

# Listening strategy for auditory rhythms modulates neural correlates of expectancy and cognitive processing

JOEL S. SNYDER,<sup>a</sup> AMANDA C. PASINSKI,<sup>a</sup> AND J. DEVIN MCAULEY<sup>b,c</sup>

<sup>a</sup>Department of Psychology, University of Nevada Las Vegas, Las Vegas, Nevada

<sup>b</sup>Department of Psychology, Bowling Green State University, Bowling Green, Ohio

<sup>c</sup>Department of Psychology, Michigan State University, East Lansing, Michigan

## Abstract

A recently described auditory tempo perception paradigm revealed individual differences in perceived stimulus timing for identical stimulus sequences. The current study takes advantage of this paradigm by recording event-related potentials (ERPs) concurrent with task performance in order to reveal brain responses that reflect individual differences in timing strategy. No strategy-related differences were observed in sensory encoding of tones, as measured by the P1-N1-P2 complex. However, the contingent negative variation (CNV) leading up to the final tone of the sequence varied as a function of strategy, as did a parietal-maximum late positive component (P3b) that occurred following the final tone. These data suggest that temporal expectancy for and cognitive processing of the final tone of rhythmic sequences underlies differences in strategy during rhythm perception.

**Descriptors:** Rhythm perception, Tempo, Expectancy, Individual differences, Event-related potentials

Processing temporal structure is a fundamental aspect of sensory, motor, and cognitive function in many complex organisms. In humans, temporal structure serves as an important cue to distinguishing different languages from early in life (Nazzi, Bertoni, & Mehler, 1998), and in identifying the beat from music in order to coordinate movement in time during tapping, dancing, and music-making (Drake, Penel, & Bigand, 2000; Hannon, Snyder, Eerola, & Krumhansl, 2004; Snyder & Krumhansl, 2001; Toiviainen & Snyder, 2003; van Noorden & Moelants, 1999). But despite a long history of empirical and theoretical work, the mechanisms of timing ability in the human brain are still not well understood (Ivry & Schlerf, 2008; Zatorre, Chen, & Penhune, 2007). For example, the precise role of individual brain structures such as the cerebellum, basal ganglia, premotor cortex, supplementary motor area, and sensory cortices and how they interact in timing tasks is still a matter of debate. And what type of computational mechanism is used by humans to encode time is also unclear (Keele, Nicoletti, Ivry, & Pokorny, 1989; Martin, Egly, Houck, Bish, Barrera, et al., 2005; McAuley & Jones, 2003; Pashler, 2001; Schulze, 1978).

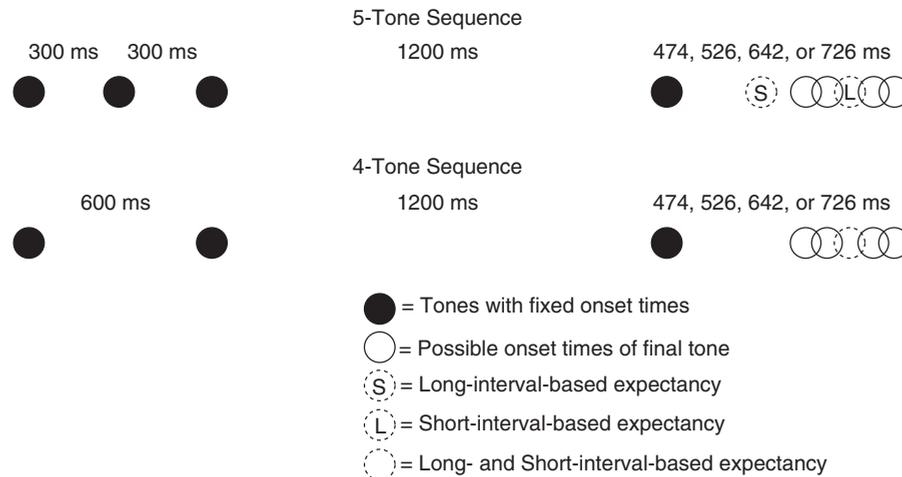
One new approach to studying timing mechanisms is to identify neural correlates of individual differences in timing strategy.

This work was supported by a summer research stipend from the College of Liberal Arts and a research development award from the Office for Research and Graduate Studies at the University of Nevada Las Vegas (J.S.S.) and grants from the GRAMMY Foundation and the National Science Foundation (J.D.M.). The authors thank Jessica Grahn for comments on an earlier draft of the paper.

Address correspondence to: Joel S. Snyder, Department of Psychology, University of Nevada Las Vegas, 4505 Maryland Parkway—Mail Stop 5030, Las Vegas, NV 89154-5030. E-mail: Joel.Snyder@unlv.edu

Once identified, such neural correlates can provide clues as to the importance of particular stages of processing or particular brain structures in timing behavior. McAuley, Frater, Janke, and Miller (2006) studied perception of a five-tone rhythm (see Figure 1), predicted to be perceived differently depending on strategy. One group of listeners heard the pattern speeding up or slowing down when the final interval was shorter or longer than 600 ms, respectively, possibly the result of imposing a 600-ms beat on the three tones at the beginning of the pattern, continuing the beat, and comparing it with the timing of the final tone. Another group of listeners always heard the pattern slowing down, possibly the result of implicitly calculating a pair of 300-ms intervals at the beginning of the pattern and comparing this stored 300-ms interval with the much longer final interval. The two groups did not differ on a four-tone rhythm (see Figure 1), predicted to result in the same pattern of behavior regardless of strategy. Although McAuley et al. (2006) interpreted the observed individual differences as reflecting the extent of engaging beat-based (Large & Jones, 1999; Large & Snyder, 2009; McAuley, 1995; Schulze, 1978) versus interval-based (Treisman, 1963) mechanisms, it should be noted that it is also possible that the individual differences arose from different tempo preferences. For example, it is possible that, when presented with a sequence containing an ambiguous reference interval duration (i.e., 300 or 600 ms), some participants prefer to attend to the long interval whereas others prefer to attend to the short interval.

Regardless of the precise nature of the individual differences revealed by the ambiguous tempo perception paradigm, these individual differences present a unique opportunity to identify brain activity related to different timing strategies, while controlling for stimulus effects. Recently, Grahn and McAuley (2009)



**Figure 1.** Stimuli. Two types of sequences were presented to participants, consisting of an initial group of three or two tones (for the five-tone and four-tone sequences, respectively) and a final group of two tones with a variable final IOI (474, 526, 642, or 726 ms). The task was to indicate whether the sequence was speeding up or slowing down at the end. If a listener adopts a weakly long-interval-based strategy, five-tone sequences would be expected to be perceived as always slowing down. In contrast, if a listener adopts a strongly long-interval-based strategy, then five-tone sequences would be expected to be heard as speeding up or slowing down depending on the final interval. Regardless of strategy, four-tone sequences would be expected to be heard as speeding up or slowing down depending on the final interval.

used this tempo paradigm in a functional magnetic resonance imaging (fMRI) study. They identified a number of specific brain areas activated during the task and found differences between long- and short-interval attending groups in mostly left-hemisphere structures (Grahn & McAuley, 2009). In particular, the inferior frontal gyrus, supplementary motor area, medial prefrontal cortex, and insula/ventrolateral prefrontal cortex, all on the left side, showed more activity for the long-interval group. The left superior temporal gyrus, left middle temporal gyrus, and right premotor cortex showed more activity for the short-interval group. The activity differences occurred while participants listened to both the five- and four-tone sequences, providing strong evidence that differences in neural activity related to timing strategy reflect general processing differences, even when no behavioral differences are apparent in the four-tone sequence. The limited temporal resolution of fMRI, however, prevented any strong conclusions about when, in the time course of processing, strategy-related differences occurred. Although group differences in prefrontal brain regions suggest the importance of higher-order cognitive aspects of processing, differences in the superior temporal lobe suggest the importance of lower-level auditory sensory processing in explaining differences in perception.

The current study used the ambiguous tempo paradigm while recording event-related brain potentials (ERPs) in human listeners. Using ERPs specifically allowed us to examine whether strategy-related differences in neural processing were the result of sensory processing of stimulus events, forming expectancies for the final (variable) interval, or comparing the final interval with previous intervals and making a decision. We hypothesized that ERP correlates of early sensory processing, namely the P1-N1-P2 long-latency auditory responses, might explain group differences in superior temporal lobe activations observed using fMRI (Grahn & McAuley, 2009). This could arise from differential attention to particular tones of the rhythmic sequence (Picton & Hillyard, 1974). For example, listeners with a preference for the long 600 ms intervals might attend more to the first and third tones of the five-tone sequence because these tones mark the

perceived 600-ms beat; this would result in larger long-latency sensory-evoked responses to the first and third tones compared to the other groups.

On the other hand, the differential activations in the supplementary motor area suggest the importance of expectancy processes, which can be measured using the contingent negative variation (Macar, Vidal, & Casini, 1999; Pfeuty, Ragot, & Pouthas, 2003; Walter, Winter, Cooper, McCallum, & Aldridge, 1964). In particular, it is possible that attending to a 300-ms interval might result in contingent negative variation (CNV) time course that ends earlier or is smaller overall than if attending to a 600-ms interval. Finally, differential activations observed in other frontal areas suggest the involvement of expectancy violation or memory updating mechanisms, as indexed by P3a and P3b, respectively (Polich, 2007), which have been previously observed to be important in rhythm perception tasks (e.g., Besson, Faita, & Requin, 1994; Brochard, Abecasis, Potter, Ragot, & Drake, 2003; Ford & Hillyard, 1981; Jongsma, Desain, & Honing, 2004).

## Materials and Methods

### Participants

Thirty-eight adults (15 men and 23 women, age range = 18–42 years, mean age = 23.76 years) with normal hearing ( $\leq 30$  dB HL from 250–4000 Hz) from the University of Nevada, Las Vegas Psychology subject pool participated for course credit after giving written informed consent. An additional eight participants were tested but were not included in the final sample because of poor behavioral performance in the four-tone control condition (5 participants) or because of poor-quality electrophysiological data (3 participants).

### Stimuli, Design, and Procedure

The stimulus used to construct the rhythmic sequences was generated off-line in MATLAB (The MathWorks, Inc., Natick, MA) and consisted of a single pure tone (424 Hz, 50 ms in

duration, including 10-ms rise/fall times). Sequences using this tone were presented binaurally through ER3A headphones (Etymotic Research, Inc., Elk Grove Village, IL) at 70 dB SPL. Behavioral responses were made on a RB-830 button box (Cedrus Corporation, San Pedro, CA). All aspects of stimulus presentation and behavioral response collection were controlled by a custom program written in Presentation (Neurobehavioral Systems, Inc., Albany, CA), running on a Pentium 4 computer with a SB X-Fi sound card (Creative Technology, Ltd.).

Two types of sequences were presented using the stimulus tone (Figure 1). Five-tone sequences consisted of 3 initial tones marking two 300-ms inter-onset intervals (IOIs) followed by 2 tones that specified a variable final IOI (600 ms  $\pm$   $\Delta T$ ). The key aspect of the five-tone sequence is that a periodic 600-ms beat is implied (but not explicitly emphasized) by the temporal structure of the sequence (Povel & Essens, 1985). Four-tone sequences, in contrast, consisted of 2 initial tones that specified a 600-ms IOI followed by 2 tones marking the same variable final IOI (600 ms  $\pm$   $\Delta T$ ). Thus, the only physical difference between the two types of sequences is that the four-tone sequence condition does not include the 2nd tone from the five-tone sequence condition. For both sequence types, 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 ms  $\pm$   $\Delta T$ , where  $\Delta T$  equaled  $\pm$  7% or  $\pm$  21% of the final IOI (474, 558, 642, or 726 ms).

Each participant listened to both the five-tone and four-tone sequences with all four final IOIs, resulting in two within-subjects factors (sequence type and final IOI). Six blocks were presented, each containing 136 trials (17 of each trial type). Thus, each trial type was presented 102 times to each participant. Eight practice trials, one of each trial type, were presented prior to beginning the experiment. Participants were seated in a comfortable chair in a single-walled sound-attenuated room (Industrial Acoustics Corp., Bronx, NY) and were asked to maintain fixation on a white cross on a black background in the center of a computer screen throughout the experiment. Participants were asked to listen to the stimuli during electrophysiological recording, and to avoid moving their eyes, head, or other body parts while the sequences were 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.” There was a 2-s interstimulus interval during which participants made their responses before a new trial began.

### Electrophysiological Recording

Electroencephalographic (EEG) signals were 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 (see <http://www.biosemi.com/faq/cms&drl.htm>), 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. EEG signals were digitized continuously (512 Hz sampling rate and a 104 Hz bandwidth) using a Biosemi ActiveTwo system (<http://www.biosemi.com>). 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 fit into place at each site. Sintered Ag-AgCl flat-type electrodes were attached with adhesive to sites below the hair line. No abrading of the skin was performed. Voltage offsets were below 40 mV prior to recording

and the resting EEG was checked for any problematic electrodes prior to and throughout the recording session.

### Data Analysis

Proportion of ‘speeding up’ responses was calculated for each participant for each of the 8 trial types (2 sequence types  $\times$  4 final IOIs). To quantify the extent of long-interval-based responding, Response proportions for five-tone (test) sequences were fit with a simple contrast model in order to assess the extent to which participants’ tempo judgments about the five-tone (test) sequences were based on a 300-ms referent interval or a 600-ms referent interval; see Grahn and McAuley (2009) for full details of the model. In the model, binary (‘speeding up’/‘slowing down’) judgments on a given trial are assumed to be based on one of two temporal referents:  $P = 300$  ms corresponding to the explicit time interval marked by the first three tones of the sequences (short temporal referent), and  $P = 600$  ms corresponding to the implied beat (long temporal referent). For each final IOI of the sequence,  $T_i$ , a temporal contrast metric, and  $C_i$ , is calculated, which measures the normalized difference between the final IOI and each referent,  $P$ :

$$C_i = \frac{(T_i - P)}{P}$$

Previous work has shown that the temporal contrast metric is a good index of the information that participants use to make time judgment decisions (McAuley & Jones, 2003). Because there are two possible temporal referents, each final IOI,  $T_i$ , results in two values of  $C_i$ , labeled here as  $C_{i300}$  for the  $P = 300$ -ms referent and  $C_{i600}$  for the  $P = 600$ -ms referent. In line with standard signal detection assumptions (Macmillan & Creelman, 1991), values of temporal contrast for each referent are assumed to be normally distributed with standard deviation,  $\sigma$ ; the values of  $C_{i300}$  and  $C_{i600}$  were then z-transformed and combined using a simple weighted average:

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

Predicted proportions of ‘speeding up’ responses,  $P$ (‘Speeding Up’), for each final interval,  $T_i$ , are then generated using cumulative normal distribution function:

$$P(\text{‘Speeding Up}) = 1 - \phi(z)$$

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

Most important for the present purposes, the continuous value of  $w$  provided an estimate of the extent to which tempo judgments for each participant about the five-tone (test) sequences were based on the explicit 300-ms (short) temporal referent or the implied 600-ms (long) temporal referent, with the latter consistent with a beat-based listening strategy. Thus, a participant with larger values of  $w$  shows greater tendency to judge tempo using the long temporal referent interval.

Participants were divided into three roughly equal groups according to  $w$ , yielding strongly long-interval based (SLI,  $n = 12$ ), moderately long-interval based (MLI,  $n = 13$ ) and weakly long-interval based (WLI,  $n = 13$ ) listener groups. Response proportions were then entered into a mixed-measures analysis of variance (ANOVA) to test for differences in perception depending on the within-subjects variables sequence (4 tones vs. 5 tones) and final IOI ( $-21, -7, +7, +21\%$ ), and the between-subjects variable group (SLI, MLI, WLI). The degrees of freedom were adjusted with the Greenhouse-Geisser epsilon ( $\epsilon$ ) when

appropriate, and all reported probability estimates were based on the reduced degrees of freedom. This adjustment was applied to all ANOVAs.

All off-line ERP analyses were performed using Brain Electrical Source Analysis software (BESA, MEGIS Software GmbH, Gräfelfing, Germany), except for baseline correction and amplitude measurements, which were performed by custom scripts in MATLAB (The MathWorks, Inc.). Electrodes that were noted during the recording as being noisy throughout the experiment were interpolated prior to analysis. Ocular artifacts (blinks, saccades, and smooth movements) were corrected automatically with a Principal Component Analysis method. Epochs contaminated by artifacts (amplitude  $> 150 \mu\text{V}$ , gradient  $> 75 \mu\text{V}$ , low signal  $< 0.10 \mu\text{V}$ ) were automatically rejected before averaging. EEG epochs were averaged separately across all non-artifact trials for each of the sequence types (five-tone and four-tone) and for each electrode site, and re-referenced to the average of all electrodes.

To examine ERPs related to processing the initial three or two tones of the sequences (for the five- and four-tone sequences, respectively), epochs were segmented with time 0 at the onset of the first 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).

To examine ERPs related to processing the final two tones 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). To quantify the CNV (which was expected to occur leading up to the final tone), these 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 tone 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. Mean amplitudes were averaged across a small set of 6 or 9 (depending on the particular ERP component) electrode sites for each participant and submitted to mixed-measures ANOVAs with Greenhouse-Geisser corrections when appropriate, with sequence (five-tone vs. four-tone) as the within-subjects factor and group (SLI, MLI, and WLI) as the between-subjects factor.

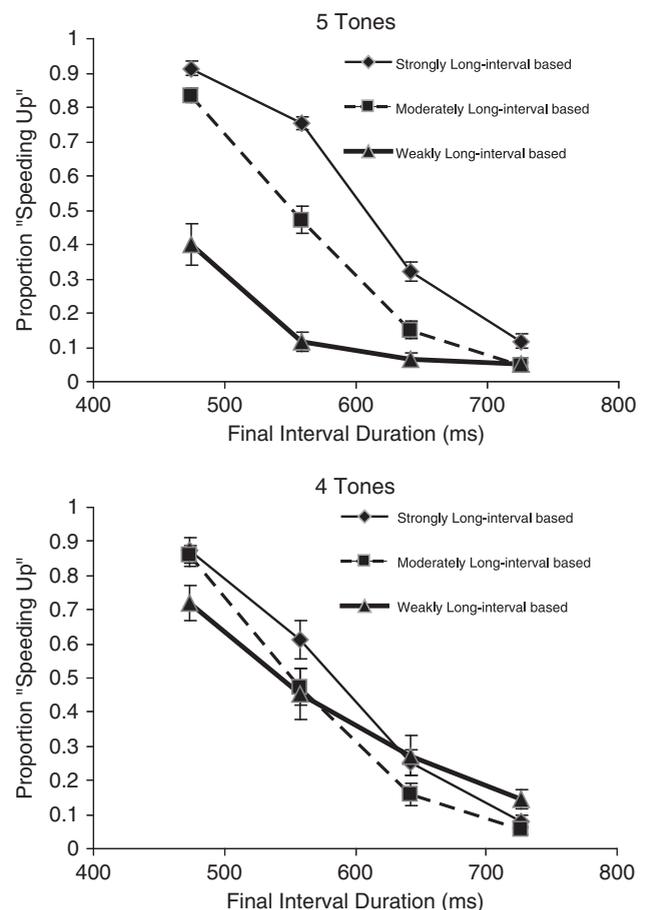
## Results

### Behavioral Data

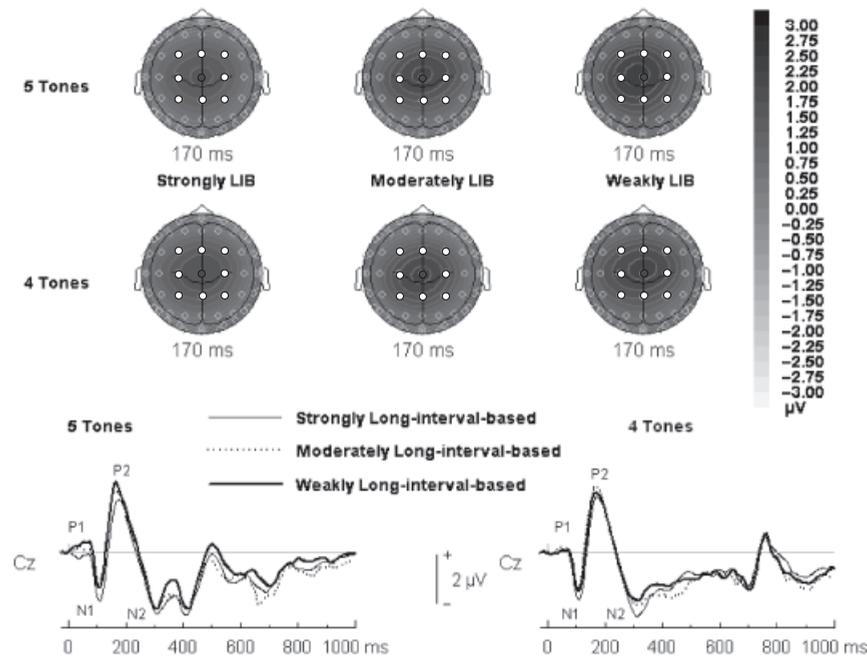
Estimates of  $w$  showed a range of values across participants but an overall tendency for most participants to perceive the five-tone sequences in a long-interval-based manner. Despite the overall tendency for long-interval-based responding, the three groups

had statistically different  $w$  values,  $F(2,35) = 24.11$ ,  $p < .001$  (SLI =  $0.99$ , MLI =  $0.92$ , WLI =  $0.55$ ). In order to assess overall temporal sensitivity regardless of any strategy differences, discrimination thresholds were measured for each participant for the four-tone (control) sequences. Discrimination thresholds correlated negatively with  $w$ ,  $r(36) = -0.675$ ,  $p < .001$ , indicating that shorter-interval listeners had poorer temporal discrimination than longer-interval listeners in the four-tone condition, even though no differences were expected depending on strategy in this condition. Because differences in discrimination thresholds for four-tone (control) sequences represents a possible confound in assessing effects of listening strategy on behavioral and brain responses to the five-tone (test) sequences, analyses reported below were run with and without the discrimination threshold included as a covariate. Tests that became non-significant when the covariate was included are indicated below. Otherwise, results are reported from analyses without the covariate.

Figure 2 shows the proportion of trials in which participants perceived speeding up for the five-tone and four-tone sequences. The behavioral results are best understood by dissecting the significant three-way interaction between the factors,  $F(6,105) = 7.09$ ,  $p < .001$ . This occurred because the stronger long-interval-based listening groups showed a steeper decline in



**Figure 2.** Behavioral results. Participants were divided into strongly long-interval-based (SLI), moderately long-interval-based (MLI), or weakly long-interval-based (WLI) according to an index of long-interval-based listening. The three groups showed a similar pattern of behavior in the four-tone condition, but distinct patterns of behavior in the five-tone condition.



**Figure 3.** Sensory-evoked ERPs for initial tones. Responses are averaged across participants in the three groups for the five-tone and four-tone sequences. (A) Scalp topographical patterns of voltage at the P2 peak, showing maxima at central midline electrodes in all groups (electrodes labeled as filled circles were used for ERP amplitude measurement; filled black circle indicates electrode shown in panel B). (B) ERP traces showing responses at Cz, with P1, N1, P2, and N2 responses to Tone 1 labeled.

perceiving speeding up as a function of final IOI, especially for five-tone sequences. When the data from the five-tone and four-tone sequences were analyzed separately, significant interactions occurred between group and final IOI for five-tone sequences,  $F(6,105) = 28.58$ ,  $p < .001$ , and four-tone sequences,  $F(6,105) = 4.65$ ,  $p < .005$ , but the interaction for four-tone sequences became non-significant when the covariate was added,  $F(6,102) = 1.77$ ,  $p = .15$ . This result simply indicates that the difference in slope of the psychometric curve for the four-tone condition reflects a difference in temporal discrimination ability.

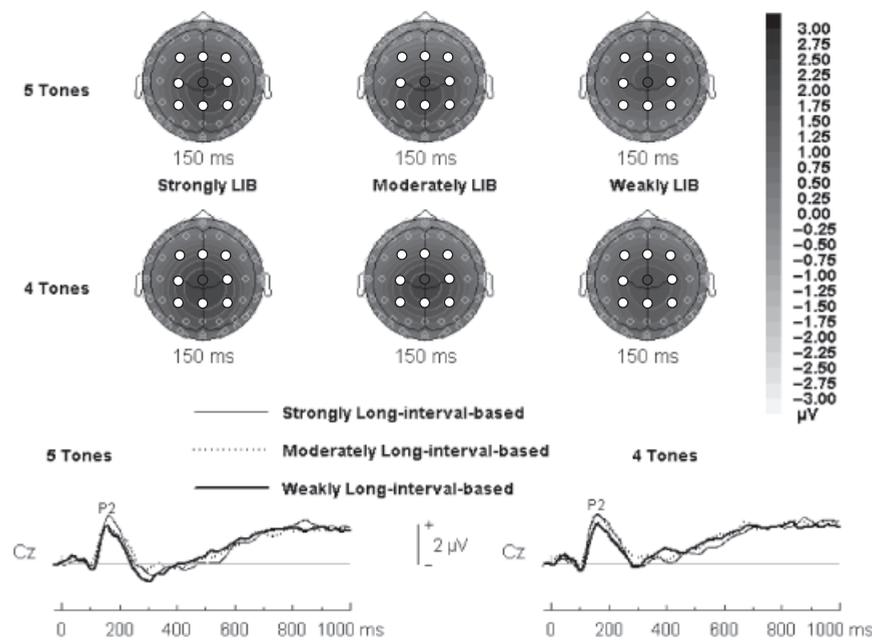
The results for the SLI (and to a lesser extent the MLI) listeners in the five-tone condition can be explained if these listeners were more likely to judge tempo using the implied 600-ms beat and consistently expected events to occur every 600 ms, causing them to perceive speeding up when the final IOI was shorter than 600 ms and slowing down when the final IOI was longer than 600 ms. The results for the WLI listeners in the five-tone condition can be explained if these participants tended to perceive the mean of the first two IOIs to be 300 ms and perceived the final IOI to be longer than 300 ms. In summary, even when controlling for differences in sensitivity to temporal changes for the four-tone control pattern, the two groups of participants showed reliable differences in perception of the five-tone pattern. Specifically, the SLI and MLI listeners were much more affected by the final IOI in the five-tone sequence than the WLI listeners; SLI and MLI listeners appeared to base tempo judgments for five-tone sequences on the implied 600-ms beat, whereas WLI listeners did so to a much less extent.

#### Electrophysiological Data

One possible difference between groups in neural processing of the rhythms is the initial sensory encoding of the stimuli. Figure 3 shows grand-averaged ERPs of the three groups overlaid on top

of each other for the five-tone and four-tone sequences. Clear long-latency N1, P2, and N2 responses (measured at electrodes FCz/1/2, Cz/1/2, and CPz/1/2 from 95–120, 145–200, and 290–330 ms, respectively) at central midline electrodes occurred in response to the first tone of both sequences. No significant differences were observed as a function of group for any of these responses,  $F(2,35) = 0.65, 0.29, 0.53$  and,  $p = .53, .75, .60$  for the N1, P2, and N2, respectively. Additional smaller responses occurred to subsequent tones in the two stimulus sequences but no group differences were apparent, and lack of reliable measurement of these responses precluded quantitative analysis. As shown in Figure 4, the P2 (measured at electrodes FCz/1/2, Cz/1/2, and CPz/1/2 from 140–200) to the final tone of the sequence showed similar amplitude as a function of group,  $F(2,35) = 0.58$ ,  $p = .57$ , as with the sensory evoked responses to the initial tone of the sequences. These negative results suggest that group behavioral differences are not likely to be related to modulation of sensory encoding.

Another possible difference between groups is how they compute time intervals and form expectancies for the final tone in the sequences. As shown in Figure 5, a CNV at central midline electrodes appeared between the onsets of the penultimate and final tones of the sequence in all groups of listeners (measured at electrodes FCz/1/2, Cz/1/2, and CPz/1/2 from  $-400$ – $0$  ms relative to the final tone). The CNV we observed likely reflected expectancy for the final tone evoked by the penultimate tone and the preceding initial tones of the sequence, as a result of explicit computation of temporal intervals (Pfeuty et al., 2003; Walter et al., 1964). There was a trend in the data for a larger CNV in listeners more likely to judge tempo using a long temporal referent, but the main effect of group was not significant,  $F(2,35) = 0.86$ ,  $p = .43$ . However, there was a significant group  $\times$  sequence interaction,  $F(2,35) = 5.25$ ,  $p < .025$ , due to larger CNV with more long-



**Figure 4.** Sensory-evoked ERPs for final tone. (A) Scalp topographical patterns of voltage at the P2 peak, showing maxima at central midline electrodes in all groups (electrodes labeled as filled circles were used for ERP amplitude measurement; filled black circle indicates electrode shown in panel B). (B) ERP traces showing responses at Cz, with P2 response labeled.

interval-based listening for the five-tone condition, but less of an effect of group in the four-tone condition.

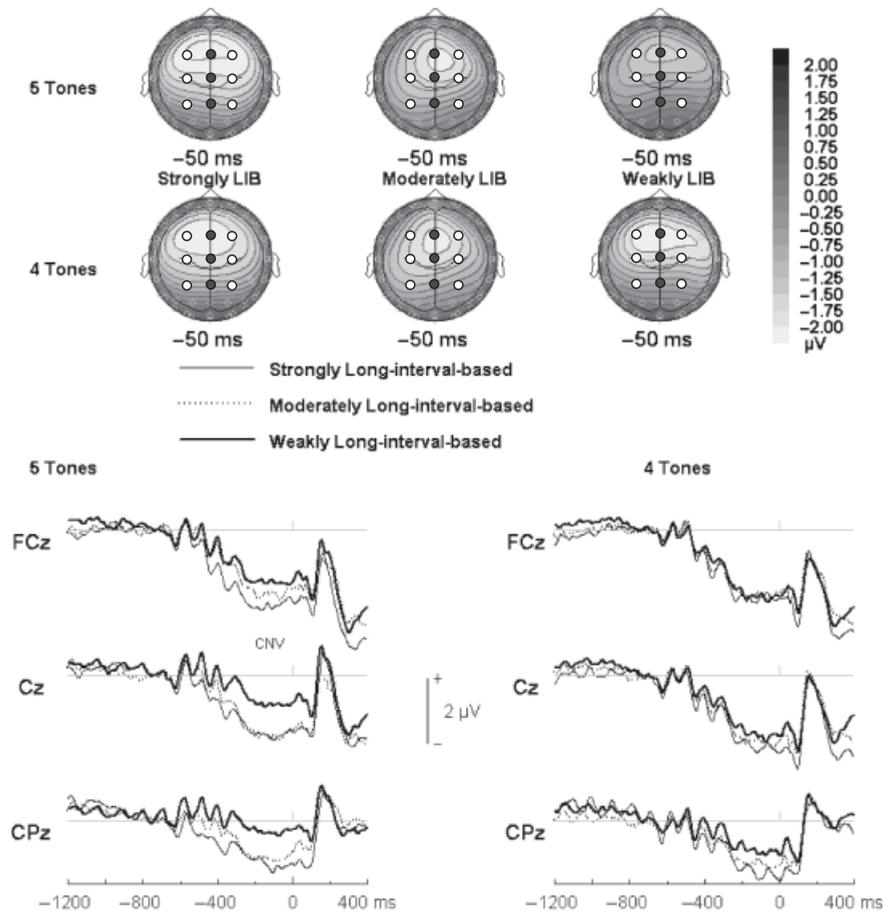
A final possible difference between listeners is in brain activity related to detecting the deviant in the final interval, prior to responding speeding up or slowing down. Such activity would be predicted to occur after the final tone in the sequences. In response to the final tone of both sequences, Figure 6 shows the presence of a late positive component occurring in all three listener groups (Donchin & Coles, 1988; Polich, 2007; Sutton, Braren, Zubin, & John, 1965). Prominent differences between the groups (measured at electrodes Pz/1/2 and POz/3/4 from 270–430 ms),  $F(2,35) = 7.78$ ,  $p < .005$ , occurred at parietal electrodes, consistent with a P3b response (Polich, 2007), with more positive activity as a function of stronger long-interval-based listening in both the five-tone and four-tone sequences. The extended temporal course of the positive wave suggests that other components besides P3b might also be activated. No group  $\times$  sequence interaction was present. Thus, the P3b reflects differences in the strength of long-interval-based listening that are present for both the four-tone and five-tone sequence conditions, unlike the CNV differences that were only present in the five-tone condition. Finally, the topographic distribution of scalp voltage for the P3b response showed larger amplitude in right-hemisphere electrodes, as indicated by a significant main-effect of hemisphere (measured at electrodes P1/2 and PO3/4 from 270–430 ms),  $F(1,35) = 13.73$ ,  $p < .001$ . However, this right-hemisphere bias was similar for the three groups, as indicated by a non-significant group  $\times$  hemisphere interaction term.

## Discussion

Differences between participants that can be attributed to rhythm processing strategy were reflected in brain activity related to timing and expectancy for the final tone of the

five-tone sequence that was designed to elicit strategy-related differences in behavioral judgments. The CNV has been clearly linked with temporal processing because it is elicited by a stimulus that temporally predicts a later stimulus (Besson, Faita, Czternasty, & Kutas, 1997; Macar et al., 1999; Martin, Houck, Kicic, & Tesche, 2008; Pfeuty et al., 2003; Pouthas, Garnero, Ferrandez, & Renault, 2000; Walter et al., 1964). Furthermore, the CNV correlates with the well-known behavioral observation that presenting multiple time intervals in succession enhances the precision of temporal judgments (Pfeuty et al., 2003). The finding that CNV differences between groups were only observed for the five-tone condition suggests that it indexes the active engagement of a particular strategy during rhythmic pattern processing. Whether the larger CNV in more long-interval-based participants is due to the engagement of stronger beat-based processing or simply due to greater attention to the implied 600-ms temporal referent per se is difficult to resolve from the current data. Although previous researchers have suggested that the CNV is likely to index the buildup of a time estimation process during interval-based processing (Martin et al., 2008; Pfeuty et al., 2003), it is equally likely that the CNV could reflect the activation of an oscillatory mechanism in beat-based processing. Besides the CNV, other EEG-based measures of temporal expectancy are long-latency ERPs (Hughes, Darcey, Barkan, Williamson, Roberts, & Asline, 2001; Janata, 2001; Raij, McEvoy, Makela, & Hari, 1997; Simson, Vaughan, & Ritter, 1976; Weinberg, Walter, Cooper, & Aldridge, 1973) and high-frequency activity (Fujioka, Large, Trainor, & Ross, 2009; Iversen, Repp, & Patel, 2009; Snyder & Large, 2005; Zanto, Large, Fuchs, & Kelso, 2005; also see Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008) evoked by missing expected events, which future studies may also show to distinguish between different listening strategies.

The current study took an individual-differences approach to distinguishing activity related to timing strategy, specifically by dividing participants into groups depending on a behavioral



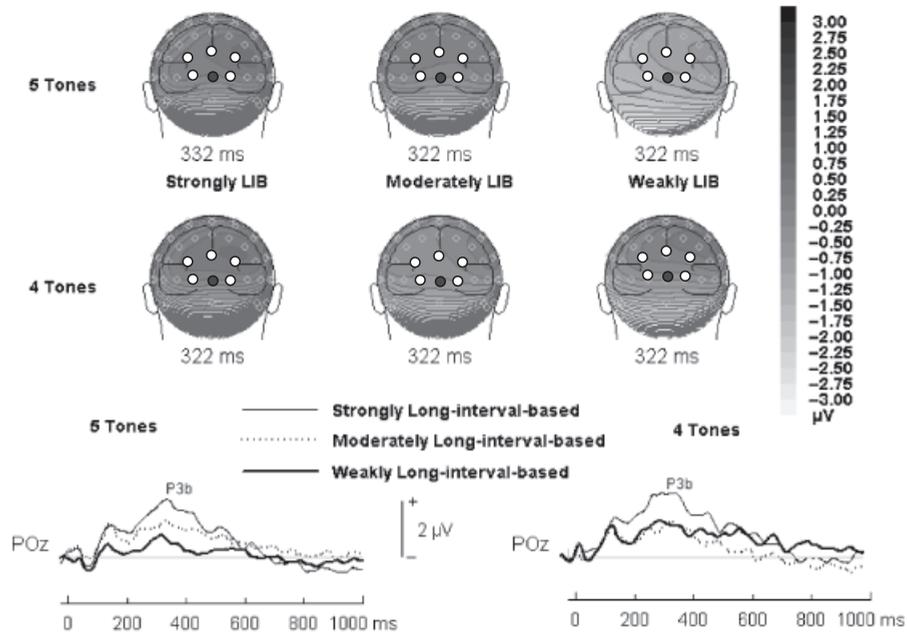
**Figure 5.** Contingent negative variation (CNV) following penultimate tone. (A) Scalp topographical patterns of voltage at  $-50$  ms, prior to the final tone, showing maxima at frontocentral midline electrodes in all groups (electrodes labeled as filled circles were used for ERP amplitude measurement; filled black circles indicate electrodes shown in panel B). (B) ERP traces showing responses at FCz, Cz, and CPz electrodes, with CNV labeled. There is a larger CNV in stronger long-interval-based listening groups, mainly due to the difference in the five-tone sequence.

measure of sensitivity to an implied 600-ms temporal referent. However, another approach to distinguish between different processes in the brain related to timing is to look for physiological markers of these processes in different brain areas. For example, a recent study used source modeling of magnetoencephalography (MEG) data to identify activity that correlated with the positive effect of a prior warning stimulus during a visual timing task (Martin et al., 2008). The authors interpreted observed CNV-like slow-wave activity as reflecting the buildup of an interval-based mechanism (Treisman, 1963), and stimulus-related phase of activity as reflecting a phase-correcting beat-based mechanism (Large & Jones, 1999; McAuley, 1995; McAuley & Jones, 2003). Sources in the parietal lobe and cerebellum exhibited slow-wave activity similar to a CNV that correlated with behavioral performance enhancement, while the cerebellum and somatomotor cortex exhibited stimulus-related phase that also correlated with behavior. These results suggest not only that beat- and interval-based timing strategies are both observable in different individuals, but that both types of mechanism may be operating in parallel in the same individuals.

In the current study, strategy-related differences were observed in late positive brain activity (i.e., the P3b response) following the final tone of the two sequences. This was likely

indicative of a difference in cognitive processing of the final interval. A previous fMRI study using the same paradigm as the current study found brain activation differences in participants with high vs. low  $w$  values (Grahn & McAuley, 2009). Consistent with the current P3b differences, these differences occurred while participants listened to both the five- and four-tone sequences, providing strong converging evidence that some differences in neural activity related to timing strategy reflect general processing differences, even when no behavioral differences are apparent. The P3b difference between groups is also consistent with the previous fMRI data showing group differences in frontal and superior temporal brain regions, major generators of the P3b (Halgren, Marinkovic, & Chauvel, 1998).

The group difference in P3b occurred for both five-tone and four-tone sequences and was present even when controlling for discrimination threshold in the four-tone condition (which is not predicted to differ depending on listening strategy). These results suggest that P3b differences do not simply reflect overall differences in temporal sensitivity or differences in temporal sensitivity specifically for the five-tone sequence. It is also unlikely that the difference in P3b occurred because weakly long-interval-based listeners were simply experiencing much larger deviations from the expected final interval than strongly long-interval-based



**Figure 6.** Late positive component (P3b) following final tone. (A) Scalp topographical patterns of voltage at P3b peak, showing maxima at parietal electrodes (electrodes labeled as filled circles were used for ERP amplitude measurement; filled black circle indicates electrode shown in panel B). (B) ERP traces showing responses at POz electrodes, with P3b labeled. There is a larger P3b in stronger long-interval-based listening groups for both five- and four-tone sequences.

participants because this would actually predict larger P3b amplitude in weakly long-interval-based participants and only in the five-tone condition, which was not observed. Finally, it is unlikely that individual differences occurred because the final interval was so far from an expected 300-ms interval that weakly long-interval-based listeners did not need to pay attention, because this would also predict group differences only for the five-tone condition.

The fact that the P3b showed reliable differences between groups suggests the importance of this component in processing temporally-structured patterns, consistent with previous ERP studies. Following intervals that were shorter or longer than the previous intervals in isochronous sequences or following a break of a pattern of intervals, late positive responses occurred (Ford & Hillyard, 1981; Nordby, Roth, & Pfefferbaum, 1988a, 1988b) that were larger in participants who performed better at detecting temporal change (Jongsma, Meeuwissen, Vos, & Maes, 2007). Studies also showed that late positive components during rhythm tasks were larger in participants with extensive training in musical rhythms (Jongsma et al., 2004), although one study found no difference between musicians and non-musicians during detection of late notes in familiar and unfamiliar melodies (Besson et al., 1994). In addition to indicating detection of temporal irregularities and individual differences in rhythm perception, late positive components index the illusory perception of alternating accents in non-accented isochronous sequences (Brochard et al., 2003), and is enhanced when a temporal interval is accurately cued by a warning stimulus (Miniussi, Wilding, Coull, & Nobre, 1999). Late positive components also have been observed in response to deviants of non-rhythmic aspects of musical structure such as melody, harmony, and lyrics (Besson, Faita, Peretz, Bonnel, & Requin, 1998; Janata, 1995; Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Trainor, McDonald, & Alain, 2002), suggesting a general role in processing sequential patterns.

It is perhaps surprising that the sensory-evoked P1-N1-P2 responses did not differ between listeners, especially for the first group of tones in the five-tone sequence because it was predicted that listeners sensitive to the 600-ms interval might attend more to the first and third tones compared to the second tone, which would result in enhanced activity (Picton & Hillyard, 1974). This negative result is all the more surprising given that sensory-evoked responses are sensitive to temporal structure, showing larger responses for larger time intervals (Carver, Fuchs, Jantzen, & Kelso, 2002; Hari, Kaila, Katila, Tuomisto, & Varpula, 1982; Lu, Williamson, & Kaufman, 1992; Mayville et al., 2001; Snyder & Large, 2004), sensitivity to the grouping structure of simple acoustic sequences (Loveless, Levanen, Jousmaki, Sams, & Hari, 1996; Loveless & Hari, 1993; Skrandies & Rammsayer, 1995), and modulation by early or late events during sensory-motor synchronization (Praamstra, Turgeon, Hesse, Wing, & Perryer, 2003; Tecchio, Salustri, Thaut, Pasqualetti, & Rossini, 2000). It is possible that the lack of individual differences in sensory-evoked responses is related to the fact that even participants in the weakly long-interval-based group had relatively high  $w$  scores. Thus, future ERP studies could screen participants in order to have listeners with maximally different rhythm processing strategies.

The behavioral data from our study and the recent fMRI study using the same paradigm (Grahn & McAuley, 2009) suggest that there is no clear-cut distinction between listener groups. Rather, listeners appear to vary continuously in listening strategy, as revealed by the range of  $w$  scores. The current ERP data further suggest a lack of clear distinction between groups because similar neural responses were observed regardless of listening strategy, with only quantitative differences in amplitude for the elicited responses. The ERP data showed highly similar time-courses of brain activity at all stages of processing, with similar-sized sensory-evoked responses to the tones and CNV and P3b responses that showed strategy-related processing of the final

time interval. Importantly, no qualitatively different pattern of brain responses was observed as a function of listening strategy. Thus, although the behavioral data do suggest that listeners can

have distinct patterns of responding, the ERP data suggest that such a pattern of behavioral data is generated by modulating the amount of activity in similar brain processes.

## REFERENCES

- Besson, M., Faita, F., Czernasty, C., & Kutas, M. (1997). What's in a pause: Event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology*, *46*, 3–23.
- Besson, M., Faita, F., Peretz, I., Bonnel, A. M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, *9*, 494–498.
- Besson, M., Faita, F., & Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neuroscience Letters*, *168*, 101–105.
- Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The “ticktock” of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science*, *14*, 362–366.
- Carver, F. W., Fuchs, A., Jantzen, K. J., & Kelso, J. A. S. (2002). Spatiotemporal analysis of the neuromagnetic response to rhythmic auditory stimulation: Rate dependence and transient to steady-state transition. *Clinical Neurophysiology*, *113*, 1921–1931.
- Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating. *Behavioral and Brain Sciences*, *11*, 357–374.
- Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception*, *18*, 1–23.
- Ford, J. M., & Hillyard, S. A. (1981). Event-related potentials (ERPs) to interruptions of a steady rhythm. *Psychophysiology*, *18*, 322–330.
- Fujioka, T., Large, E. W., Trainor, L. J., & Ross, B. (2009). Beta and gamma rhythms in human auditory cortex during musical beat processing. *Annals of the New York Academy of Sciences*, *1169*, 89–92.
- Grahn, J. A., & McAuley, J. D. (2009). Neural bases of individual differences in beat perception. *NeuroImage*, *47*, 1894–1903.
- Hannon, E. E., Snyder, J. S., Eerola, T., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 956–974.
- Halgren, E., Marinkovic, K., & Chauvel, P. (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalography and Clinical Neurophysiology*, *106*, 156–164.
- Hari, R., Kaila, K., Katila, T., Tuomisto, T., & Varpula, T. (1982). Interstimulus-interval dependence of the auditory vertex response and its magnetic counterpart: Implications for their neural generation. *Electroencephalography and Clinical Neurophysiology*, *54*, 561–569.
- Hughes, H. C., Darcey, T. M., Barkan, H. I., Williamson, P. D., Roberts, D. W., & Aslin, C. H. (2001). Responses of human auditory association cortex to the omission of an expected acoustic event. *NeuroImage*, *13*, 1073–1089.
- Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, *12*, 273–280.
- Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences*, *1169*, 58–73.
- Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*, *7*, 153–164.
- Janata, P. (2001). Brain electrical activity evoked by mental formation of auditory expectations and images. *Brain Topography*, *13*, 169–193.
- Jongsma, M. L., Desain, P., & Honing, H. (2004). Rhythmic context influences the auditory evoked potentials of musicians and non-musicians. *Biological Psychology*, *66*, 129–152.
- Jongsma, M. L., 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.
- Keele, S. W., Nicoletti, R., Ivry, R. B., & Pokorný, R. A. (1989). Mechanisms of perceptual timing: Beat-based or interval-based judgments? *Psychological Research*, *50*, 251–256.
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*, 110–113.
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*, 119–159.
- Large, E. W., & Snyder, J. S. (2009). Pulse and meter as neural resonance. *Annals of the New York Academy of Sciences*, *1169*, 46–57.
- Loveless, N., Levanen, S., Jousmaki, V., Sams, M., & Hari, R. (1996). Temporal integration in auditory sensory memory: Neuromagnetic evidence. *Electroencephalography and Clinical Neurophysiology*, *100*, 220–228.
- Loveless, N. E., & Hari, R. (1993). Auditory evoked fields covary with perceptual grouping. *Biological Psychology*, *35*, 1–15.
- Lu, Z. L., Williamson, S. J., & Kaufman, L. (1992). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science*, *258*, 1668–1670.
- Macar, F., Vidal, F., & Casini, L. (1999). The supplementary motor area in motor and sensory timing: Evidence from slow brain potential changes. *Experimental Brain Research*, *125*, 271–280.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. New York: Cambridge University Press.
- Martin, T., Egly, R., Houck, J. M., Bish, J. P., Barrera, B. D., Lee, D. C., et al. (2005). Chronometric evidence for entrained attention. *Perception and Psychophysics*, *67*, 168–184.
- Martin, T., Houck, J. M., Kicic, D., & Tesche, C. D. (2008). Interval timers and coupled oscillators both mediate the effect of temporally structured cueing. *NeuroImage*, *40*, 1798–1806.
- Mayville, J. M., Fuchs, A., Ding, M. Z., Cheyne, D., Deecke, L., & Kelso, J. A. S. (2001). Event-related changes in neuromagnetic activity associated with syncopation and synchronization timing tasks. *Human Brain Mapping*, *14*, 65–80.
- McAuley, J. D. (1995). *Perception of time as phase: toward an adaptive-oscillator model of rhythmic pattern processing*. Unpublished doctoral dissertation, Indiana University, Bloomington.
- McAuley, J. D., Frater, D., Janke, K., & Miller, N. S. (2006). Detecting changes in timing: Evidence for two modes of listening. *The Proceedings of the 9th International Conference on Music Perception and Cognition*, 188–189.
- 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. *J Exp Psychol Hum Percept Perform*, *29*, 1102–1125.
- Miniussi, C., Wilding, E. L., Coull, J. T., & Nobre, A. C. (1999). Orienting attention in time. Modulation of brain potentials. *Brain*, *122*, 1507–1518.
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 756–766.
- Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988a). Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. *Psychophysiology*, *25*, 262–268.
- Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988b). Event-related potentials to time-deviant and pitch-deviant tones. *Psychophysiology*, *25*, 249–261.
- 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., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*, 717–733.
- Pfeuty, M., Ragot, R., & Pouthas, V. (2003). Processes involved in tempo perception: A CNV analysis. *Psychophysiology*, *40*, 69–76.
- Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials 2: 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.

- Pouthas, V., Garnero, L., Ferrandez, A. M., & Renault, B. (2000). ERPs and PET analysis of time perception: Spatial and temporal brain mapping during visual discrimination tasks. *Human Brain Mapping, 10*, 49–60.
- Povel, D. J., & Essens, P. J. (1985). Perception of temporal patterns. *Music Perception, 2*, 411–440.
- Praamstra, P., Turgeon, A., Hesse, C. W., Wing, A. M., & Perryer, L. (2003). Neurophysiological correlates of error correction in sensorimotor-synchronization. *NeuroImage, 20*, 1283–1297.
- Raij, T., McEvoy, L., Makela, J. P., & Hari, R. (1997). Human auditory cortex is activated by omissions of auditory stimuli. *Brain Research, 745*, 134–143.
- Schulze, H. H. (1978). The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research, 40*, 173–181.
- Simson, R., Vaughan, H. G., & Ritter, W. (1976). Scalp topography of potentials associated with missing visual or auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 40*, 33–42.
- Skrandies, W., & Rammsayer, T. (1995). The perception of temporal structure and auditory evoked brain activity. *Biological Psychology, 40*, 267–280.
- Snyder, J., & Krumhansl, C. L. (2001). Tapping to ragtime: Cues to pulse finding. *Music Perception, 18*, 455–489.
- Snyder, J. S., & Large, E. W. (2004). Tempo dependence of middle- and long-latency auditory responses: Power and phase modulation of the EEG at multiple time-scales. *Clinical Neurophysiology, 115*, 1885–1895.
- Snyder, J. S., & Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research, 24*, 117–126.
- Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science, 150*, 1187–1188.
- Tecchio, F., Salustri, C., Thaut, M. H., Pasqualetti, P., & Rossini, P. M. (2000). Conscious and preconscious adaptation to rhythmic auditory stimuli: A magnetoencephalographic study of human brain responses. *Experimental Brain Research, 135*, 222–230.
- Trainor, L. J., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience, 14*, 430–442.
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the “internal clock.” *Psychological Monographs, 77*, 1–31.
- Toiviainen, P., & Snyder, J. S. (2003). Tapping to Bach: Resonance-based modeling of pulse. *Music Perception, 21*, 43–80.
- van Noorden, L. P. A. S., & Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research, 28*, 43–66.
- Walter, W. G., Winter, A. L., Cooper, R., McCallum, W. C., & Aldridge, V. J. (1964). Contingent negative variation: Electric sign of sensorimotor association and expectancy in the human brain. *Nature, 203*, 380–384.
- Weinberg, H., Walter, W. G., Cooper, R., & Aldridge, V. J. (1973). Emitted cerebral events. *Electroencephalography and Clinical Neurophysiology, 34*, 752–752.
- Zanto, T. P., Large, E. W., Fuchs, A., & Kelso, J. A. S. (2005). Gamma-band responses to perturbed auditory sequences: Evidence for synchronization of perceptual processes. *Music Perception, 22*, 531–547.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*, 547–558.

(RECEIVED January 12, 2010; ACCEPTED February 22, 2010)