An added intensity condition found that salience of the stimuli did not matter: Auditory distracters were just as effective when they were soft as when they were loud. Spatial congruence did not matter either: Results for congruent and incongruent spatial stimuli were similar, although congruent stimuli may have contributed to stronger visual distracter effects. Finally, after a certain relative phase, auditory distracters became less effective (past a 100 ms integration window), but variability increased due to shifts in phase. This inability of the visual system to compete with the auditory modality suggests that either the auditory system is itself more efficient at temporal processing, or the visual

modality uses a less direct, much slower path to process information prior to arriving at a central timer.

However, a study by Shih and colleagues suggests that differences in modalities might be due to visual temporal information needing to be “translated” through auditory areas (Shih, Kuo, Yeh, Tzeng, and Hsieh, 2009). Participants were presented with duration discrimination tasks using both empty and filled intervals in both auditory and visual modalities. MRI results for the auditory condition revealed activation in bilateral preSMA/SMA and bilateral superior temporal gyri during discrimination. The visual task activated bilateral preSMA/SMA and right prefrontal cortices. There was also significant parietal activation in the bilateral precuneus and the left inferior parietal lobule. During visual duration discrimination, activation was found in the bilateral superior temporal gyri, left cingulate cortex, and bilateral anterior lobe of the cerebellum. A conjunction analysis reveals common activations in the bilateral preSMA/SMA, left basal ganglia, and dorsolateral prefrontal cortices for both tasks. The authors argued that the preSMA/SMA and basal ganglia are dedicated timers, and the temporal lobe activation found in the visual task points to a routing of temporal information from the more spatial visual areas to the more temporally efficient auditory areas. This would suggest modality-general timing, with dedicated timers in subcortical and cortical areas.

Karabanov and colleagues also found temporal lobe activation for auditory and visual sequence reproduction (Karabanov, Blom, Forsman, and Ullén, 2009). Participants were trained on two rhythms (one auditory, one visual), and were asked to reproduce them during fMRI scans. Significant activation was found in the frontal motor and premotor areas, superior temporal and parietal regions, the basal ganglia, and the

cerebellum during rhythm reproduction. There was no evidence that modality-specific training lead to modality-specific long-term representations. There seemed to be modality-general activation in the posterior part of the left superior temporal gyrus and the inferior parietal cortex around the temporo-parietal junction (which may be important to auditory-motor integration and musical improvisation). This area is a key node in the dorsal auditory pathway, which appears to be important in translating auditory sequences into motor activity.

If visual rhythms become translated to the auditory modality, presenting incongruent auditory information during visual processing should impair visual discrimination. However, if visual rhythms are processed visually, then the auditory distracters should have minimal effect. Guttman, Gilroy, and Blake (2005) performed three experiments in which participants judged standard and comparison sequences of paired, matched or mismatched auditory and visual stimuli as the same or different, and measured the extent to which task-irrelevant visual and auditory information interfered with performance. Auditory distracters had a clear effect on visual discrimination: When the sound coincided with the visual changes, performance increased, and when the sound was incongruent, performance decreased (both with regards to a no-sound control). When provided with both auditory and visual task-irrelevant information, both types of irrelevant input decreased performance, with the irrelevant auditory information producing the greatest reduction. The third experiment varied whether the irrelevant auditory information affecting the encoding of the sequences by placing distracters in the standard, the control, both sequences, or neither sequence. When the distracters were placed during the standard only, performance decreased significantly, and when

distracters were placed in the comparison only, visual discrimination decreased slightly (but not significantly). This suggests that visual information is automatically and involuntarily encoded in an auditory manner, and that auditory distracters can interfere with this encoding.

Beat Versus Interval Strategies and the Current Paradigm

It has been argued that both beat-based and interval-based models of timing represent dedicated timers (Ivry & Schlerf, 2008). Beat-based models of internal timekeeping posit an internal oscillator that maintains a beat. This oscillation is usually depicted as a sine wave, with peaks in amplitude corresponding to onsets and the space between onsets as the period. This oscillator entrains to (synchronizes with) a series of external beats, and can then hold that beat internally without continued external cues. The earlier, more traditional view of beat-based models (Keele, Nicoletti, Ivry, & Pokorny, 1989; Pashler, 2001; Schulze, 1978) does not allow the oscillator to change the period or phase of its oscillations, which may not be the best approach to understanding beat perception in complex or changing rhythms. Despite these earlier, more easily refuted forms of beat-based models, recent studies that allowed for period and phase correction have strengthened the arguments for this model (McAuley & Jones, 2003).

In interval-based models, the listener encodes the time it takes for two events and compares that memory to all future events. Interval-based models need three parts: an internal clock that estimates the time between two events, a store for that information, and a mechanism that compares the stored interval with intervals formed by the current external onset. If beat-based models can be thought of as using an oscillator-like

mechanism, interval-based models can be thought of using a stopwatch- or hourglass-like mechanism. The amount of sand in an hourglass represents the time between two onsets, and at each subsequent onset the hourglass is turned and empties its sand in time for the next event. There are two basic types of interval models: The first is an adjacency model, where the listener compares adjacent intervals to form a standard interval. The second is a memory model, where the perception of the first two intervals is averaged and then stored, with all subsequent intervals compared to that memory trace (Keele, 1989). Both models have strengths and weaknesses, and lead to different arguments against the beat-based models.

To properly quantify and examine the differences between interval- and beat-based models, McAuley and Jones developed and tested computational models of each (2003). They developed four cardinal models: a interval-based full-reset model, where the internal clock is reset every event, and each estimate is independent of successive time intervals (context free); an interval-based phase-reset model, that adjusts phase but not period and is more dependent on context; a beat-based period-reset model, which alters period but does not adjust phase; and a beat-based no-reset model, with neither phase nor period correction, and which corresponds to the earlier proposed beat based model (as presented by Keele, Pashler, and Schulze). After four experiments, they concluded that while the beat-based phase-reset is best of all four models, it over-predicts assimilation, or the ability to match standard and context IOIs despite a possible asynchronous gap between the two sequences. However, a modified phase-reset model with an interval-averaging algorithm (an interval model that does not fully reset when off the beat and averages previous intervals) and an entrainment variation with partial phase

and period correction (a beat-based model that does not completely reset and corrects phase and period when off the beat) both predict better than the strict four cardinal models. Constraining the parameters to strictly beat-based or interval-based did not account for behavioral responses, suggesting a more graded use of both strategies.

To examine this issue in more depth, McAuley and Jones's computational model also allows for the calculation of individual differences in how beat-based or interval-based a listener is. This uses the parameter, W , which can range from 0 (completely interval-based) to 1 (completely beat-based). McAuley, Frater, Janke, and Miller (2006) developed a stimulus pattern (discussed in further detail in procedures) that was predicted to be perceived differently depending on the listening strategy. A control sequence was presented consisting of four tones: two initial tones separated by 600 ms, a pause of 1200 ms, followed by two final tones separated by a variable interval that was either shorter or longer than an implied 600 ms. All listeners heard the pattern speeding up or slowing down at the end based upon whether the sequence was shorter or longer than expected, respectively. This was contrasted with an experimental sequence of five tones. The five-tone sequence inserted a tone that bisected the initial two tones; otherwise, the sequence was identical to the control sequence.

McAuley and colleagues found that one group of listeners heard the pattern speeding up or slowing down when the final interval was shorter or longer than 600 ms, similar to the perception of the control sequence. However, another group of listeners perceived the pattern as always slowing down. The authors attributed this to differences in beat-based versus interval-based strategies, and follow-up studies found differences in brain responses dependent on listening strategy (Grahn, Henry, & McAuley, 2011;

Snyder, Pasinski, & McAuley, 2011). McAuley and Henry (2010) examined whether these differences exist within the visual modality as well using fMRI, and used a modified version of the original paradigm using visual flashes of the same duration and temporal position. Therefore, this modified auditory and visual paradigm should allow the examination of listening strategy in detail; it also provides a well-documented design that is known to elicit strong brain potentials. By adapting a paradigm that consistently elicits both beat-based and interval-based strategies to both auditory and visual stimuli, we can establish how deeply ingrained these strategies are both within participants and across modalities. The use of this paradigm should illuminate neural processes underlying rhythm perception, regardless of whether or not these individual differences in timing strategy actually cross modalities.

Sensory-Evoked Potentials

Using EEG can help illuminate the brain responses required for rhythm perception. Differences in cognitive ERP components elicited by auditory and visual stimuli might suggest modality-specific contributions, whereas similar components across modalities might suggest dedicated timers. Differences in sensory-evoked potentials between the 4- and 5-event patterns may suggest differences in interval encoding strategies. Both single tones and trains of tones elicit reliable ERP components. Picton, Hillyard, Krausz, and Galambos (1974) found 15 different sensory ERPs to simple click tones, and classified each into early, middle, and late latency components. Early components (in the first 8-10 ms after stimulus) arise from the inferior colliculi, cochlear nucleus, and acoustic nerve. The fields from these early responses are measured

maximally at the vertex in the form of a far-field potential (Buchwald & Huang, 1975). Middle latency components (between 8-50 ms) are more widespread and are maximally recorded over fronto-central regions (Yvert, Crouzeix, Bertrand, Seither-Preisler, & Pantev, 2001).

N1 and P2 late components are maximally distributed over fronto-central regions and at electrode Cz; due to the large amplitude of these late components, they most likely arise from cortical areas. These N1 and P2 responses are especially of interest; Näätänen, Sams, Alho, Paavilainen, Reinikainen, and Sokolov (1988) found highly specifically tuned independent neuronal populations code for frequency and location. Participants passively listened to 77 different stimulus combinations of location and frequency (11 possible frequencies by 7 different locations) interspersed between a standard of 1,000 Hz centrally located (but the authors found no intensity difference between ears). There was a main effect of both frequency and location (with identical stimuli resulting in lower amplitudes), but no interaction between the two, suggesting no mutual dependence of frequency and location effects. The authors suggested that the N1 could consist of a modality non-specific component that is sensitive to inter-stimulus-interval (which contributes to it diminishing) and a second component reflecting sensory-specific processing (being more stable).

Early and middle latency components are usually stable despite fluctuations in subjective arousal and environmental conditions; however, N1 and P2 late components are enhanced during directed attention towards an auditory stimulus (Picton & Hillyard, 1974). The N1-P2 complex is also sensitive to changes in intensity and spectral frequency (Martin & Boothroyd, 2000). While the N1 component can be reliably elicited, it can also

be made to vary under certain circumstances (Näätänen & Picton, 1987). For instance, in a repeated sequence of tones, amplitudes for both N1 and P2 diminish after each tone; however, if the pitch or temporal properties of the stimulus is altered, the amplitude of the N1 may partially recover (Picton, Campbell, Baribeau-Braun, & Proulx, 1978; Butler, 1968). When participants were presented with deviants in isochronous sequences that occurred earlier than an expected tone, the N1 response to that deviant nearly doubled in amplitude, and the latency elongated (Ford & Hillyard, 1981). This effect is modulated by attention, with higher levels of attention increasing the amplitude of the N1 to early tones by 48%. However, the P2 following these deviants does not change significantly, nor does the N1 following deviants that occur later than expected, suggesting that there should only be differences in N1 responses in sequences that “speed up” in the McAuley paradigm.

Visual sensory responses follow similar patterns to their auditory counterparts, but are too distinct to compare directly (as they display slightly longer time-courses with respect to stimulus onset). Jeffreys and Axford (1972) found early visual evoked potentials (VEPs) to patterned stimulus fields. They found a primary component (CI) that started 40-50 ms after stimulus onset and peaked at approximately 60-70 ms. This was followed by a second component (CII) of opposite polarity peaking at 90-100 ms. CI appeared to be independent of changes in luminance, the type of pattern used, and the effects of sequence and timing. They found that CI and CII have different scalp distributions, and these distributions vary with the stimulated retinal area. CII also undergoes adaptation much like the later latency auditory components, while CI does not.

The authors concluded that CI had a striate cortical origin (BA 17), and the origin of CII was extrastriate.

Maier, Dagnelie, Spekreijse, and van Dijk (1987) conducted a principle components analysis on VEPs recorded to a variety of visual stimuli. They found a primary (positive) principal component to a number of different stimuli: high frequency luminance flickers, pattern onset and offset, pattern reversal, and motion onset. This component appeared to arise from a single current dipole in the primary visual area, and corresponds to a CI. A second (negative) principal component was maximal to pattern onset, but was also elicited by pattern offset and reversal stimuli. This component appeared to originate in higher visual areas. Maier and colleagues argue that the CII component arises from the tail end of their primary principal component and the entirety of their second negative component.

Di Russo, Martínez, Sereno, Pitzalis, and Hillyard, (2001) further examined these early components using EEG and fMRI. Using patterned stimuli, they recorded a CI with a peak latency of 60 and 100 ms, which appeared to reverse in polarity for upper versus lower visual-field stimulation (due to the retinotopic organization of the striate cortex). Source localization placed the origin of this component in the primary visual cortex. A CII (which the authors refer to as a P1) was recorded with a peak latency of between 100-130ms and a source localization to V3 and V3a. Di Russo and colleagues also measured a complex, later-latency negative response (N1). This N1 complex is a series of components from 150-200 ms, and appeared to arise from multiple generators. It can be separated into four temporarily overlapping subcomponents: an occipital N150, an anterior centro-parietal N155, and a difficult to localize N180 and N200.

These P1 and N1 waves can be modulated by attention. Luck, Heinze, Mangun, and Hillyard (1990) presented participants with unilateral and bilateral letter fields, and participants were told to attend to only one side. Participants were instructed to push a button when a target letter appeared on the attended side, and to ignore it when it appeared on the unattended side. They found that unilateral presentation of non-target stimuli elicited larger P1 and N1 components when attended than when unattended. Bilateral presentation of non-target stimuli produced larger positivities at scalp sites contralateral to the attended visual field; this component extended over the range of P1, N1, and P2 peaks. Attended bilateral target stimuli elicited similar P1 and N1 components to the non-target arrays, but elicited a long-latency N2 wave (200-250 ms) specific to attended-field targets. Luck and colleagues argue that the N1 might reflect orientation to task-relevant stimuli or the engaging of an attentional mechanism. The P1 enhancement might reflect facilitation of sensory input due to attention. The long-latency N2 might be modality specific, and the authors found larger N2 amplitudes in the hemisphere contralateral to the attended-target's visual field; this suggests that target classification may be carried out preferentially in the hemisphere receiving direct sensory information. Examining these basic sensory responses are important to understanding modality-specific brain responses required for rhythm perception. If higher-level responses are modality-specific, they may resemble these early sensory responses.

The Contingent Negative Variation as an Index of Timing

The Contingent Negative Variation (CNV) is a slow, late, negative component that has been shown to reflect expectancy of tones. Work by Walter and colleagues

(1964) showed that attention and expectancy sustain CNV latency. Therefore expectancy for the final tone in a sequence should produce a slow buildup of the response leading up to the last tone. The CNV also reflects the association of a warning (conditional) and indicative (unconditional) stimuli in both the visual and auditory modalities (see both Pfeuty, Ragot, & Pouthas, 2003, and Walter, Cooper, Aldridge, McCallum, & Winter, 1964). The warning stimulus conveyed information about a rule (for example, a tone in the left ear means press a button with the right hand) and participants stored that rule until the presentation of an indicative stimulus; the participants then produced the required behavior. Studies have shown that the CNV occurs between a warning & indicative stimulus for single intervals or series of intervals (Pfeuty et al., 2003) but not self-paced tapping or decision-making (Ikeda, Lüders, Collura, Burgess, Morris, Hamano, & Shibasaki, 1996).

The CNV is closely related to the readiness or Bereitschaftspotential (BP), which has been argued to be a component of the CNV proper. The BP is only recorded with self-paced voluntary movement, and arises from generators in the primary sensorimotor and supplementary motor areas (SMA), but not the premotor or prefrontal areas that are usually associated with decision making (Ikeda, et al. 1996). The CNV appears to have two phases: an early phase related to orientation towards the warning stimulus, and a late phase related to motor response preparation. This late phase is most similar to the BP, but includes additional activation of posterior sensory-related sites (Gómez, Fernández, Maestú, Amo, González-Rosa, Vaquero, & Ortiz, 2004). Similar work was done by Cui, Egkher, Huter, Lang, Lindinger, and Deecke, in 2000: They found an early, task-dependent CNV between a visual warning stimulus and an auditory indicative stimulus.

The late CNV seemed to be motor-preparatory, and was absent when participants were given a non-motor control task. Because our paradigm requires a behavioral motor response, we can expect to find a distinct early and late CNV.

A study by Pfeuty and colleagues (2003) suggests that different strategies could result in group differences of CNV responses for encoding intervals. They hypothesized that if listeners use an interval-based strategy, systematic increases in the CNV will reflect encoding of timing intervals. They also posited that after a certain temporal window, the interval will be maximally encoded, and the CNV will no longer increase. They tested sequences of three and six intervals (standard sequence IOIs of 600 ms with target sequences of $\pm 4\%$ of 600 ms) and asked participants if the test sequences speeded up, slowed down, or stayed constant as compared to the preceding standard.

Based on the calculated CNV amplitude slope at electrode FCZ, participants were separated into two groups: G1 showed an overall positive slope and G2 showed an overall negative slope (for z units per time interval). G1 showed an increased CNV throughout the encoding phase, and G2 showed a high CNV for the first interval, but a decreased response to subsequent intervals. G2 also displayed higher accuracy behaviorally. Increases in CNV only occurred during the standard and not the test phases, and a steady decrease in CNV was even observed in the 6-interval sequence, suggesting a similar amplitude decrease or habituation found in the N1-P2 complex to repeated tones (Ford & Hillyard, 1981).

According to an interval-based model, the increase in the CNV for successive intervals might aid in developing a memory trace. When that trace is established, the load on attention decreases, and the CNV plateaus and then decreases; this fits the

assumption that G1 was an interval-based group. If G2 was employing a beat-based strategy, that group may only need one interval to start the entrainment process, and successive intervals do not help strengthen the entrainment. A second hypothesis suggests that an increased CNV reflects storage of successive tones. Pfeuty and colleagues argue a lack of CNV for the test tones suggest that listeners encoded the standard interval then used a beat-based strategy to compare that to the external test tones.

However, these results can be misleading. Both groups may have used an interval-based strategy, and G2 may just have been more efficient in forming memory traces. Beat-based strategies may also require more than one interval to fully entrain, suggesting that the decrease in CNV found in G2 might not be related to entrainment or interval encoding. Thus, CNV differences can be difficult to explain and might not be the best indicator of strategies between groups. A previous study using McAuley's paradigm (Snyder, Pasinski, & McAuley, 2011) examined group differences in the CNV. We found a larger CNV for beat-based listeners when presenting the 5-tone condition, but no group differences when presenting the 4-tone condition. This group by sequence interaction may reflect active engagement of strategy.

A study by N'Diaye, Ragot, Garnero, and Pouthas, (2004) examined the modality specificity of the CNV using simultaneous EEG and MEG. Participants were given 6 presentations of a standard stimulus before given a test stimulus of 490, 595, 700, 805, or 910 ms. Participants were asked if the test stimulus was the same or different from the standard. Auditory stimuli were sustained 500 Hz pure tones, and visual stimuli were sustained green lights presented for identical durations.

N'Diaye and colleagues found a P1-N1-P2 complex in both modalities, with an

N1m and P2m measured through MEG. There was also CNV for both modalities; however, this was larger for the auditory modality. CNV topography as measured through EEG appeared to be modality-independent; however, MEG topography showed activation similar to early modality-specific sensory responses. The CNV measured also appeared to have an early and late component; the early component appeared to reflect temporal encoding of the stimulus, while the late may reflect a decision-making component. The authors also argue that the CNV is not a component of interval timing, as the resolution of the CNV is dependent on the standard interval, and not the comparison.

Macar and Vidal (2004) examined the CNV as a possible index of timing using the framework of the pulse accumulator theory. If the CNV is an on-line index of timing, it might represent a pulse accumulator. They argued that an increased CNV amplitude might reflect longer estimates of the interval, as more pulses are needed for longer intervals. Participants were asked to judge successive intervals, and EEG was measured leading to the final tone of the comparison interval. They found that CNV amplitude is increased as a function of interval duration, and source localization suggested that a temporal accumulator might exist in the SMA. However, the late phase of the CNV appeared to be affected by components of decision-making; memory and context updating potentials like the P3 may have muddied the late stage of their CNV.

Novelty and Context-Updating P3

Another component closely tied to expectation is the P300 - also known by P3 or Late Positive Component (LPC), which can be further divided into the P3a and P3b

(Polich, 2007). Its discovery stemmed from the use of the oddball paradigm; early uses of these paradigms suggest the P3 is an important component of stimulus probability and task relevance. P3 topography is usually measured over the midline at Fz, Cz, and Pz. The context updating theory of the P3 states that the component indexes brain activities that revise mental representations of incoming stimuli. If no change to the representation is detected, only sensory evoked potentials are recorded, but if a change forces an update of the representation, a P3 emerges. This can be seen for simple sound, word, or object representations.

This hypothesis is very robust, and has resisted refutation for over 25 years. Attention plays a large role in this process, and the overall system is modulated by overall arousal level, which controls the amount of attention available for performing the task. The P3 has ties to longer-term memory as well, as encoded words elicit larger components than do words not recalled, and can be diminished by rehearsal strategies and enlarged by rote memorization. This suggests that stimulus encoding that promotes successful integration into storage and facilitates retrieval and recognition enlarges P3-like amplitude. If this is true, and if auditory strategies are more efficient at encoding and storing intervals, than there should be modality differences with the auditory modality showing larger P3 components, as it allows more allocation of memory resources.

A comprehensive review by Linden examined the novelty versus context updating P3 and possible source locations (2005). The novelty P3a appears to be a neural correlate of the orienting response; it also appears to reflect inhibition of a prepotent response to a salient but task-irrelevant stimulus. The P3b appears to reflect context updating or cognitive processing; the P3b amplitude decreases with increased memory load,

suggesting a shared reliance on attention. The context-P300 appears to be affected by temporoparietal junction lesions, while the novelty P300 is affected by medial temporal, frontal, and parietal lesions; however, the P3b system seems to be more redundant than the P3a. The P3b appears to be a much better candidate for indexing temporal processing than the earlier novelty P3a.

Other factors such as expectedness, probability, and salience of targets can influence the amplitudes, latencies, and scalp topographies of the P3 subcomponents (Jongsma, Meeuwissen, Vos, & Maes, 2007). When isochronous sequences were accelerated or decelerated at the end, changes in both early and late P3 responses were found based on whether participants perceived speeding up or slowing down of the sequence. When compared to purely isochronous sequences, trials that sped up at the end resulted in a slightly smaller early P3 component but a markedly increased late P3 response. For trials that slowed down at the end, there was a marked increase in early P3, with no changes in late P3 response. This early versus late P3 difference might be explained by a beat-based model: an oscillator can predict when attention should be optimally allocated in the future. A “too early” beat leads to a surprise effect that leads to an increase in late P3, and a “too late” beat arrives when attention is already maximal, leading to a larger early P3.

Gontier, Dantec, Paul, and Bernard (2008) argue that the P300 reflects a late positive, time-related component (LPCt) that reflects decision making in duration tasks, and may be generated during the inhibition of non-relevant information. They used a visual duration and size discrimination task to examine whether this component reflects timing specifically or more general decision-making processes. The LPCt increased in

amplitude as a function of stimulus durations, with higher LPC amplitude and greater behavioral accuracy for longer intervals. The component was also right-lateralized, consistent with previous findings (mentioned above) emphasizing a right hemisphere importance to temporal processing. They found that the LPCt was also present during a size discrimination task; they argued that the LPC may reflect general decision-making. However, the size task still included a temporal factor – the size stimuli were presented at the same variable durations as in the duration discrimination task –so it is difficult to fully dissociate the LPCt from temporal processing.

According to the context updating theory of the P3, our paradigm should elicit distinct P3 responses after the final tone of the sequence, as this tone forces an update of the established beat. As this response is closely tied to expectancy and attention, there also should be differences in P3 amplitude based on whether the tone occurs too early or too late. Where the CNV might index timing directly, the P3 can be elicited by simple sounds, words and objects, suggesting that it instead reflects the more memory-dependent, cognitively-controlled aspect of timing.

Possible Sources And Generators For Event-Related Potentials

The CNV is measured maximally at fronto-central electrodes, and should be measured to both visual and auditory patterns (Pfeuty, et al., 2003, and Walter, et al., 1964). However, sources for the auditory and visual could require different brain areas. If both the auditory and visual CNV reflect the same process (using a dedicated timer), then both should show similar sources. If both CNVs show different sources, this would suggest an intrinsic model of timing. A study by Diaconescu, Kovacevic, and McIntosh

(2008) used independent component analysis to examine the CNV, and found a warning stimulus-related response with a posterior positivity and parietal-frontal negativity that may reflect an early visual CNV. Another component was facilitated by the presentation of the warning stimulus prior to the indicative, and was modulated by modality. These components seemed to reflect modality independent brain activity; however, there are some concerns about what constitutes a visual vs. auditory CNV.

For all the studies mentioned above that use auditory and visual stimuli, the warning and indicative stimuli are presented in different modalities. For instance, the Walter et al., study used an auditory click to cue a visual flash, and a visual flash to cue an auditory click. There seems to be a lack of studies that present both the warning and indicative stimuli in the same modality and then compare across modalities. A benefit of the proposed study is that CNVs elicited from the auditory patterns should reflect pure auditory CNVs that can be directly compared to the pure visual CNVs elicited from the visual patterns.

Le Dantec and colleagues examined possible ERPs and source locations associated with visual duration discriminations (Le Dantec, Grontier, Paul, Charvin, Bernard, Lalonde, and Rebaï, 2007). They argue that the prefrontal cortex and the parietal cortex share functional connections along with pathways to the basal ganglia; these circuits may influence the relationship between the P3a and P3b. They presented four visual stimulus pairs (consisting of short or long durations, and separated by an ISI of 1 sec), and half of participants were asked if the first duration was longer or shorter than the second, while the other half were asked if the second duration was longer or shorter

than the first. They observed two CNVs during the stimulus presentation, as well as two P300s following each stimulus.

The CNV was larger during the first interval when instruction 1 was used, and CNV amplitudes were higher when behavior was more accurate; the same pattern was seen for the P300 to the first interval. P300 as measured by electrode P3 varied as a function of duration and order, in line with a left PAC contribution. During the second interval, both CNV and P300 were sensitive to duration and order, but not to instruction; suggesting less short-term memory load. The authors suggest that the PFC and other subcortical structures form an accumulator function emitted by a pacemaker. The PAC stores the memory, the PFC compares the stored memory, and decides whether the comparison is shorter or longer (in line with Basso et al, 2003). This may suggest distributed cortical timing rather than a localized pacemaker/accumulator in the cortex or basal ganglia.

Motivations for the Current Study

Despite the large role perception of timing information plays in our every day activities, we still do not understand what mechanisms underlie these perceptions. Using EEG can help illuminate both sensory and sustained responses that might contribute to rhythm perception. The CNV appears to be an on-line index of timing, and may index modality-specific timing. Modality-specificity may support intrinsic models of timing; sensory-specific areas could process time independently without the help of a central clock mechanism.

However, the context-updating P3b may reflect a more attention-dependent

temporal decision-making process. This P3b appears to be a much better candidate for indexing temporal processing than the earlier novelty P3a; this component may also reflect modality-general timing processes. As intrinsic models become less viable for longer intervals, a more cognitive component may be needed to act as a central timer. Whether dedicated timers must allocate memory to the intrinsic timers or read out the outputs from intrinsic networks as a type of accumulator pulse still remains unclear; however, modality-general activation may strengthen the claims of dedicated models.

CHAPTER 3

MATERIALS AND METHODS

Participants

Thirty-one participants (14 male) with normal hearing (≤ 30 dB from 250-4000 Hz) were recruited from the University of Nevada, Las Vegas Psychology subject pool. Participants were aged 18-47 (mean = 22.9) years old with no prior history of substance abuse, and received course credit.

Materials and Procedure

The auditory stimulus used in the rhythmic sequences was generated off-line in Matlab and consisted of a single pure tone (440 Hz, 50 ms in duration, including 5 ms rise/fall times). Two types of auditory sequences were created using this one tone (Figure 1). The five-tone sequence consisted of 3 initial tones marked by two 300-ms inter-onset intervals (IOIs) followed by 2 tones that marked a variable final IOI ($600 \text{ ms} \pm \Delta T$). This resulted in a five-tone sequence with a periodic 600-ms beat that was implied (but not explicitly emphasized) by the temporal structure of the first three tones of the sequence (Povel & Essens, 1985). A four-tone sequence consisted of 2 initial tones that specified a 600-ms IOI followed by 2 tones marking the same variable final IOI ($600 \text{ ms} \pm \Delta T$) as the five-tone sequence. Thus, the only physical difference between the two sequences was that the 2nd tone from the five-tone sequence was missing in the four-tone sequence. For both sequences, the initial group of tones was separated from the final group of tones by an IOI of 1200 ms. Final IOIs of the sequences were $600 \text{ ms} \pm \Delta T$, where ΔT equals $\pm 4\%$, $\pm 12\%$ or $\pm 20\%$ of the implied 600ms IOI (480, 528, 576, 624, 672, or 720 ms).

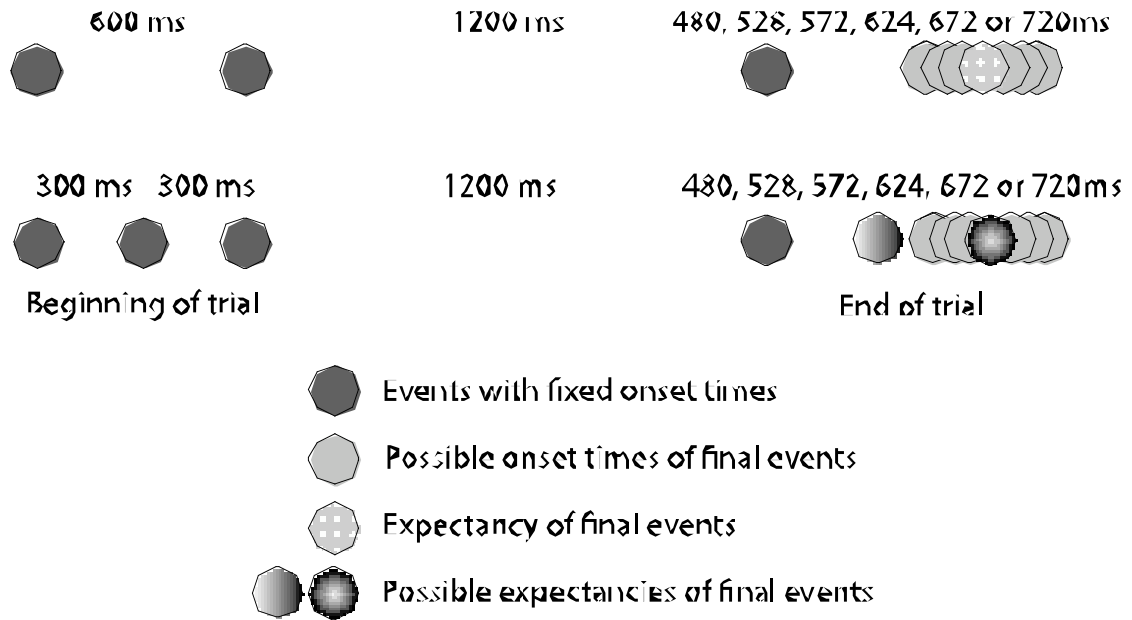


Figure 1: Five-tone and four-tone patterns. Final tone intervals reflect $\pm 4\%$, $\pm 12\%$ or $\pm 20\%$ of the implied 600ms IOI. Visual patterns follow this same presentation.

The visual stimulus used in the rhythmic sequences consisted of a black 60 x 47 pixel box flashed on a white screen for 50 ms. Visual sequence structure and IOIs were identical to those of both the auditory sequences, except there was no rise/fall time.

Participants were assigned to either an auditory first or visual first condition. For each condition, six blocks were presented. In the auditory first condition, participants heard two blocks of auditory 5-event sequences followed by two visual 5-event blocks, an auditory 4-event block, and a visual 4-event block. Participants in the visual first condition saw two blocks of visual 5-event sequences followed by two auditory 5-event blocks, a visual 4-event block, and an auditory 4-event block. Each block contained 132 trials, with 22 trials each of the six trial types ($\pm 4\%$, $\pm 12\%$ or $\pm 20\%$) for that condition. Eight practice trials (using final IOIs of $\pm 40\%$) were presented prior to the start of the

experiment, and verbal instructions were given before each 4-event block to inform the participant as to the changing condition type.

Participants were seated in a comfortable chair in a single-walled sound-attenuated room and were asked to maintain fixation on a black cross on a white background in the center of a computer screen for the auditory conditions only. No fixation cross was provided for the visual conditions. Participants were seated approximately 80 cm from the screen. Participants were asked to listen to or watch the stimuli during electrophysiological recording, and to avoid moving their eyes, head, or other body parts while the stimuli are presented. At the end of each sequence, participants indicated by pressing one of two buttons whether they perceived the pattern “slowing down” or “speeding up” at the end. Participants were allowed a 2 s inter-trial interval to make their responses before a new trial began.

The auditory stimulus patterns were presented binaurally through ER3A headphones at 70 dB SPL. The auditory and visual sequences were presented and behavioral responses were collected by a custom program written in Presentation. Behavioral responses were collected using an RB-830 button box.

Electrophysiological Recording

Electroencephalographic (EEG) signals were digitized continuously (512 Hz sampling rate and a 104 Hz bandwidth) using a Biosemi ActiveTwo system. According to the Nyquist theorem, this is an adequate sampling rate because it is at least twice as great as the highest frequency in the signal. The EEG was recorded from an array of 72 electrodes, with a Ag-AgCl Common Mode Sense (CMS) active electrode and a Ag-

AgCl Driven Right Leg (DRL) passive electrode serving as ground, placed at 64 points based on the 10/20 system in a Biosemi electrode cap and 8 additional points below the hair line (both mastoids, both pre-auricular points, outer canthus of each eye, and inferior orbit of each eye) and recorded onto a PC desktop computer for offline analysis. Before EEG recording, conducting gel was applied to the skin at each electrode site with the cap on and sintered Ag-AgCl pin-type electrodes were fitted into place at each site in the cap. Sintered Ag-AgCl flat-type electrodes were attached with adhesive to sites below the hairline to correct for eye and muscle movements. No abrading of the skin was necessary. Voltage offsets were adjusted to below 40 mV prior to recording and the resting EEG was checked for any problematic electrodes prior to and throughout the recording session. If large amounts of drift were present while recording, a small fan was placed in the sound booth during the breaks between blocks.

Data Analysis

Proportions of ‘speeding up’ responses were calculated for each participant for each of the 24 trial types (4 conditions x 6 final IOIs). This allowed us to check that participants were responding correctly and were paying adequate attention. In order to determine whether participants used an explicit 300 ms or implied 600 ms referent interval for the 5-event sequences, the proportions of “speeding up” responses were fit with a simple contrast model (see McAuley & Jones, 2003). For each of the 6 possible final intervals (T_i) we calculated a temporal contrast metric (C_i), which measures the normalized difference between the final interval and the referent interval, P (either 300 or 600 ms):

$$C_i = (T_i - P)/P$$

This results in two possible values of C_i for each of the final intervals (one for each referent) for each modality. Values of temporal contrast for each referent were assumed to be normally distributed with a standard deviation, σ ; and the values of C_{i300} and C_{i600} were then z-transformed and combined using a simple weighted average:

$$z = (1-w)z_{i300} + wz_{i600}$$

Where w is the beat-sensitivity index, and z_{i300} and z_{i600} are the z-transformed values of each temporal contrast referent. The predicted proportions of “speeding up” responses, $P(\text{“Speeding up”})$, for each final interval, T_i , was then generated using a cumulative normal distribution function:

$$P(\text{“Speeding up”}) = 1 - \Phi(z)$$

The model fits allowed both $w \in [0,1]$ and σ to vary, minimizing the root-mean-square error (RMSE) between the observed and predicted response proportions.

This resulted in auditory and visual continuous values of w , ranging from 0 (completely interval-based strategy) to 1 (completely beat-based strategy). Participants were also grouped by median split values of w (the point of split falling between .95-.96 for both auditory and visual conditions). As only the w values from the 5-event sequence indicate differences in strategy, the w values for the 4-event sequences were not used in any of the analyses.

All off-line ERP analyses including baseline correction and amplitude measurements were performed using Brain Electrical Source Analysis software. Electrodes that were noted during the recording as being noisy throughout the experiment were automatically interpolated prior to analysis. Ocular artifacts (blinks, saccades, and

smooth movements) were corrected automatically with a Principal Component Analysis (PCA) method. Epochs contaminated by artifacts (amplitude exceeds 150 uV, gradient exceeds 75 uV, or low signal below 0.10 uV) were automatically rejected before averaging. EEG epochs were averaged separately across all non-artifact trials for each of the 24 trial types and for each electrode site, and re-referenced to the average of all electrodes.

To examine ERPs related to processing the final two events of the sequence, epochs were segmented with time 0 at the onset of the last tone of the sequence, with a 1226 ms pre-trigger baseline period and a 1000 ms post-trigger active period, and baseline corrected by subtracting the mean of the -26 to 0 ms portion of the baseline from each point in the epoch. These epochs were digitally band-pass filtered to attenuate frequencies below 0.5 Hz (6 dB/octave attenuation, forward) and above 30 Hz (24 dB/octave attenuation, symmetrical). Low-pass filtering was necessary to ensure that the sampling rate was at least twice the highest frequency in the signal, and high-pass filtering removed slow skin potentials and other non-neural potentials. To quantify the CNV (which occurred leading up to the final tone), those epochs were digitally band-pass filtered to attenuate frequencies below 0.1 Hz (6 dB/octave attenuation, forward) and above 30 Hz (24 dB/octave attenuation, symmetrical) and baseline corrected by subtracting the mean of the -752 to -726 ms portion of the baseline (time before penultimate event onset) from each point in the epoch.

ERP mean amplitudes were calculated in time ranges showing maximal differences in the grand averaged waveforms between conditions of interest at electrodes showing the maximal difference (from -200 to 0 ms for the CNV and from 325 to 630

ms; based on known latencies for the CNV and P3). Mean amplitudes were averaged across a small number of electrode sites where ERP differences were most prominent for each participant and submitted to repeated-measures ANOVAs with Greenhouse-Geisser corrections when appropriate.

Scalp topographies were obtained in BESA at single time points showing the maximal difference between conditions (-200 ms for the CNV and 400 ms for the P3). After vector normalizing the mean amplitude data for both CNV and P3 (based on the method described by McCarthy and Wood, 1985), a repeated-measures ANOVA was run to identify condition differences in activity patterns across the scalp by testing for electrode x condition interactions.

CHAPTER 4

RESULTS

Behavioral Data

Two repeated measures ANOVA were conducted to examine the possible effects of w values and participants' perception the sequences speeding up. Factors included the size of the deviant (± 4 , 12 or 20% of 600 ms), modality (auditory or visual), and sequence type (5-event or 4-event), and both analyses included order (visual sequences first or auditory sequences first) as a between subjects variable. Values of w were calculated separately for each type of w value (auditory 5- and 4-event or visual 5- and 4-event).

When examining participants' probability of responding that the sequence was speeding up, there was a main effect of modality $F(1, 30) = 8.548, p=.007$, with a greater chance of responding "speeding up" for the visual modality. There was a significant effect of deviant size $F(1, 30) = 173.117, p<.001$, with participants correctly identifying that the sequence sped up for deviants occurring before the implied 600 ms beat (see Figures 2 and 3). This pattern of responding is consistent with previous research using this paradigm (Grahn, Henry, & McAuley, 2011, and Snyder, Pasinski, & McAuley, 2011). There was a significant interaction between modality and sequence $F(1, 30) = 13.351, p<.001$, with little change between proportions in the auditory modality, but a large decrease in the probability of responding "speeding up" from the four- to the five-event visual sequences. This suggests that a larger number of subjects changed between a beat-based to an interval-based strategy for visual sequences only, which may be supported by a three-way interaction trend of modality by sequence by deviant size

($F(1,30)=2.179, p=.060$). There was also an interaction with modality and deviant size $F(1, 30) = 27.954, p<.001$, and sequence and deviant size $F(1, 30) = 4.787, p<.001$. The auditory sequences and the 5-event sequences showed higher proportions of “speeding up” responses for only the earliest two deviants (at 480 and 528 ms); the curves for both types of sequences were steeper than their counterparts. The response curves for the 5-event auditory and 4 event auditory sequences were very similar, and matched what was previously found for 5-tone auditory patterns. However, the response curves for the visual sequences showed more variability between five- and four-event sequences, and both were shallower than their auditory counterparts. A paired samples t-test was conducted to compare just noticeable differences (JNDs) in the auditory and visual 4-event conditions. There was a significant difference in the JNDs for the visual condition ($M=30.09, SD=37.02$) and the auditory condition ($M=11.62, SD=6.62$); $t(30)=-2.887, p = 0.007$. This suggests that performance was poorer in the visual condition. There was no significant main effect of whether the participants were presented with auditory or visual sequences first, $F(1, 30) = 1.306, p>.05$, nor were there any interactions involving order.

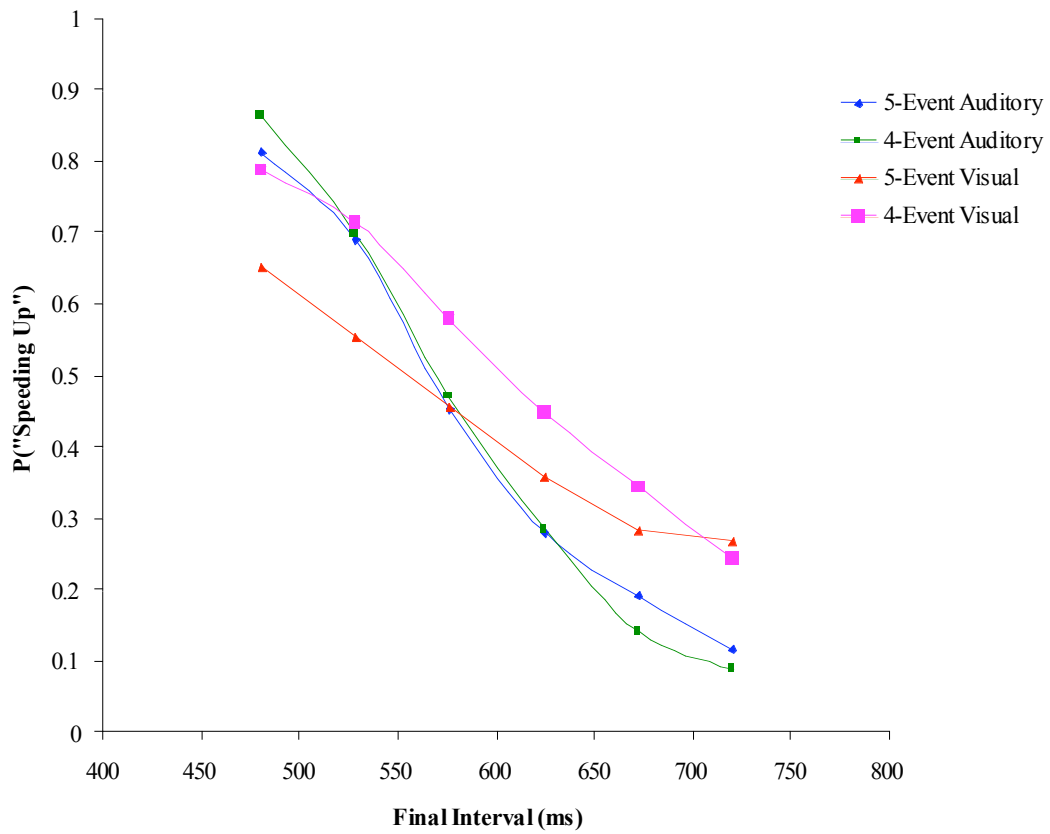


Figure 2: Proportion of “speeding up” responses by final interval. Final intervals reflect $\pm 4\%$, $\pm 12\%$ or $\pm 20\%$ of the implied 600ms IOI

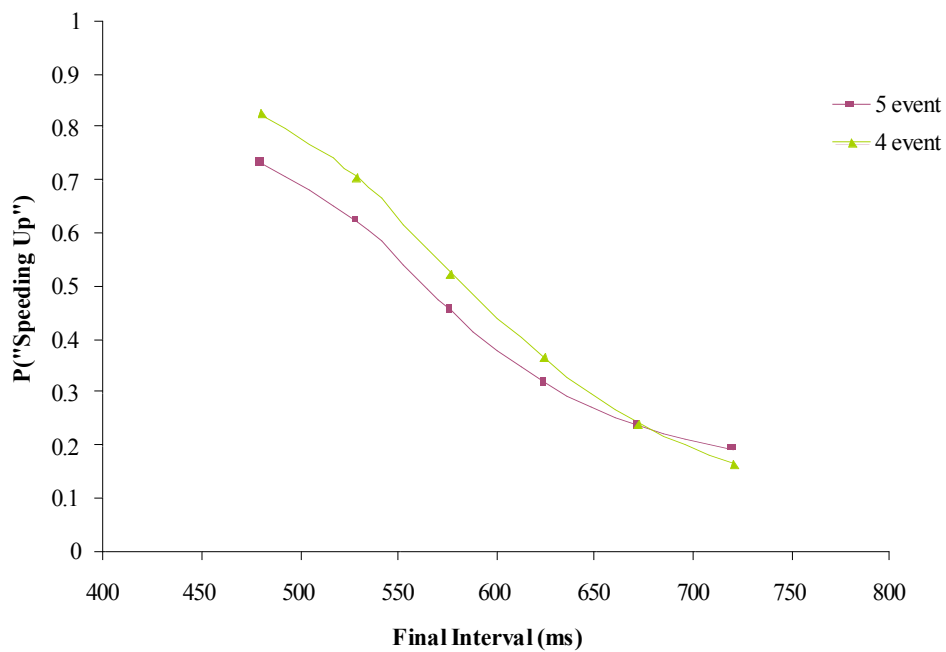
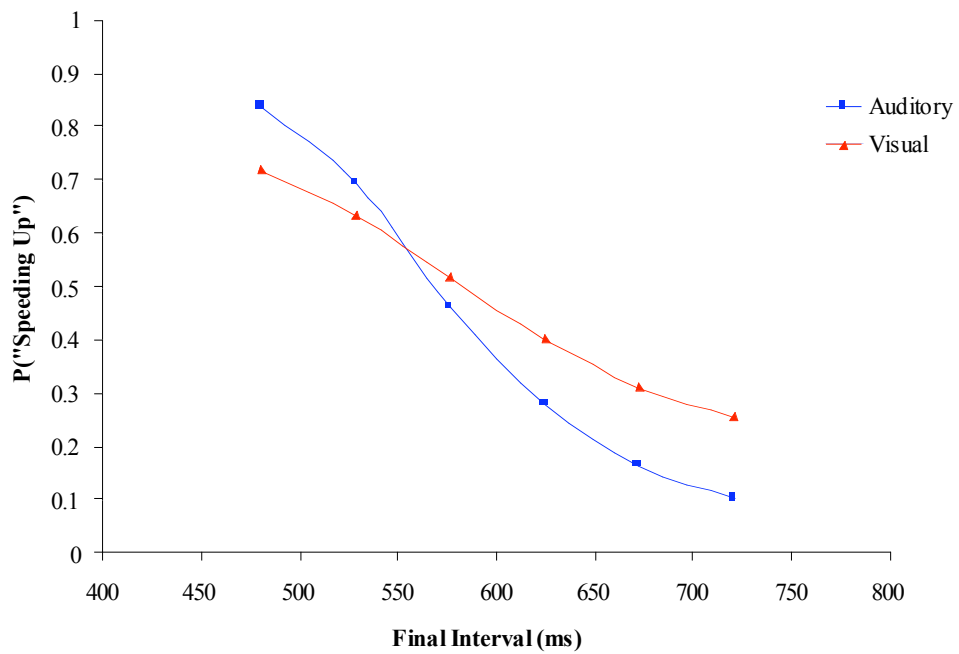


Figure 3: Proportion of “speeding up” responses by final interval. Proportions have been collapsed across sequence type and modality, and final intervals reflect $\pm 4\%$, $\pm 12\%$ or $\pm 20\%$ of the implied 600ms IOI.

For the analysis using the four types of w, there was a main effect of sequence $F(1, 30) = 5.534, p=.026$, with larger w values for the four-event sequence. There was also a trend towards a main effect of modality $F(1, 30) = 3.171, p=.085$, with higher w scores in the auditory modality. There was an interaction of sequence by modality $F(1, 30) = 8.082, p<.01$, with little difference between w values for the four-event sequence across the two modalities, but much larger values for auditory five-event than the visual five-event sequence. This shows that the 5-event sequences and not the 4-event sequences elicit different perceptual strategies in participants. As with the previous behavioral analysis, there was no main effect of whether the participants were presented with auditory or visual sequences first $F(1, 30) = 1.202, p>.05$.

Electrophysiological Data

It is important to examine differences in sensory-evoked potentials when looking for differences in temporal processing. For example, some listeners might attend more to the first and third tones of the 5-event pattern (a “1 and 2” count) because these stimuli reinforce the implied 600 ms beat; this would result in larger long-latency sensory-evoked responses to the first and third stimuli compared to the other groups (Picton, 1974). There were no significant differences as a function of group (based on auditory and visual dichotomous w) or sequence (5- versus 4-event) for any of the long-latency N1, P2, or N2 sensory-evoked responses as measured at electrode Cz. Amplitude for auditory responses diminished consistently after the first tone; the N1-P2-N2 complex

decreases in amplitude for subsequent tones presented in stimulus trains with IOIs of less than 10 seconds (Ford & Hillyard, 1981).

Four repeated measures ANOVA were performed for both the CNV and P3 mean amplitudes with different w values as a between subjects variable (using the auditory continuous w , auditory dichotomous w , visual continuous w , and visual dichotomous w values calculated from the 5-event condition, as mentioned previously). Factors included whether the final event occurred earlier or later than the implied 600 ms beat (minus/plus), whether the sequence had five or four events (5/4), and the modality of the sequence (auditory/visual). For the observed CNV, there was a main effect of modality for each of the ANOVAs that used different w variables (see Table 1), with larger CNV amplitudes in the auditory modality, suggesting the CNV indexes modality-specific processing. There was also a main effect of whether the final event came before or after the expected time (see Table 1), with larger CNV amplitudes when the final event occurred earlier than expected (See Figure 4).

Table 1

Results of CNV Analysis of Variance for w Values

Source	<i>df</i>	<i>F</i>	<i>p</i>
Auditory continuous w			
Modality	30	48.327	0.000*
Minus/Plus	30	20.501	0.003*
Auditory dichotomous w			
Modality	30	25.118	0.000*
Minus/Plus	30	15.172	0.001*
Visual continuous w			
Modality	30	8.227	0.013*
Minus/Plus	30	27.081	0.000*
Visual dichotomous w			
Modality	30	21.874	0.000*
Minus/Plus	30	18.93	0.000*

For the auditory continuous and auditory dichotomous ANOVAs, there was an interaction between modality and w ($F(1,30) = 3.782, p=.038$ and $F(1,30) = 7.920, p=.019$, respectively), with larger w values corresponding to a larger difference between auditory and visual CNV amplitudes, and smaller w values corresponding to a very small difference between auditory and visual CNV amplitudes. This suggests that the difference

between auditory and visual processing is much larger in beat-based than interval-based participants. There was also an interaction between modality and minus/plus for the auditory continuous and dichotomous conditions as well as the visual dichotomous ANOVAs ($F(1,30) = 9.265, p=.019$, $F(1,30) = 7.634, p=.010$, and $F(1,30) = 5.636, p=.024$, respectively), with CNV amplitudes that occurred earlier being much closer in size and being much larger than CNV amplitudes that occurred later. Visual CNV amplitudes showed a much larger amplitude difference between minus and plus than auditory CNV amplitudes, which showed little amplitude differences. There was a three-way modality by minus/plus by w interaction for the auditory dichotomous ANOVA only, $F(1,30)=8.337, p=.007$. When events occurred too early, there was little difference between auditory and visual CNV mean amplitude for interval based participants only. However, when events occurred too late, the difference between auditory and visual amplitudes matched that of the beat-based group. Finally, there was a minus/plus by w interaction for the visual dichotomous ANOVA only $F(1,30)=7.160, p=.012$, with the interval-based group showing much larger amplitudes when the events occurred too early.

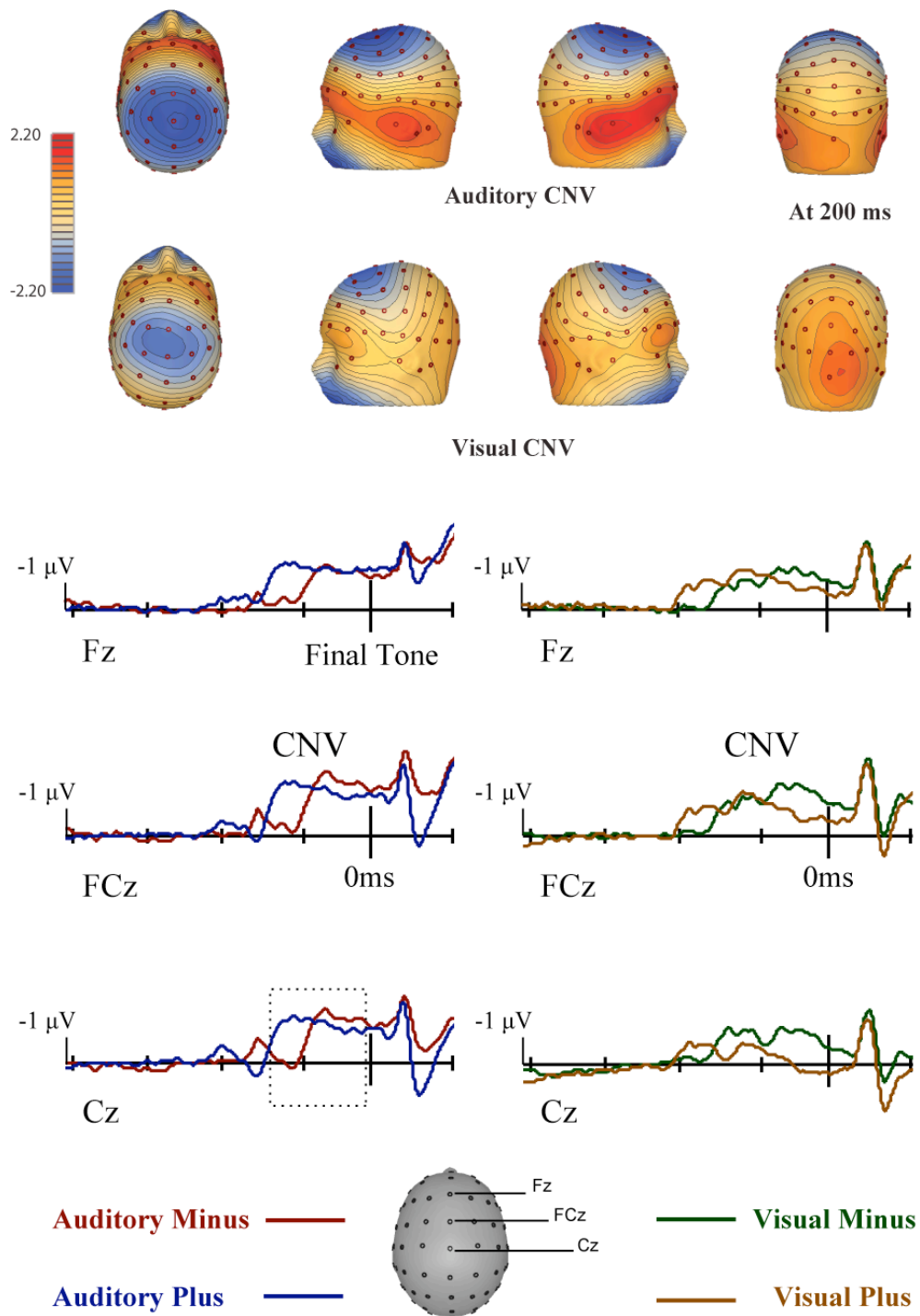


Figure 4: ERP traces and topographies for Auditory and Visual CNVs. ERPs have been collapsed across 4- and 5-event conditions, while topographies have been collapsed across 4- and 5-events, as well as minus and plus conditions.

When observing the voltage pattern across the scalp, the auditory CNV appears to show contributions from temporal areas, while the visual CNV appears to show contributions from occipital areas. To determine whether there were condition differences in pattern activity independent of amplitude differences, the data from all 72 electrodes was vector-normalized. A repeated-measures ANOVA was run to identify these differences across the scalp by testing for electrode x condition interactions. There was a significant interaction between modality and electrode $F(1, 30) = 7.16, p < .001$ that remained significant when the eye electrodes were removed (to exclude a possible confound of ocular artifacts), $F(1, 30) = 7.36, p < .001$ (see Figure 5).

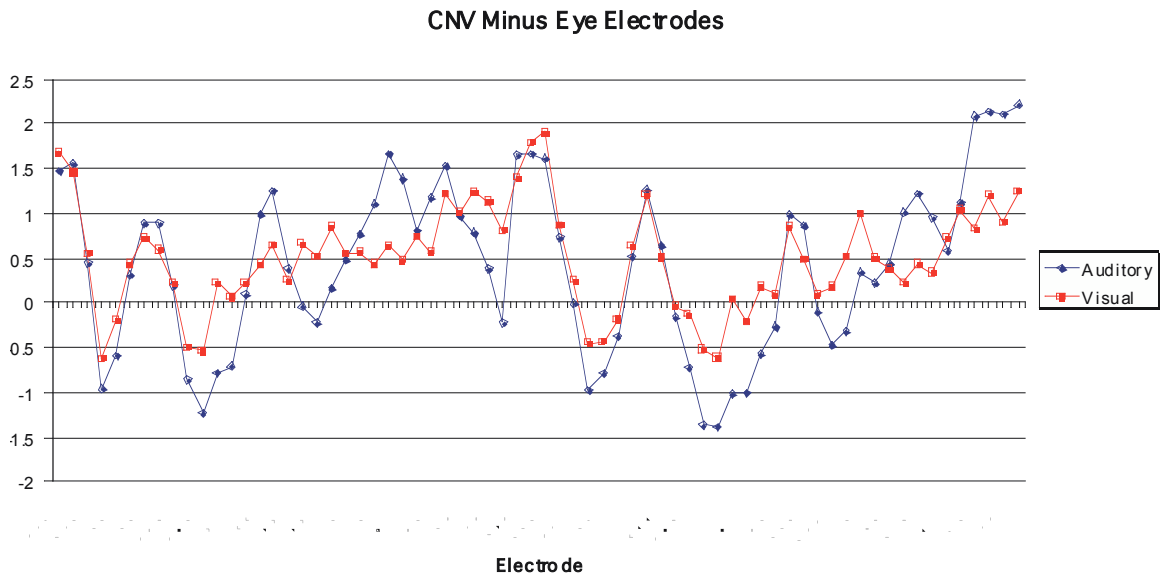


Figure 5: Mean amplitude across the 68 electrodes (64 scalp electrodes plus four non-ocular face electrodes).

While the CNV appears to be modality specific, the later, more cognitively controlled P3 might be expected to be modality general. After the final event, participants

must decide if the sequence speeds up or slows down, and this requirement to make a decision results in late components (Bender, 2008; Polich, 2007; Sutton, 1965). A distinct P3b occurred at the parietal electrodes (maximally at POz) for both auditory and visual stimuli. As with the CNV, four repeated measures ANOVAs were performed with different *w* values as a between subjects variable (using auditory continuous *w*, auditory dichotomous *w*, visual continuous *w*, and visual dichotomous *w* values). All four showed a main effect of minus/plus ($F(1,30)=22.356, p=.002$, $F(1,30)=5.994, p=.021$, $F(1,30)=8.886, p=.011$, and $F(1,30)=6.300, p=.018$, respectively). There were larger P3 amplitudes when the final event occurred earlier than expected, suggesting a partial novelty enhancement for events before the expected time, or decreased memory and cognitive resources when the event was recognized as occurring after the expected time. There was no main effect of modality ($p>.05$ for all conditions), suggesting the P3 may reflect modality-general processing (see Figure 6).

For the auditory continuous *w* value condition, there was a minus/plus by *w* ($F(1,30)=4.513, p=.024$) and a minus/plus by modality interaction ($F(1,30)=6.085, p=.043$). The auditory P3 amplitudes were much larger than the visual amplitudes when the event occurred too early, but there was little difference between modalities when the event occurred later than expected. The minus/plus by modality interaction was seen as a trend in both the auditory dichotomous and visual dichotomous conditions, but neither reached significance ($F(1,30)=3.645, p=.066$ and $F(1,30)=4.024, p=.054$, respectively). Finally, there was a significant modality by sequence by *w* interaction for the visual continuous condition $F(1,30)=2.979, p=.026$.

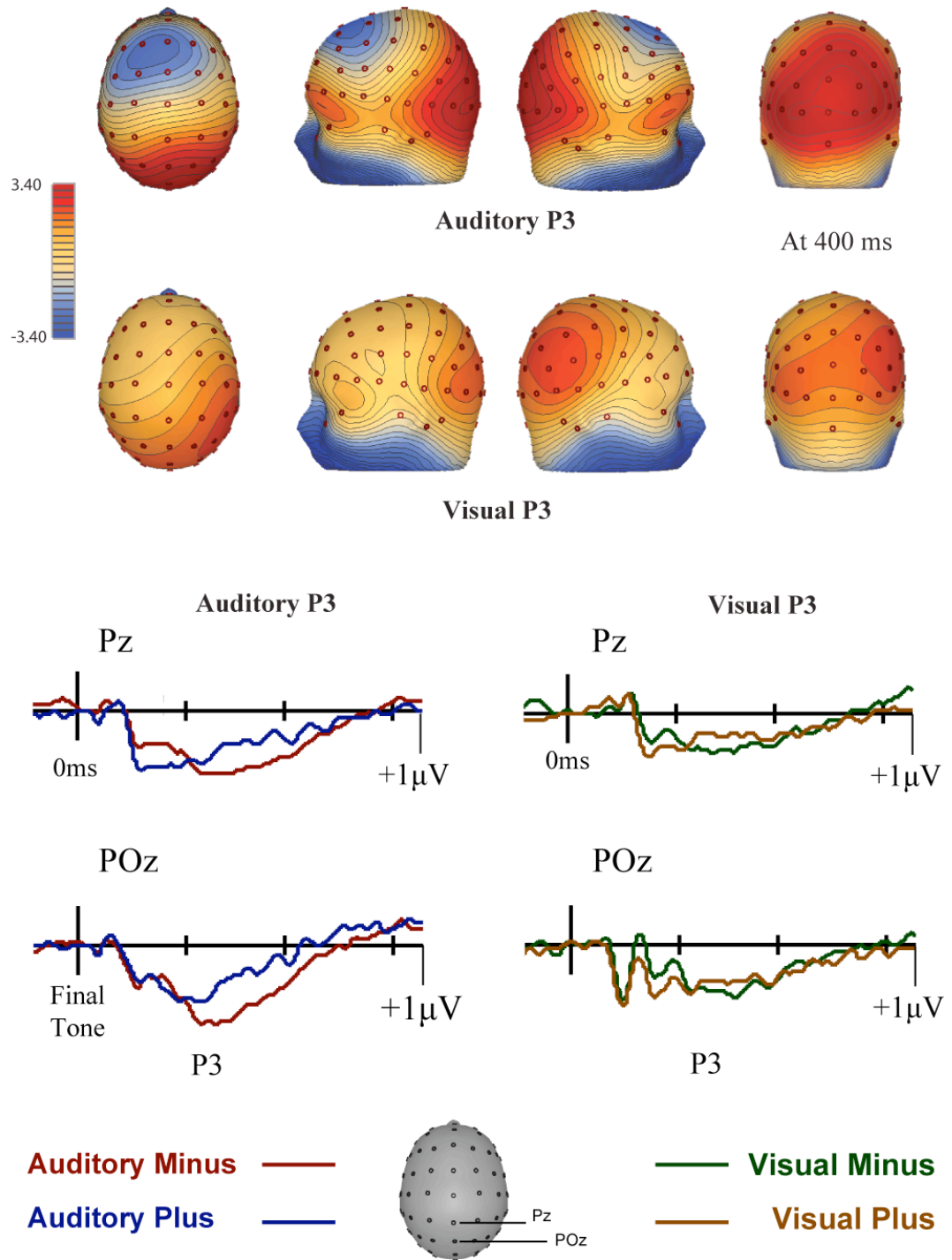


Figure 6: ERP traces and topographies for Auditory and Visual P3s. ERPs have been collapsed as with the CNV.

In contrast to the pattern of activity differences found for the CNV, activity across the scalp did not appear to come from modality-specific sources for the P3. Data from all 72 electrodes were again vector-normalized, and a repeated-measures ANOVA was run to testing for electrode x condition interactions. There was a significant interaction between modality and electrode $F(1, 30) = 2.08, p < .001$ that remained significant when the eye electrodes were removed; $F(1, 30) = 1.95, p < .001$ (see Figure 7).

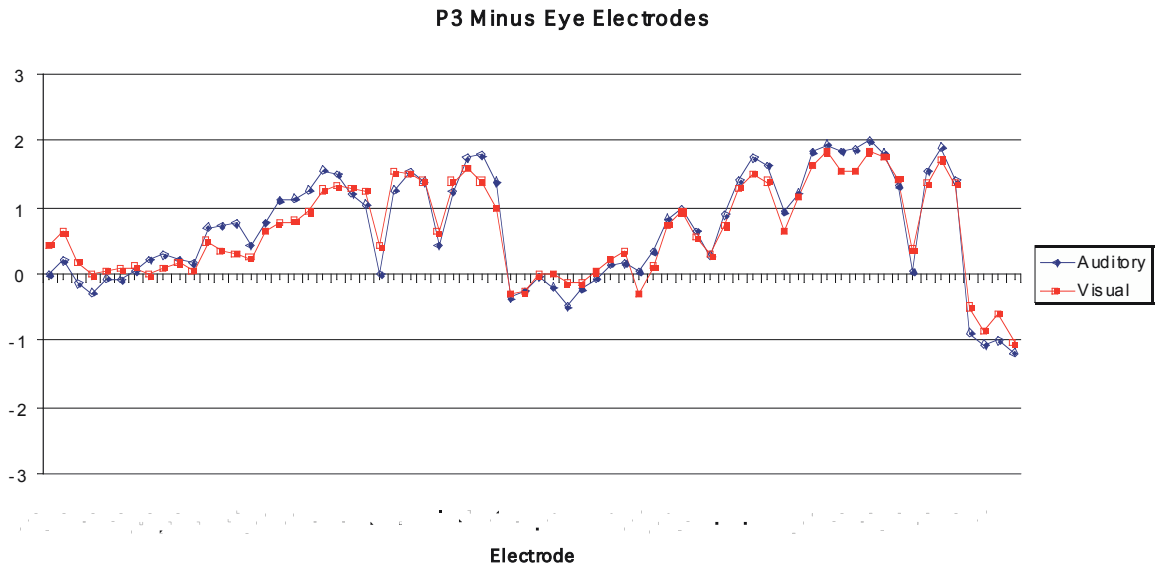


Figure 7: Mean amplitude across the 68 electrodes (64 scalp electrodes plus four non-ocular face electrodes).

Discussion

The current study used EEG to examine early sensory and later cognitive brain responses thought to contribute to rhythm perception. We used a previously established perceptual paradigm to elicit these brain responses; this paradigm is also known to behaviorally elicit different perceptual strategies (McAuley, Frater, Janke & Miller, 2003, McAuley & Henry, 2010, Grahn, Henry, & McAuley, 2011, and Snyder, Pasinski, & McAuley, 2011). Consistent with that research, participants in this study showed differing perceptual strategies for the test sequences but not the control sequences. Participants also correctly identified that the sequences sped up when deviants occurred before the implied 600 ms beat. There also appeared to be some flexibility in the use of perceptual strategies. Participants reported different proportions of the patterns speeding up; there was a greater chance of responding “speeding up” for the visual modality. An interaction between modality and sequence and a three-way interaction trend of modality by sequence by deviant size showed that a larger number of subjects changed between a beat-based to an interval-based strategy for visual sequences only, suggesting that the visual sequences were perceptually different enough from the auditory sequences for some participants to require a shift in strategy. This may be related to the decreased performance shown in the visual condition. This further suggests local intrinsic timing in the visual cortex that is not as detailed as in the auditory modality, or intrinsic visual information that takes a less efficient path to a central timer, requiring a cognitive shift in strategy.

There were also distinct effects of perceptual strategy on brain responses related to timing and decision-making. The first of these brain responses is the CNV, which is a late

negative component shown to reflect expectancy for the final stimulus of a pattern (Pfeuty, 2005; Walter, 1964). The CNV appears to be an on-line index of timing (Macar and Vidal, 2004), and the results of this study suggest it also indexes modality-specific timing. We found topography differences suggesting that the CNV arises from sensory specific areas: A visual CNV was elicited from occipital areas while an auditory CNV was elicited from temporal areas. There were larger CNV amplitudes in the auditory modality, suggesting that the auditory cortex may be more efficient in encoding timing information than the visual cortex. An interaction between modality and the type of perceptual strategy (as measured by w values) showed that the difference between auditory and visual processing is larger in participants with beat-based rather than interval-based strategies. Interval-based participants showed moderate CNV amplitudes for both visual and auditory modalities as compared to beat-based participants. The beat-based participants, however, showed greatly increased CNV amplitudes for the auditory modality and greatly diminished amplitudes for the visual modality as compared to the interval-based participants. It may be that the switch in perceptual strategy mentioned above from beat-based to interval-based for visual sequences may contribute to this cost to CNV amplitude.

There were also differences in whether the final event in the pattern occurred earlier or later than expected. There were larger CNV amplitudes when the final event occurred earlier than expected; studies by Pfeuty and colleagues (2003) and N'Diaye (2004) and colleagues showed that the CNV diminished sharply after the time of an expected tone, suggesting that the minus/plus difference we observed might be due to expectancy. Since our data were collapsed using the final tone as time 0 (with the CNV

being observed from -200 ms to 0), the CNV may have started to decrease before the final tone in the late sequences, leading to the amplitude differences we found. There were also differences in CNV amplitude based on listening strategy; when events occurred too early, there was little difference between auditory and visual CNV mean amplitude for interval-based participants only. However, when events occurred too late, the difference between auditory and visual amplitudes matched that of the beat-based group. This, paired with the interval-based group showing much larger amplitudes when the events occurred too early, suggests that participants with interval-based strategies are more susceptible to changes early on in expectation for the sequence. This is related to their perception of these sequences as almost always slowing down; there is more perceptual variability closer to their expected 300 ms final interval, and this may result in larger CNV amplitudes that diminish continuously after that point.

The modality-specificity of the CNV may also be seen in the observed interaction between minus/plus and modality. The CNV amplitudes were much closer in size for the auditory modality, but showed a much larger voltage difference in the visual modality as a function of minus/plus; the visual CNVs appeared to decay much quicker than the auditory CNVs, suggesting differences in processing between the modalities. Visual CNVs with smaller amplitudes that decay quicker than their auditory counterparts suggest that the visual modality may be less equipped to form temporal expectancies.

In contrast to the compelling evidence for a modality-specific CNV, the context-updating P3b may reflect a more attention-dependent and modality-general temporal decision-making process (Linden, 2005, and Polich, 2007). Unlike the CNV, there were no topography differences in the P3 nor were there any main effects of modality,

suggesting modality-generalty. However, there was a significant effect of whether the final event occurred earlier or later than expected; the P3 amplitudes were larger when the final event occurred earlier than expected. This could suggest a partial novelty enhancement for events before the expected time, consistent with the novelty P3, a variation of the more fronto-central P3a (Linden, 2005). Despite the more anterior generation of the P3a, overlap may contribute to an enhancement of the P3b (Polich, 2007). This novelty P3 may reflect more of an orienting response than an actual index of timing, so differences in P3 between early and late events should be interpreted differently than those found in the CNV. However, this difference in P3 amplitude might also reflect a decreased load to processing; more easily discriminated targets result in larger P3 amplitudes, as resources are not being tapped by working memory. Polich's context-updating version of the P3 states that as the drain on memory resources increases, the P3 amplitude decreases; since participants are holding the initial beat in memory, final tones that occur too late only have a small pool of resources left, and P3 amplitude decreases.

Both interpretations may be supported by an observed trend towards an interaction between minus/plus and modality. There was little to no difference between P3 amplitude to events occurring after the expected time across modalities; however, the auditory P3 amplitudes showed much larger amplitudes for auditory events occurring before the expected time. If the auditory modality is better at temporal processing, this may carry over to an increased novelty response for early P3s in the auditory modality that does not appear for early visual P3s. But because both auditory and visual P3s index decision-making, both suffer from decreased cognitive resources when events occur later than

expected. However, because the exact mechanisms of the P3 are still unknown, further research should focus on the exact contributions of memory, decision-making and novelty detection to this modality-general component.

The differences observed between the CNV and P3 suggest a mixture of sensory-level intrinsic activation and more cortical, dedicated, timing processes. Timing information appears to be encoded at a sensory level (consistent with the SDN model promoted by Buonomano, Bramen, Khodadadifar and others) and this short-term local activity appears to contribute to the timing-related CNV component. This modality-specific information appears to contribute to two CNV components: a larger, more efficient auditory CNV and a much smaller, less defined visual CNV. However, intrinsic models become less viable for intervals over a few hundred milliseconds, and a central timer may be required to consolidate or modulate this information via attention or other cortical contributions. This allows for modality-specific timing areas to fall under the cognitive control of a higher brain area (Lewis & Miall, 2006; Noguchi & Kakigi, 2006; Bueti, van Dongen, & Walsh, 2010). This central timer may allocate the memory and novelty-detection resources responsible for producing the decision-related P3b. This component does not show the modality differences found in the CNV, and reflects activity from this modality-general, memory-dependent timer.

APPENDIX 1
IRB APPROVALS

Social/Behavioral

IRB – Expedited

Review

Continuing Review Approved

NOTICE TO ALL RESEARCHERS:

Please be aware that a protocol violation (e.g., failure to submit a modification for any change) of an IRB approved protocol may result in mandatory remedial education, additional audits, re-consenting subjects, researcher probation, suspension of any research protocol at issue, suspension of additional existing research protocols, invalidation of all research conducted under the research protocol at issue, and further appropriate consequences as determined by the IRB and the Institutional Officer.

January 21, 2011

Dr. Joel Snyder, Psychology

Office of Research Integrity – Human Subjects

Notification of IRB Action by /Ramona Denby/

Dr. Ramona Denby Brinson, Chair

Protocol Title: Neural Mechanisms of Auditory and Visual Processing in Healthy Adults

Protocol #: 0710-2518

Expiration Date: January 20, 2012

Continuing review of the protocol named above has been reviewed and approved.

This IRB action will reset your expiration date for this protocol. The protocol is approved for a period of one year from the date of IRB approval. The new expiration date for this protocol is January 20, 2011. If the above-referenced project has not been completed by this date you must request renewal by submitting a Continuing Review Request form 30 days before the expiration date.

PLEASE NOTE:

Upon approval, the research team is responsible for conducting the research as stated in the protocol most recently reviewed and approved by the IRB, which shall include using the most recently submitted Informed Consent/Assent forms and recruitment materials.

The official versions of these forms are indicated by footer which contains current approval and expiration dates.

Should there be any change to the protocol, it will be necessary to submit a Modification Form through ORI - Human Subjects. No changes may be made to the existing protocol until modifications have been approved by the IRB. Modified versions of protocol materials must be used upon review and approval. Unanticipated problems, deviations to protocols, and adverse events must be reported to the ORI – HS within 10 days of occurrence.

If you have questions or require any assistance, please contact the Office of Research Integrity - Human

Subjects at IRB@unlv.edu or call 895-2794.

REFERENCES

- Basso, G., Nichelli, P., Wharton, C.M., Peterson, M. & Grafman, J. (2003). Distributed neural systems for temporal production: a functional MRI study. *Brain Res. Bull.*, 59, 405–411.
- Bender, S., Oelkers-Ax, R., Hellwig, S., Resch, F., & Weisbrod, M. (2008). The topography of the scalp-recorded visual N700. *Clinical Neurophysiology*, 119, 587-604.
- Buchwald, J.S. & Huang, C-M. (1975). Far-field acoustic response: Origins in the cat. *Science*, 189, 382-394.
- Bueti D., Bahrami B., Walsh V. (2008). Sensory and association cortex in time perception. *Journal of Cognitive Neuroscience* 20, 1054–1062.
- Bueti, D., Walsh, V., Frith, C., and Rees, G. (2008). Different Brain Circuits Underlie Motor and Perceptual Representations of Temporal Intervals. *Journal of Cognitive Neuroscience* 20, 204–214.
- Bueti D., van Dongen E.V., Walsh V. (2008) The Role of Superior Temporal Cortex in Auditory Timing. *PLoS ONE* 3(6): e2481.
- Buonomano, D.V. (2000). Decoding Temporal Information: A Model Based on Short-Term Synaptic Plasticity. *The Journal of Neuroscience*, 20, 1129–1141.
- Buonomano, D.V., Bramen, J., & Khodadadifar, M. (2009). Influence of the interstimulus interval on temporal processing and learning: testing the state-dependent network model. *Phil. Trans. R. Soc. B*, 364, 1865-1873.
- Butler, R.A., 1968. Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. *J. Acoust. Soc. Am.* 44, 945–950.

- Church, R.M., & Broadbent, H.A., (1990). Alternative representations of time, number, and rate. *Cognition*, 37: 55-81.
- Cui, R.Q., Egkher, A., Huter, D., Lang, W., Lindinger, G., & Deecke, L. (2000). High resolution spatiotemporal analysis of the contingent negative variation in simple or complex motor tasks and a non-motor task. *Clinical Neurophysiology*, 111, 1847-1859.
- Di Russo, F., Martínez, A., Sereno, M.I., Pitzalis, S., & Hillyard, S.A. (2001). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, 15, 95-111.
- Diaconescu, A.O., Kovacevic, N., & McIntosh, A.R. (2008). Modality-independent processes in cued motor preparation revealed by cortical potentials. *NeuroImage*, 42, 1255-1265.
- Ford, J.M., & Hillyard, S.A. (1981). Event-Related Potentials (ERPs) to interruptions of a steady rhythm. *Psychophysiology*, 18, 322-330.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279-325.
- Gómez, C.M., Fernández, A., Maestú, F., Amo, C., González-Rosa, J.J., Vaquero, E., & Ortiz, T. (2004). Task-specific sensory and motor preparatory activation revealed by contingent magnetic variation. *Cognitive Brain Research*, 21, 59-68.
- Gontier, Dantec, Paul, and Bernard (2008)
- Grahn, J. A. & Brett, M. (2007). Rhythm and Beat Perception in Motor Areas of the Brain. *Journal of Cognitive Neuroscience* 19(5), 893–906.

- Grahn, J. A., Henry, M. J., & McAuley, J. D. (2011). FMRI investigation of cross-modal interactions in rhythm perception: Audition primes vision, but not vice versa. *NeuroImage*, 54, 1231-1243.
- Guttman, S.E., Gilroy, L.A., & Blake, R. (2005). Hearing What the Eyes See: Auditory Encoding of Visual Temporal Sequences. *Psychological Sciences*, 16, 228-235.
- Harrington, D.L., Boyd, L.A., Mayer, A.R., Sheltraw, D.M., Lee, R.R., Huang, M., & Rao, S.M. (2004). Neural representation of interval encoding and decision making. *Cognitive Brain Research*, 21, 193-205.
- Hazeltine, E., Helmuth, L. L., & Ivry, R.B. (1997). Neural mechanisms of timing. *Trends in Cognitive Sciences*, 1, 163-169.
- Ikeda, A., Lüders, H.O., Collura, T.F., Burgess, R.C., Morris, H.H., Hamano, T., Shibasaki, H. (1996). Subdural potentials at orbitofrontal and mesial prefrontal areas accompanying anticipation and decision making in humans: a comparison with Bereitschaftspotential. *Electroencephalography and Clinical Neurophysiology*, 98, 206-212.
- Ivry, R.B. & Schlerf, J.E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, 12, 273-280.
- Jefferys, D.A. & Axford, J.G. (1972). Source locations of pattern-specific components of human visual evoked potentials. I. Component of striate cortical origin. *Experimental Brain Research*, 16, 1-21.
- Jefferys, D.A. & Axford, J.G. (1972). Source locations of pattern-specific components of human visual evoked potentials. II. Component of extrastriate cortical origin. *Experimental Brain Research*, 16, 22-40.

- Jin, D.Z., Fujii, N., & Graybiel, A.M. (2009). Neural representation of time in cortico-basal ganglia circuits. *PNAS*, *106*, 19156–19161.
- Jongsma, M.L.A., Meeuwissen, E., Vos, P.G., & Maes, R. (2007). Rhythm perception: Speeding up or slowing down affects different subcomponents of the ERP P3 complex. *Biological Psychology*, *75*, 219-228.
- Karabanov, A., Blom, Ö., Forsman, L. & Ullén, F. (2009). The dorsal auditory pathway is involved in performance of both visual and auditory rhythms. *NeuroImage*, *44*, 480-488.
- Karmarkar, U. R. & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learn. Mem.* *10*, 141–147.
- Karmarkar, U. R. & Buonomano, D. V. (2007) Timing in the absence of clocks: encoding time in neural network states. *Neuron*, *53*, 427–438.
- Keele, S. W., Nicoletti, R., Ivry, R. I., & Pokorny, R. A. (1989). Mechanisms of perceptual timing: Beat-based or interval-based judgements? *Psychological Research*, *50*, 251-256.
- Le Dantec, C., Grontier, E., Paul, I., Charvin, H., Bernard, C., Lalonde, R., & Rebaï, M. (2007). ERPs associated with visual duration discriminations in prefrontal and parietal cortex. *Acta Psychologica*, *125*, 85-98.
- Lewis, P.A., & Miall, R.C. (2006). Remembering the time: a continuous clock. *Trends in Cognitive Sciences* *10*, 401-406.
- Linden, D.E.J. (2005). The P300: Where in the brain is it produced and what does it tell us? *Neuroscientist*, *11*, 563-576.
- Luck, S.J., Heinze, H.J., Mangun, G.R., & Hillyard, S.A. (1990). Visual event-related

- potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. *Electroencephalography and Clinical Neurophysiology*, 75, 528-542.
- Macar, F. & Vidal, F. (2004). Event-Related potentials as indices of time processing: A review. *Journal of Psychophysiology*, 18, 89-104.
- Maier, J., Dagnelie, G. Spekreijse, H. & van Dijk, B.W. (1987). Principle components analysis for source localization of VEPs in man. *Vision Research*, 27, 165-177.
- Martin, B.A. & Boothroyd, A. (2000). Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude. *J. Acoust. Soc. Am.* 107, 2155-2161.
- McAuley, J.D., Frater D., Janke, K., & Miller, N.S. (2003). Detecting changes in timing: Evidence for two modes of listening. International Conference on Music Perception and Cognition - Proceedings.
- McAuley, J. D. and Henry, M.J. (2010). Modality effects in rhythm processing: Auditory encoding of visual rhythms is neither obligatory nor automatic. *Attention, Perception & Psychophysics*, 72, 1377 – 1389.
- McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception and Performance*, 29(6), 1102-1125.
- N'Diaye, K., Ragot, R., Garnero, L. & Pouthas, V. (2004). What is common to brain activity evoked by the perception of visual and auditory filled durations? A study with MEG and EEG co-recordings. *Cognitive Brain Research* 21, 250 – 268.

- Näätänen, R. & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24, 375-425.
- Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E.N. (1988). Frequency and location specificity of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology*, 69, 523-531.
- Noguchi, Y. & Kakigi, R. (2006). Time Representations Can Be Made from Nontemporal Information in the Brain: An MEG Study. *Cerebral Cortex*, 16, 1797-1808.
- Pashler, H. (2001). Perception and production of brief durations: Beat-based versus interval-based timing. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 485-493.
- Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. (2005). The influence of metricity and modality on synchronization with a beat. *Experimental Brain Research*, 163, 226-238.
- Penhune, V. B., Zatorre, R. J., & Evans, A. C. (1998). Cerebellar Contributions to Motor Timing: A PET Study of Auditory and Visual Rhythm Reproduction. *Journal of Cognitive Neuroscience* 10, 752–765
- Pfeuty, M., Ragot, R., & Pouthas, V. (2003). Processes involved in tempo perception: A CNV analysis. *Psychophysiology*, 40, 69–76.
- Picton, T. W., Campbell, K. B., Baribeau-Braun, J. & Proulx, G. B. The neurophysiology of human attention: A tutorial review. In J. Requin (Ed.), *Attention and performance VII*, Hillsdale. New Jersey: Lawrence Erlbaum. 1978. Pp. 429-467.

- Picton, T.W., Hillyard, S.A., Krausz, H.I. & Galambos, R. (1974). Human auditory evoked potentials. I: Evaluation of components. *Electroencephalography and Clinical Neurophysiology*, 36, 179-190.
- Picton, T.W., & Hillyard, S.A. (1974). Human auditory evoked potentials. II: Effects of attention. *Electroencephalography and Clinical Neurophysiology*, 36, 191-199.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128–2148.
- Povel, D., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics*, 30, 565-572.
- Repp, B.H. (2003). Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision. *Journal of Motor Behavior*, 35, 355-370.
- Repp, B.H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1085-1099.
- Schulze, H-H. (1978). The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research*, 40, 172-181.
- Shih, L.Y.L., Kuo, W.-J., Yeh, T.-C., Tzeng, O.J.L., & Hsieh, J.-C. (2009). Common neural mechanisms for explicit timing in the sub-second range. *NeuroReport*, 20, 897 -901.

Snyder, J.S., Pasinski, A.C., & McAuley, J.D., (2011). Listening strategy for auditory rhythms modulates neural correlates of expectancy and cognitive processing.

Psychophysiology, 48, 198-207.

Thurlow, W. R. & Jack, C. E. (1973). Certain determinants of the ventriloquism effect.

Perceptual and Motor Skills, 36, 1171-1184.

Walter, W.G., Cooper, R., Aldridge, V.J., McCallum, W.C., & Winter, A.L. (1964).

Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature*, 203, 380-384.

Yvert, B., Crouzeix, A., Bertrand, O., Seither-Preisler, A., & Pantev, C. (2001). Multiple supratemporal sources of magnetic and electric auditory evoked middle latency components in humans. *Cerebral Cortex*, 11, 411-423.

VITA

Graduate College
University of Nevada, Las Vegas

Amanda Claire Pasinski

Degrees:

Bachelor of Art, Psychology, 2007
University of Nevada, Las Vegas

Thesis Title: Modality-Specific and Modality-General Encoding of Auditory and Visual Rhythms

Thesis Examination Committee:

Chairperson, Joel S. Snyder, Ph. D.
Committee Member, Erin Hannon, Ph. D.
Committee Member, Jefferson Kinney, Ph. D.
Graduate Faculty Representative, Eugenie Burkett, Ph. D.