

A linear oscillator model predicts dynamic temporal attention and pupillary entrainment to rhythmic patterns

Lauren K. Fink
University of California, Davis
Davis, CA, USA

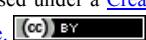
Brian K. Hurley
University of California, Davis

Joy J. Geng
University of California, Davis

Petr Janata
University of California, Davis

Rhythm is a ubiquitous feature of music that induces specific neural modes of processing. In this paper, we assess the potential of a stimulus-driven linear oscillator model (Tomic & Janata, 2008) to predict dynamic attention to complex musical rhythms on an instant-by-instant basis. We use perceptual thresholds and pupillometry as attentional indices against which to test our model predictions. During a deviance detection task, participants listened to continuously looping, multi-instrument, rhythmic patterns, while being eye-tracked. Their task was to respond anytime they heard an increase in intensity (dB SPL). An adaptive thresholding algorithm adjusted deviant intensity at multiple probed temporal locations throughout each rhythmic stimulus. The oscillator model predicted participants' perceptual thresholds for detecting deviants at probed locations, with a low temporal salience prediction corresponding to a high perceptual threshold and vice versa. A pupil dilation response was observed for all deviants. Notably, the pupil dilated even when participants did not report hearing a deviant. Maximum pupil size and resonator model output were significant predictors of whether a deviant was detected or missed on any given trial. Besides the evoked pupillary response to deviants, we also assessed the continuous pupillary signal to the rhythmic patterns. The pupil exhibited entrainment at prominent periodicities present in the stimuli and followed each of the different rhythmic patterns in a unique way. Overall, these results replicate previous studies using the linear oscillator model to predict dynamic attention to complex auditory scenes and extend the utility of the model to the prediction of neurophysiological signals, in this case the pupillary time course; however, we note that the amplitude envelope of the acoustic patterns may serve as a similarly useful predictor. To our knowledge, this is the first paper to show entrainment of pupil dynamics by demonstrating a phase relationship between musical stimuli and the pupillary signal.

Keywords: Pupil, attention, entrainment, rhythm, music, modeling, amplitude envelope, psychophysics

Received April 4th, 2018; Published November 20th, 2018.
Citation: Fink, L. K., Hurley, B. K., Geng, J. J. & Janata, P. (2018).
A linear oscillator model predicts dynamic temporal attention and pupillary entrainment to rhythmic patterns. *Journal of Eye Movement Research*, 11(2):12.
Digital Object Identifier: 10.16910/jemr.11.2.12
ISSN: 1995-8692
This article is licensed under a [Creative Commons Attribution 4.0 International license](#). 

Introduction

Though diverse forms of music exist across the globe, all music shares the property of evolving through time. While certain scales, modes, meters, or timbres may be more or less prevalent depending on the culture in question, the use of time to organize sound is universal. There-

fore, rhythm, one of the most basic elements of music, provides an excellent scientific starting point to begin to question and characterize the neural mechanisms underlying music-induced changes in motor behavior and attentional state. To remain consistent with previous literature, here rhythm is defined as patterns of duration, timing, and stress in the amplitude envelope of an auditory signal (a physical property), whereas meter is a perceptual phenomenon that tends to include the pulse (beat or tactus) frequency perceived in a rhythmic sequence, as well as slower and faster integer-related frequencies (1).

Previous studies have shown that the presence of meter affects attention and motor behavior. For instance, perceptual sensitivity is enhanced and reaction times are decreased when targets occur in phase with an on-going metric periodicity (2–4). Interestingly, this facilitation via auditory regularity is observed not only for auditory targets, including speech (5), but also for visual targets (6–12). One promising theory that accounts for these results is Dynamic Attending Theory (DAT) (13,14).

Dynamic Attending Theory

Dynamic Attending Theory (DAT) posits that the neural mechanisms of attention are susceptible to entrainment by an external stimulus, allowing for temporal predictions and therefore attention and motor coordination to specific time points (13–15). For any given stimulus, the periodicities with the most energy will capture attention most strongly. Neurobiologically, the proposed mechanism is entrainment of neuronal membrane potential of, for example, neurons in primary auditory cortex (in the case of auditory entrainment), to the external stimulus. These phase-locked fluctuations in membrane potential alter the probability of firing action potentials at any given point in time (see (16) for a review).

Similarly, the recent Active Sensing Hypothesis (16–18) proposes that perception occurs actively via motor sampling routines, that neural oscillations serve to selectively enhance or suppress input, cross-modally, and that cortical entrainment is, in and of itself, a mechanism of attentional selection. Higher frequency oscillations can become nested within lower frequency ones via phase-phase coupling, phase-amplitude coupling, or amplitude-amplitude coupling, allowing for processing of different stimulus attributes in parallel (19,20). (21) have connected the ideas of Active Sensing to those of DAT by highlighting the critical role of low frequency neural oscillations. In

summary, DAT and Active Sensing are not incompatible, as outlined in (22).

Interestingly, studies of neural entrainment are typically separate from those investigating sensorimotor synchronization, defined as spontaneous synchronization of one's motor effectors with an external rhythm (23). However, a recent study confirms that the amplitude of neural entrainment at the beat frequency explains variability in sensorimotor synchronization accuracy, as well as temporal prediction capabilities (24). Although motor entrainment, typically referred to as sensorimotor synchronization, is not explicitly mentioned as a mechanism of DAT, the role of the motor system in shaping perception is discussed in many DAT papers (25–29) and is a core tenet of the Active Sensing Hypothesis (22).

In this paper, we test whether a computational model of Dynamic Attending Theory can predict attentional fluctuations to rhythmic patterns. We also attempt to bridge the gap between motor and cortical entrainment by investigating coupling of pupil dynamics to musical stimuli. We consider the pupil both a motor behavior and an overt index of attention, which we discuss in more detail below.

Sensori(oculo)motor coupling

Though most sensorimotor synchronization research has focused on large-scale motor effectors, the auditory system also seems to have a tight relationship with the ocular motor system. For instance, (30) show that eye movements can synchronize with a moving acoustic target whether it is real or imagined, in light or in darkness. With regard to rhythm, a recent paper by (31) suggests that the tempo of rhythmic auditory stimuli modulates both fixation durations and inter-saccade-intervals: rhythms with faster tempi result in shorter fixations and inter-saccade-intervals and vice versa. These results seem to fit with those observed in audiovisual illusions, which illustrate the ability of auditory stimuli to influence visual perception and even enhance visual discrimination (32,33). Such cross-modal influencing of perception also occurs when participants are asked to engage in purely imaginary situations (34).

Though most studies have focused on eyeball movements, some (outlined below) have begun to analyze the effect of auditory stimuli on pupil dilation. Such an approach holds particular promise, as changes in pupil size reflect sub-second changes in attentional state related to

locus coeruleus-mediated noradrenergic (LC-NE) functioning (35–37). The LC-NE system plays a critical role in sensory processing, attentional regulation, and memory consolidation. Its activity is time-locked to theta oscillations in hippocampal CA1, and is theorized to be capable of phase-resetting forebrain gamma band fluctuations, which are similarly implicated in a broad range of cognitive processes (38).

In the visual domain, the pupil can dynamically follow the frequency of an attended luminance flicker and index the allocation of visual attention (39), as well as the spread of attention, whether cued endogenously or exogenously (40). However, such a pupillary entrainment effect has never been studied in the auditory domain. Theoretically though, pupil dilation should be susceptible to auditory entrainment, like other autonomic responses, such as respiration, heart rate, and blood pressure, which can become entrained to slow periodicities present in music (see (41), for a recent review on autonomic entrainment).

In the context of audition, the pupil seems to be a reliable index of neuronal auditory cortex activity and behavioral sensory sensitivity. (42) simultaneously recorded neurons in auditory cortex, medial geniculate (MG), and hippocampal CA1 in conjunction with pupil size, while mice detected auditory targets embedded in noise. They found that pupil diameter was tightly related to both ripple activity in CA1 (in a 180 degree antiphase relationship) and neuronal membrane fluctuations in auditory cortex. Slow rhythmic activity and high membrane potential variability were observed in conjunction with constricted pupils, while high frequency activity and high membrane potential variability were observed with largely dilated pupils. At intermediate levels of pupil dilation, the membrane was hyperpolarized and the variance in its potential was decreased. The same inverted U relationship was observed for MG neurons as well. Crucially, in the behavioral task, (42) found that the decrease in membrane potential variance at intermediate pre-stimulus pupil sizes predicted the best performance on the task. Variability of membrane potential was smallest on detected trials (intermediate pupil size), largest on false alarm trials (large pupil size), and intermediate on miss trials (small pupil size). Though this study was performed on mice, it provides compelling neurophysiological evidence for using pupil size as an index of auditory processing.

The same inverted U relationship between pupil size and task performance has been observed in humans during

a standard auditory oddball task. For instance, (43) showed that baseline pupil diameter predicts both reaction time and P300 amplitude in an inverted U fashion on an individual trial basis. Additionally, (43) found that baseline pupil diameter is negatively correlated with the phasic pupillary response elicited by deviants. Because of the well-established relationship between tonic neuronal activity in locus coeruleus and pupil diameter, it is theorized that both the P300 amplitude and pupil diameter index locus coeruleus-norepinephrine activity (35,37,43–45).

Moving towards musical stimuli, pupil size has been found to be larger for: more arousing stimuli (46,47), well-liked stimuli (48), more familiar stimuli (47), psychologically and physically salient auditory targets (49–51), more perceptually stable auditory stimuli (52), and chill-evoking musical passages (53). A particularly relevant paper by (54) showed that the pupil responded to unattended omissions in on-going rhythmic patterns when the omissions coincided with strong metrical beats but not weak ones, suggesting that the pupil is sensitive to internally generated hierarchical models of musical meter. While Damsma and van Rijn's analysis of difference waves was informative, the continuous time series of pupil size may provide additional dynamic insights into complex auditory processing.

Of particular note, (55) demonstrated a relationship between attention and the time course of the pupillary signal while listening to music. To do this, they had participants listen to 30-sec clips of classical music while being eye-tracked. In the first phase of the experiment, participants listened to each clip individually (diotic presentation); in the second phase participants were presented with two different clips at once (dichotic presentation) and instructed to attend to one or the other. Kang & Wheatley compared the pupil signal during dichotic presentation to the pupil signal during diotic presentation of the attended vs. ignored clip. Using dynamic time warping to determine the similarity between the pupillary signals of interest, they showed that in the dichotic condition, the pupil signal was more similar to the pupil signal recorded during diotic presentation of the attended clip than to that recorded during diotic presentation of the unattended clip (55). Such a finding implies that the pupil time series is a time-locked, continuous dependent measure that can reveal fine-grained information about an attended auditory stimulus. However, it remains to be determined whether it is possible for the pupil to become entrained to rhythmic auditory stimuli and whether such oscillations would reflect attentional

processes or merely passive entrainment.

Predicting dynamic auditory attention

Because the metric structure perceived by listeners is not readily derivable from the acoustic signal, a variety of algorithms have been developed to predict at what period listeners will perceive the beat. For example, most music software applications use algorithms to display tempo to users and a variety of contests exist in the music information retrieval community for developing the most accurate estimation of perceived tempo, as well as individual beats, e.g. the Music Information Retrieval Evaluation eX-change (MIREX) Audio Beat Tracking task (56). The beat period, however, is just one aspect of the musical meter. More sophisticated algorithms and models have been developed to predict all prominent metric periodicities in a stimulus, as well as the way in which attention might fluctuate as a function of the temporal structure of an audio stimulus, as predicted by Dynamic Attending Theory.

For instance, the Beyond-the-Beat (BTB) model (57) parses audio in a way analogous to the auditory nerve and uses a bank of 99 damped linear oscillators (reson filters) tuned to frequencies between 0.25 and 10 Hz to model the periodicities present in a stimulus. Several studies have shown that temporal regularities present in behavioral movement data (tapping and motion capture) collected from participants listening to musical stimuli correspond to the modeled BTB periodicity predictions for those same stimuli (57–59). Recent work (60) suggests that an additional model calculation of time-varying temporal salience can predict participants' perceptual thresholds for detecting intensity changes at a variety of probed time points throughout the modeled stimuli, i.e. participants' time-varying fluctuations in attention when listening to rhythmic patterns.

In the current study, we further tested the BTB model's temporal salience predictions by asking whether output from the model could predict the pupillary response to rhythmic musical patterns. We hypothesized that the model could predict neurophysiological signals, such as the pupillary response, which we use as a proxy for attention. Specifically, we expected that the pupil would become entrained to the rhythmic musical patterns in a stimulus specific way.

We also expected to see phasic pupil dilation responses to intensity deviants. As in (60), we used an adaptive thresholding procedure to probe participants' perceptual

thresholds for detecting intensity increases (dB SPL) inserted at multiple time points throughout realistic, multi-part rhythmic stimuli. Each probed position within the stimulus had a corresponding value in terms of the model's temporal salience predictions. We hypothesized that detection thresholds should be lower at moments of high model-predicted salience and vice versa. If perceptual thresholds differ for different moments in time, we assume this reflects fluctuations in attention, as predicted by DAT.

Methods

Participants

Eighteen people participated in the experiment (13 female; mean age: 26 years (min: 19; max: 52; median 23 years)). Student participants from UC Davis received course credit for participation; other volunteer participants received no compensation. The experimental protocol was approved by the Institutional Review Board at UC Davis.

Materials

The five rhythmic patterns used in this study (Fig. 1, left column, top panels) were initially created by Dr. Peter Keller via a custom audio sequencer in Max/MSP 4.5.7 (Cycling '74), for a previous experiment in our lab. Multitimbre percussive patterns, each consisting of the snap, shaker, and conga samples from a Proteus 2000 sound module (E-mu Systems, Scotts Valley, CA) were designed to be played back in a continuous loop at 107 beats per minute, with a 4/4 meter in mind. However, we remain agnostic as the actual beat periodicity and metric periodicities listeners perceived in the stimuli, as we leave such predictions to the linear oscillator model. Each stimulus pattern lasted 2.2 s. We use the same stimulus names as in (60) for consistency. All stimuli can be accessed in the supplemental material of (60).

Please note that the intensity level changed dynamically throughout the experiment based on participants' responses. The real-time, adaptive presentation of the stimuli is discussed further in the *Adaptive Thresholding Procedure* section below.

Linear oscillator model predictions

All stimuli were processed through the Beyond-the-Beat model (57) to obtain mean periodicity profiles and

temporal salience predictions. For full details about the architecture of the model and the periodicity surface calculations, see (57). For details about the temporal salience calculations, please see (60).

In short, the model uses the Institute for Psychoacoustics and Electronic Music toolbox (61) to transform the incoming audio in a manner analogous to the auditory nerve, separating the signal into 40 different frequency bands, with center frequencies ranging from 141 to 8877 Hz. Then, onset detection is performed in each band by taking the half-wave rectified first order difference of the root mean square (RMS) amplitude. Adjacent bands are averaged together to reduce redundancy and enhance computational efficiency. The signal from each of the remaining five bands is fed through a bank of 99 reson filters (linear oscillators) tuned to a range of frequencies up to 10 Hz. The oscillators driven most strongly by the incoming signal oscillate with the largest amplitude (Fig. 2A). A windowed RMS on the reson filter outputs results in five periodicity surfaces (one for each of the five bands), which show the energy output at each reson-filter periodicity (Fig. 2B). The periodicity surfaces are averaged together to produce an Average Periodicity Surface (Fig. 2C). The profile plotted to the right of each stimulus (Fig. 1, right column; Fig. 2D) is termed the Mean Periodicity Profile (MPP) and represents the energy at each periodicity frequency, averaged over time. Periodicities in the MPP that exceed 5% of the MPP's amplitude range are considered peak periodicities and are plotted as dark black lines against the gray profile (Figure 1, right column).

After determining the peak periodicities for each stimulus, we return to the output in each of the five bands from the reson filters. We mask this output to only contain activity from the peak frequencies. Taking the point-wise mean resonator amplitude across the peak-frequency reson filters in all five bands yields the time series shown directly beneath each stimulus pattern in Figure 1 (also see Fig. 2E). We consider this output an estimate of salience over time.

In deciding the possible time points at which to probe perceptual thresholds, we tried to sample across the range of model-predicted salience values for each stimulus by choosing four temporal locations (dotted lines in lower panels of Figure 1). We treat the model predictions as a continuous variable.

To predict the temporal and spectral properties of the

pupillary signal, we 1) extend the temporal salience prediction for multiple loop iterations 2) convolve the extended temporal salience prediction with a canonical pupillary response function (62) and 3) calculate the spectrum of this extended, convolved prediction. The pupillary response function (PRF) is plotted in Figure 2F. Its parameters have been empirically derived, first by (63) then refined by (62). The PRF is an Erlang gamma function, with the equation:

$$h = t^n e^{(-nt/t_{max})}$$

where h is the impulse response of the pupil, with latency t_{max} . (63) derived n as 10.1 which represents the number of neural signaling steps between attentional pulse and pupillary response. They derived t_{max} as 930ms when participants responded to suprathreshold auditory tones with a button press. More recently, (62) estimated t_{max} to suprathreshold auditory tones in the absence of a button press to be 512ms. They show that this non-motor PRF is more accurate in correctly deconvolving precipitating attentional events and that it can be used even when there are occasional motor responses involved, e.g. responses to deviants, as long as they are balanced across conditions. Hence, in our case, we model the continuous pupillary response to our stimuli using the non-motor PRF and simply treat any motor responses to deviants as noise that is balanced across all of our conditions (stimuli).

Though previous studies have taken a deconvolution approach (deconvolving the recorded pupil data to get an estimate of the attentional pulses that elicited it), note that we here take a forward, convolutional approach. This allows us to generate predicted pupil data (Fig. 2G) which we compare to our recorded pupil data. With this approach, we avoid the issue of not being certain when exactly the attentional pulse occurred, i.e. with deconvolution it is unclear what the relationships are between the stimulus, attentional pulse, and the system's delay (see discussion in (63) p. 24); also note that deconvolution approaches often require an additional temporal alignment technique such as an optimization algorithm, e.g. (64), and/or dynamic time-warping, e.g. (55). Here, we take the empirically derived delay, t , of the pupil to return to baseline from (62), Figure 1a as 1300ms.

Alternative models

An important consideration is whether the complexity of the linear oscillator model is necessary to accurately

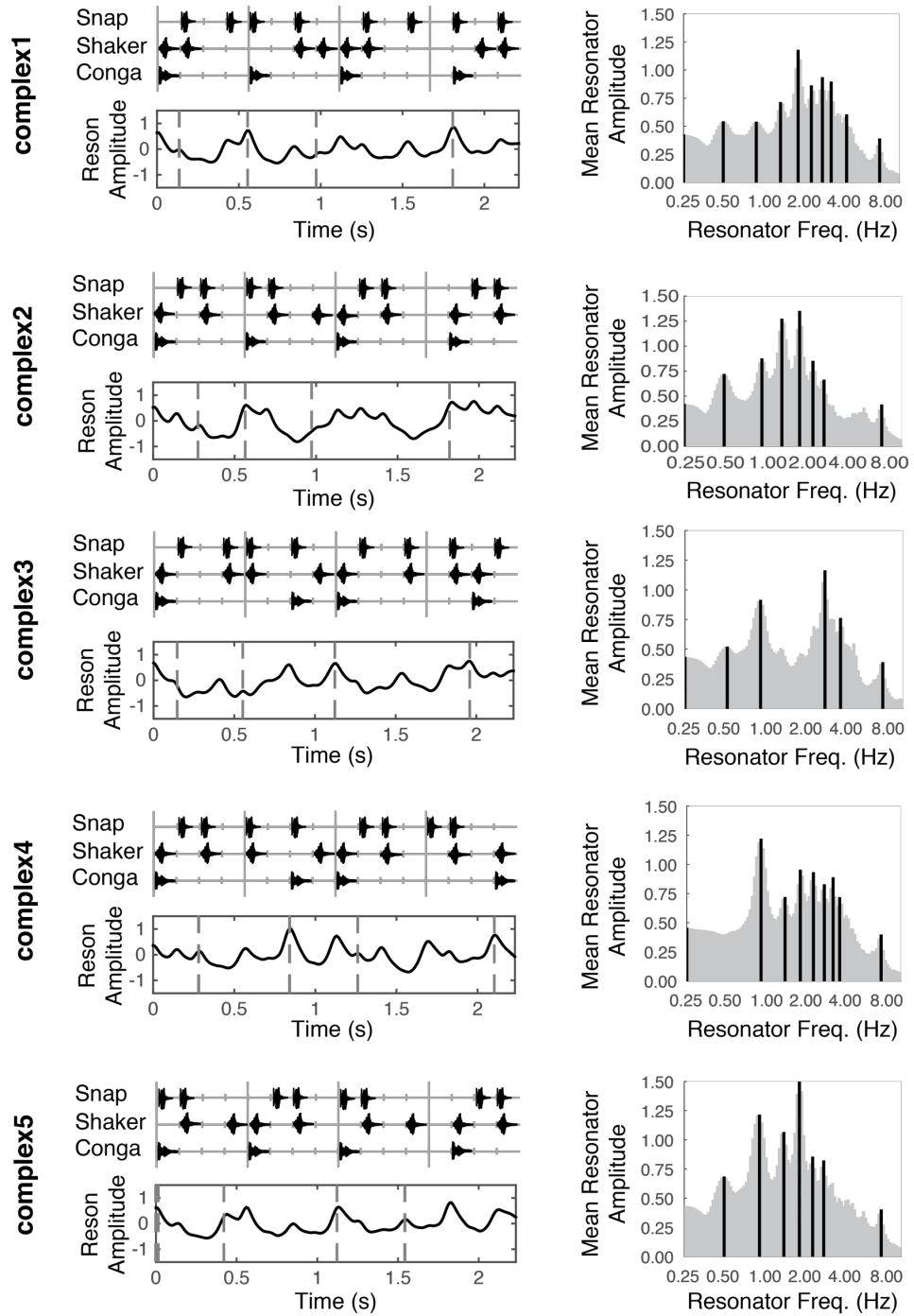


Figure 1. Stimulus patterns (left column, top panels), temporal salience predictions (left column, bottom panels), and mean periodicity profiles (right column) for each of the five stimuli used in this experiment. Vertical tick marks underlying the stimulus patterns (upper panels, left column) correspond to time in 140 ms intervals. Dotted vertical lines (bottom panels, left column), indicate moments in time that were probed with a deviant. These moments correspond to the musical event onset(s) directly above in the top panels. Dark vertical lines in the right column indicate peak periodicities in the mean periodicity profile.

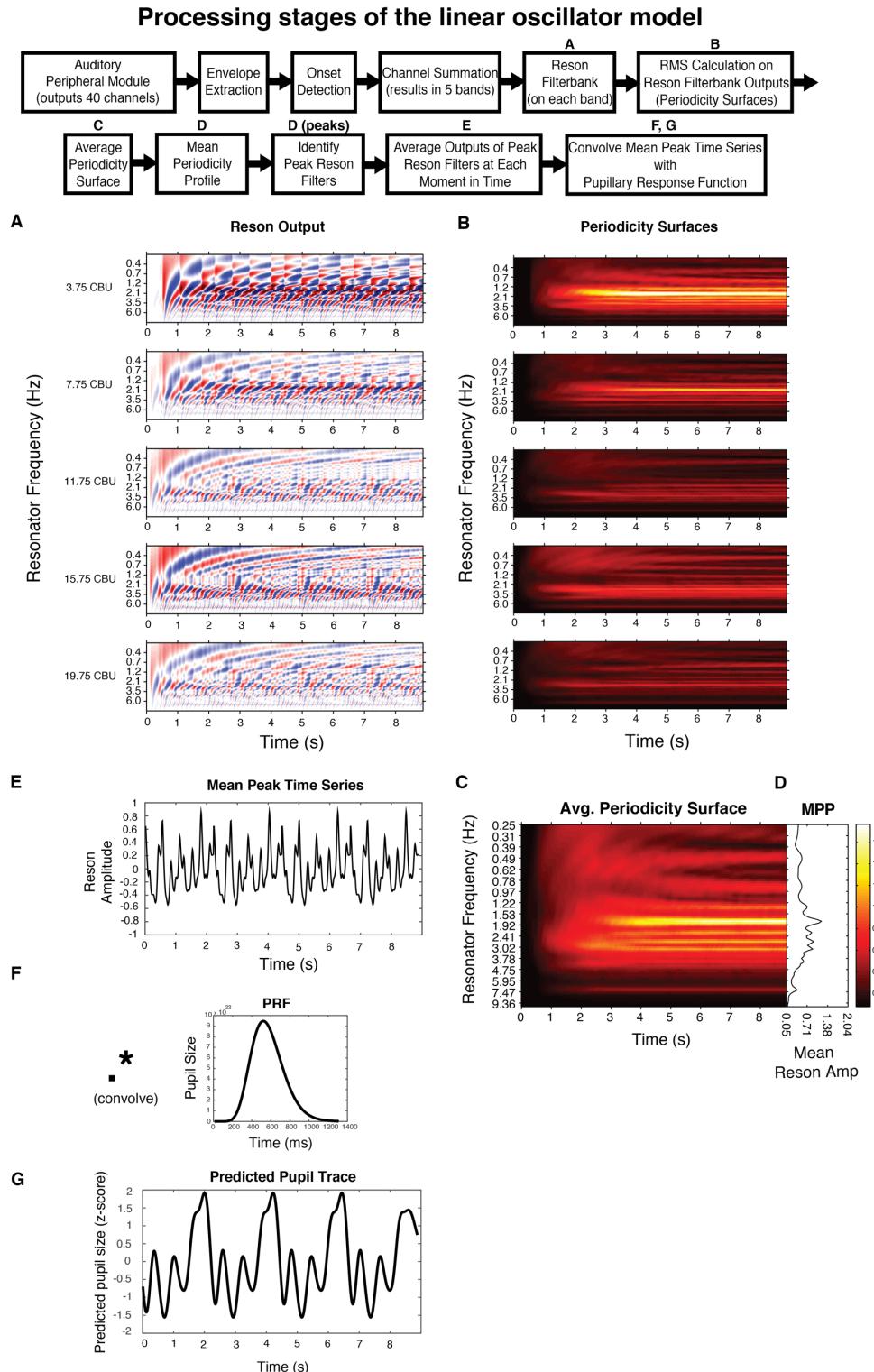


Figure 2. Boxes represent processing stages of the linear oscillator model; those labeled with a letter have a corresponding plot below. Please see *Linear oscillator model predictions* for full details. All plots are for stimulus *complex1*.

predict behavioral and pupillary data for rhythmic stimuli. To address this question, two alternative models are considered in our analyses, each representing different, relevant aspects of the acoustic input sequence.

Full resonator output: Rather than masking the resonator output at the peak periodicities determined by the Mean Periodicity Profile to get an estimate of salience over time that is driven by the likely relevant metric frequencies, it is possible to just average the output from all reson filters over time. Such a prediction acts as a nice alternative to our filtered output and allows for a comparison of whether the prominent metric periodicities play a role in predicting attention over time.

Amplitude Envelope: The spectrum of the amplitude envelope of a sound signal has been shown to predict neural entrainment frequencies, e.g. (65) show cortical steady-state evoked potentials at peak frequencies in the envelope spectrum. Hence, as a comparison to the linear oscillator model predictions, we also used the amplitude envelope of our stimuli as a predictor. To extract the amplitude envelope of our stimuli, we repeated each stimulus for multiple loops and calculated the root mean square envelope using MATLAB's *envelope* function with the 'rms' flag and a sliding window of 50ms. Proceeding with just the upper half of the envelope, we low-pass filtered the signal at 50 Hz using a 3rd order Butterworth filter then down-sampled to 100 Hz to match the resolution of the oscillator model. To predict pupil data, we convolved the envelope with the PRF, as previously detailed for the linear oscillator model.

Apparatus

Participants were tested individually in a dimly lit, sound-attenuating room, at a desk with a computer monitor, infrared eye-tracker, Logitech Z-4 speaker system, and a Dell keyboard connected to the computer via USB serial port. Participants were seated approximately 60 cm away from the monitor. Throughout the experiment, the screen was gray with a luminance of 17.7 cd/m², a black fixation cross in the center, and a refresh rate of 60 Hz. The center to edge of the fixation cross subtended 2.8° of visual angle. Pupil diameter of the right eye was recorded with an Eye-link 1000 (SR Research) sampling at 500 Hz in remote mode, using Pupil-CR tracking and the ellipse pupil tracking model. Stimuli were presented at a comfortable listening level, individually selected by each participant, through speakers that were situated on the right and left sides of the computer monitor. During the experiment, auditory stimuli were adaptively presented through Max/MSP (Cycling '74; code available at:

<https://github.com/janatalab/attmap.git>), which also recorded behavioral responses and sent event codes to the eye-tracking computer via a custom Python socket.

Procedure

Participants were instructed to listen to the music, to maintain their gaze as comfortably as possible on the central fixation cross, and to press the "spacebar" key any time they heard an increase in volume (a deviant). They were informed that some increases might be larger or smaller than others and that they should respond to any such change. A 1 min practice run was delivered under the control of our experimental web interface, Ensemble (66), after which participants were asked if they had any questions.

During the experiment proper, a run was approximately 7 min long and consisted of approximately 190 repetitions of a stimulus pattern. There were no pauses between repetitions, thus a continuously looping musical scene was created. Please note that the exact number of loop repetitions any participant heard varied according to the adaptive procedure outlined below. Each participant heard each stimulus once, resulting in five total runs of approximately 7 min each, i.e. a roughly 35 min experiment.

Stimulus order was randomized throughout the experiment. Messages to take a break and continue when ready were presented after each run of each stimulus. Following the deviance detection task, participants completed questionnaires assessing musical experience, imagery abilities, genre preferences, etc. These questionnaire data were collected as part of larger ongoing projects in our lab and will not be reported in this study. In total, the experimental session lasted approximately 50 min; this includes the auditory task, self-determined breaks between runs (which were typically between 5-30 s), and the completion of surveys.

Adaptive Thresholding Procedure: Though participants experienced a continuous musical scene, we can think of each repetition of the stimulus loop as the fundamental organizing unit of the experiment that determined the occurrence of deviants. Specifically, after every standard (no-deviant) loop iteration, there was an 80% chance of a deviant, in one of the four probed temporal locations, without replacement, on the following loop iteration. After every deviant loop, there was a 100% chance of a no-deviant loop. The Zippy Estimation by Sequential Testing (ZEST) (67,68) algorithm was used to dynamically change

the decibel level of each deviant, depending on the participant's prior responses and an estimated probability density function (p.d.f.) of their threshold.

The ZEST algorithm tracked thresholds for each of the four probed temporal locations separately during each stimulus run. A starting amplitude increase of 10 dB SPL was used as an initial difference limen. On subsequent trials, the p.d.f. for each probed location was calculated based on whether the participant detected the probe or not, within a 1000 ms window following probe onset. ZEST uses Bayes' theorem to constantly reduce the variance of a posterior probability density function by reducing uncertainty in the participant's threshold probability distribution, given the participant's preceding performance. The mean of the resultant p.d.f. determines the magnitude of the following deviant at that location. The mean of the estimated probability density function on the last deviant trial is the participant's estimated perceptual threshold.

Compared to a traditional staircase procedure, ZEST allows for relatively quick convergence on perceptual thresholds. Because the ZEST procedure aims to minimize variance, using reversals as a stopping rule, like in the case of a staircase procedure, does not make sense. Here, we used 20 observations as a stopping rule because (68) showed that 18 observations were sufficient in a similar auditory task and (60) demonstrated that, on average, 11 trials allowed for reliable estimation of perceptual threshold when using a dynamic stopping rule. For the current study, 20 was a conservative choice for estimating thresholds, which simultaneously enabled multiple observations over which to average pupillary data.

In summary, each participant was presented with a deviant at each of the four probed locations, in each stimulus, 20 times. The intensity change was always applied to the audio file for 200 ms in duration, i.e. participants heard an increase in volume of the on-going rhythmic pattern for 200 ms before the pattern returned to the initial listening volume. The dB SPL of each deviant was adjusted dynamically based on participants' prior responses. The mean of the estimated probability density function on the last deviant trial (observation 20) was the participant's estimated threshold. Examples of this adaptive stimulus presentation are accessible online as part of the Supplemental Material in (60): <http://dx.doi.org/10.1037/xhp0000563.sup>.

Analysis

Perceptual Thresholds: Participants' perceptual

thresholds for detecting deviants at each of the probed temporal locations were computed via ZEST (67,68); see the *Adaptive Thresholding Procedure* section above for further details.

Reaction Time: Reaction times were calculated for each trial for each participant, from deviant onset until button press. Trials containing reaction times that did not fall within three scaled median absolute deviations from the median were removed from subsequent analysis. This process resulted in the removal of 0.12% of the data.

Pupil Preprocessing: Blinks were identified in the pupil data using the Eyelink parser blink detection algorithm (69), which identifies blinks as periods of loss in pupil data surrounded by saccade detection, presumed to occur based on the sweep of the eyelid during the closing and opening of the eye. Saccades were also identified using Eyelink's default algorithm.

Subsequently, all ocular data were preprocessed using custom scripts and third party toolboxes in MATLAB version 9.2 (70). Samples consisting of blinks or saccades were set to NaN, as was any sample that was 20 arbitrary units greater than the preceding sample. A sliding window of 25 samples (50ms) was used around all NaN events to remove edge artifacts. Missing pupil data were imputed by linear interpolation. Runs requiring 30% or more interpolation were discarded from future analysis, which equated to 9% of the data. The pupil time series for each participant, each run (~7 min), was high-pass filtered at .05 Hz, using a 3rd order Butterworth filter, to remove any large-scale drift in the data. For each participant, each stimulus run, pupil data were normalized as follows: z-scored-PupilData = (rawData - mean(rawData)) / std(rawData). See Figure S1 in the Supplementary Materials accompanying this article for a visualization of these pupil pre-processing steps.

Collectively, the preprocessing procedures and some of the statistical analyses reported below relied on the Signal Processing Toolbox (v. 7.4), the Statistics and Machine Learning Toolbox (v. 11.1), the Bioinformatics Toolbox (v. 4.4), Ensemble (66), and the Janata Lab Music Toolbox (57,71). All custom analysis code is available upon request.

Pupil Dilation Response: The pupil dilation response (PDR) was calculated for each probed deviant location in each stimulus by time-locking the pupil data to deviant onset. A baseline period was defined as 200 ms preceding

deviant onset. The mean pupil size from the baseline period was subtracted from the trial pupil data (deviant onset through 3000 ms). The mean and max pupil size were calculated within this 3000 ms window. We chose 3000 ms because the pupil dilation response typically takes around 2500 ms to return to baseline following a motor response (62); therefore, 3000 ms seemed a safe window length. Additionally, the velocity of the change in pupil size from deviant onset to max dilation was calculated as the slope of a line fit from pupil size at deviant onset to max pupil size in the window, similar to Figure 1C in (72). The latency until pupil size maximum was defined as the duration (in ms) it took from deviant onset until the pupil reached its maximum size. Trials containing a baseline mean pupil size that was greater than three scaled median absolute deviations from the median were removed from subsequent analyses (0.2% of all trials).

Time-Frequency Analyses: To examine the spectro-temporal overlap between our varied model predictions and the observed pupillary signals, we calculated the spectrum of the average pupillary signal to 8-loop epochs of each stimulus, for each participant. We then averaged the power at each frequency across all participants, for each stimulus. We compared the continuous time series and the power spectral density for the recorded pupil signal for each stimulus to those predicted by the model predictions convolved with the pupillary response function. These two average analyses are included for illustrative purposes; note that the main analysis of interest is on the level of the single participant, single stimulus, as outlined below.

To compare the fine-grained similarity between the pupil size time series for any given stimulus to the linear oscillator model prediction for that stimulus, we computed the Cross Power Spectral Density (CPSD) between the pupil time series and itself, the model time series and itself, and the pupil time series and the model time series. The CPSD was calculated using Welch's method (73), with a 4.4 s window and 75% overlap. For each participant, each stimulus, we computed 1) the CPSD between the pupil trace and the model prediction for that stimulus and 2) the CPSD between the pupil trace and the model prediction for all other stimuli, which served as the null distribution of coherence estimates.

The phase coherence between the pupil and the model for any given stimulus was defined as the squared absolute value of the pupil-model CPSD, divided by the power spectral density functions of the CPSD of the individual

signals with themselves. We then calculated a single true and null coherence estimate for each participant, each stimulus, by finding the true vs. null coherence at each model-predicted peak frequency (see Table S2) under 3 Hz and averaging.

Results

Perceptual Thresholds

We tested each of the three alternative predictors of perceptual thresholds (peak filtered resonator output, full resonator output, and amplitude envelope) using mixed-effects models via the *nlme* package (74) in R (75). Threshold was the dependent variable and the given model's predicted value at the time of each probe was the fixed effect. Random-effect intercepts were included for each participant. We calculated effect sizes of fixed effects using Cohen's f^2 , a standardized measure of an independent variable's effect size in the context of a multivariate model (76). We calculated f^2 effect sizes following the guidelines of (77) for mixed-effects multiple regression models.

We assessed the relative goodness of model fit using Akaike's information criterion (AIC; (78)). As widely recommended (e.g. (79)), we rescaled AIC values to represent the amount of information lost if choosing an alternative model, as opposed to the preferred model, with the equation:

$$\Delta_i = \text{AIC}_i - \text{AIC}_{\min}$$

where AIC_{\min} is the model with the lowest AIC value of all models considered and AIC_i is the alternative model under consideration. Given this equation, AIC_{\min} , by definition, has a value of 0 and all other models are expressed in relation. Typically, models having a value $\Delta_i < 2$ are considered to have strong support; models with $4 < \Delta_i < 7$ have less support, and models $\Delta_i > 10$, no support. This transformation of the AIC value is important as it is free of scaling constants and sample size effects that influence the raw AIC score. In our case, because each of our models being compared has the same amount of complexity, Bayesian Information Criterion (80) differences are identical to those calculated for AIC Δ_i so we do not include them here.

As Table 1 indicates, our peak filtered model yielded

Table 1. Comparison of three alternative predictive models of perceptual threshold.

Threshold Predictor	β	SE	df	f^2	AIC	Δ_i
Peak-filtered reson	-3.30 **	.431	344	.17	2082.857	0
Full reson	-0.011**	.0023	344	.064	2125.737	42.8
Amplitude env	-133.14**	20.80	344	.12	2090.672	7.8

Note. Model estimates were obtained using linear mixed-effects models to regress fixed effects of stimulus model type on threshold; participant intercept was included as a random effect. AIC Akaike's Information Criterion (a lower value indicates a more preferred model); Cohen's f^2 for effect size. ** $p < .001$.

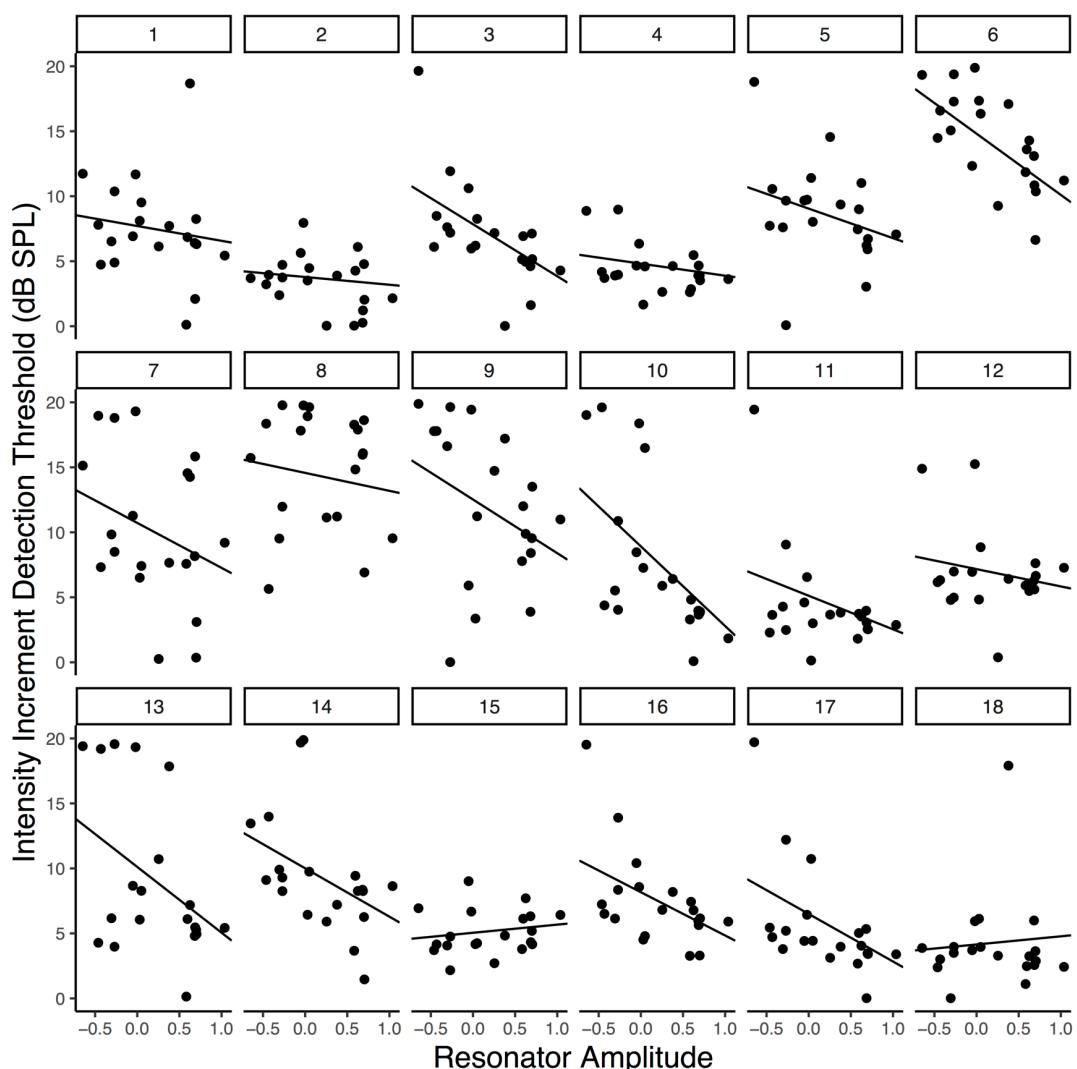


Figure 3. Increment detection thresholds as a function of averaged peak resonator level. Each panel is an individual participant's data. Lines reflect participants' slopes and intercepts as random effects in a mixed-effects model.

the lowest AIC value, suggesting that it is strongly preferred over the amplitude envelope of the audio and entirely preferred over the full resonator model output as a predictor of perceptual threshold, given this common metrics of model fit. Furthermore, although full reson output and the amplitude envelope were both significant predictors of participant thresholds, the effect size was largest for the peak-filtered reson model compared to the alternatives. However, we note that no broadly accepted significance test exists for comparing non-nested mixed-effects models (i.e., each model contains a different fixed-effect term). As such, we caution that a strong claim of model-fit superiority would require further testing. Nevertheless, these results suggest that the peak-filtered resonator model better explains variance in participants' thresholds than does the amplitude envelope of the auditory stimulus or the unfiltered reson output. We interpret this in favor of participants entraining their attention to endogenously generated metrical expectations which are represented by the peak periodicities from our model.

The negative relationship between peak resonator level and increment detection threshold is plotted in Figure 3 and visible within most participants' data. Note that random slopes were included in the final model that generated Figure 3 so that participant level fit could be visualized and because the there is growing consensus that the random effects structure of linear mixed effects models should be kept maximal (81); however, we wish to note that adding the random slope did not significantly improve the fit of the model, which is why it was not included during our model comparisons. Overall, the final peak reson filtered model had a conditional R^2 of .087 – reflecting an approximation of the variance explained by the fixed effect of peak reson output – and a marginal R^2 of .475 – reflecting an approximation of the variance explained by the overall model (fixed effect of peak reson output plus random intercepts and slopes for participants). R^2 estimates were calculated using the *MuMIn* package in R (82–84).

In the remainder of the paper, we refrain from using the alternative full reson model, as it could be confidently rejected and was weakest of the three models. We do continue to compare the peak filtered model with the amplitude envelope; however, we wish to note that a strong correlation exists between the amplitude envelope predictions and the peak-filtered predictions at each probed position ($r(20) = .90, p < .001$).

Pupil Dilation Response (PDR)

The average pupil dilation response for each trial type, for each probed position, in each stimulus, is plotted in Figure 4. Possible trial types are 1) trials in which a deviant occurred and was detected by participants (blue), 2) trials in which a deviant occurred and was not detected by participants (red), and 3) trials in which no deviant occurred (black). In all cases the data are plotted with respect to the probed time point. Please recall that the only difference between deviant and no-deviant trials is that the auditory event at the probed moment in time is increased in dB SPL, relative to the baseline volume of the standard pattern. On average, per participant, per probe position, 12 trials were used in calculating the “hit” average, and 8 trials were used in calculating the “miss” average. For a full report of the average number and standard deviation of trials used in calculating each grand average trace plotted in Figure 4, see Table S1.

As can be seen in Figure 4, the PDR to a deviant is consistent and stereotyped across all probed time points. There was a significant difference between mean pupil size on hit vs. missed trials, $t(15) = 2.14, p = .049$, and missed vs. no deviant trials, $t(15) = 4.60, p < .001$. Additional features of the PDR also varied as a function of whether the deviant was detected or missed. There was a significant difference between max pupil size on hit vs. missed trials, $t(15) = 4.48, p < .001$, and missed vs. no-deviant trials, $t(15) = 28.45, p < .001$, as well as a marginally significant difference between pupil latency until maximum on hit vs. missed trials $t(15) = -2.11, p = .052$. There was no significant difference between pupil velocity to maximum on hit vs. miss trials.

Before constructing trial-level predictive models based on specific features of the PDR, we first assessed correlations between all predictor and outcome variables. All Pearson correlation coefficients for both hit and miss trials, across all participants, are reported in Table 2. We found that baseline pupil size is negatively correlated with mean and max evoked pupil size, as well as latency to max pupil size and pupil velocity. Mean evoked pupil size – the standard metric in most cognitive studies – was strongly positively correlated with max evoked pupil size, latency to max pupil size, and pupil velocity. It is also worth noting that decibel contrast (relative to baseline volume) was negatively correlated with reson output, reflecting our perceptual threshold results, i.e. moments of low salience re-

quired higher dB contrast to be detected. There was no correlation between dB contrast and pupil dilation, adding to the literature of mixed findings on this topic. Reaction time was weakly correlated with baseline pupil size, dB contrast, and max evoked pupil size, though not mean evoked pupil size, despite the strong correlation of these two pupil variables with each other. As the coefficients indicate, these reaction time effects are very weak at best, possibly due to our use of a standard computer keyboard which may have introduced jitter in the recording of responses.

Because of the strong correlations between possible pupil metrics of interest, in subsequent analyses we used maximum evoked pupil size in our statistical models, so as not to construct models with collinear predictors. While it may be argued that baseline pupil size is a more intuitive metric to use, as it could indicate causality, we wish to point out that there were no significant differences in baseline pupil size between hit vs. missed trials $t(15) = -0.75$, $p = .463$, while there was a significant difference in max evoked pupil size between hit and missed trials, as previously indicated. Hence, though baseline pupil size is strongly correlated with mean and max evoked pupil size, it is not our predictor of interest in context of this analysis.

To test if we could predict whether a deviant was detected or not based on the evoked pupillary response, on a per-trial basis, we fit a generalized linear mixed-effects logistic regression model. The generalized linear mixed-effects model (GLMM) included max evoked pupil size and peak reson output as predictors, participant as a random intercept, and a binary *hit* (detection of the increment) or *miss* (non-detection of the increment) as the dependent variable. The GLMM was fit via maximum likelihood using the *glmer* function from the *lme4* package (85) in R. Odds ratios, z statistics, confidence intervals, and p-values are reported for all fixed effects (Table 3).

As a comparison, we ran the same model but swapped peak reson prediction for amplitude envelope prediction. We used the *pROC* package (86) to calculate the Receiver Operating Characteristic (ROC) curves. ROC curves compare the true positive prediction rate against the false positive rate. We compared the area under the ROC curves (AUC) of the peak reson model vs. the amplitude envelope model using DeLong's test for two correlated ROC curves (87). The AUC is a common metric for evaluating both the goodness of fit of a model, as well as the performance of two different models. In this case, it reflects the probability that a randomly selected 'hit' trial is correctly identified as a 'hit' rather than 'miss' (88). With a range of .5 (chance) to 1 (perfect prediction), higher AUC values indicate better model performance. The peak reson model had an AUC of .608, while the amplitude envelope model had an AUC of .606, thus there was no significant difference between the models ($Z = 1.594$, $p = .111$). Both performed significantly above chance, whether chance was defined as the standard .5 or more conservatively via shuffling of the pupil and model data (peak reson vs. shuffled chance: $Z = 3.79$, $p < .001$; amp env vs. shuffled chance: $Z = 3.53$, $p < .001$).

Given the similar performance of peak reson and amplitude envelope models, and the fact that the pupil dilation response to deviants at different moments of predicted salience is remarkably stereotyped, it is not possible to be sure whether the pupil dilation response reflects endogenous meter perception or merely a bottom-up response to the stimuli. An experiment with a wider range of stimuli, incorporating non-stationary rhythms, might be well suited to answer this question, as such stimuli would likely result in a greater difference between the amplitude envelope and the peak reson filter predictions. Nonetheless, the finding of a PDR on trials in which participants failed to report detecting a deviant has implications for future studies and is discussed in more detail below.

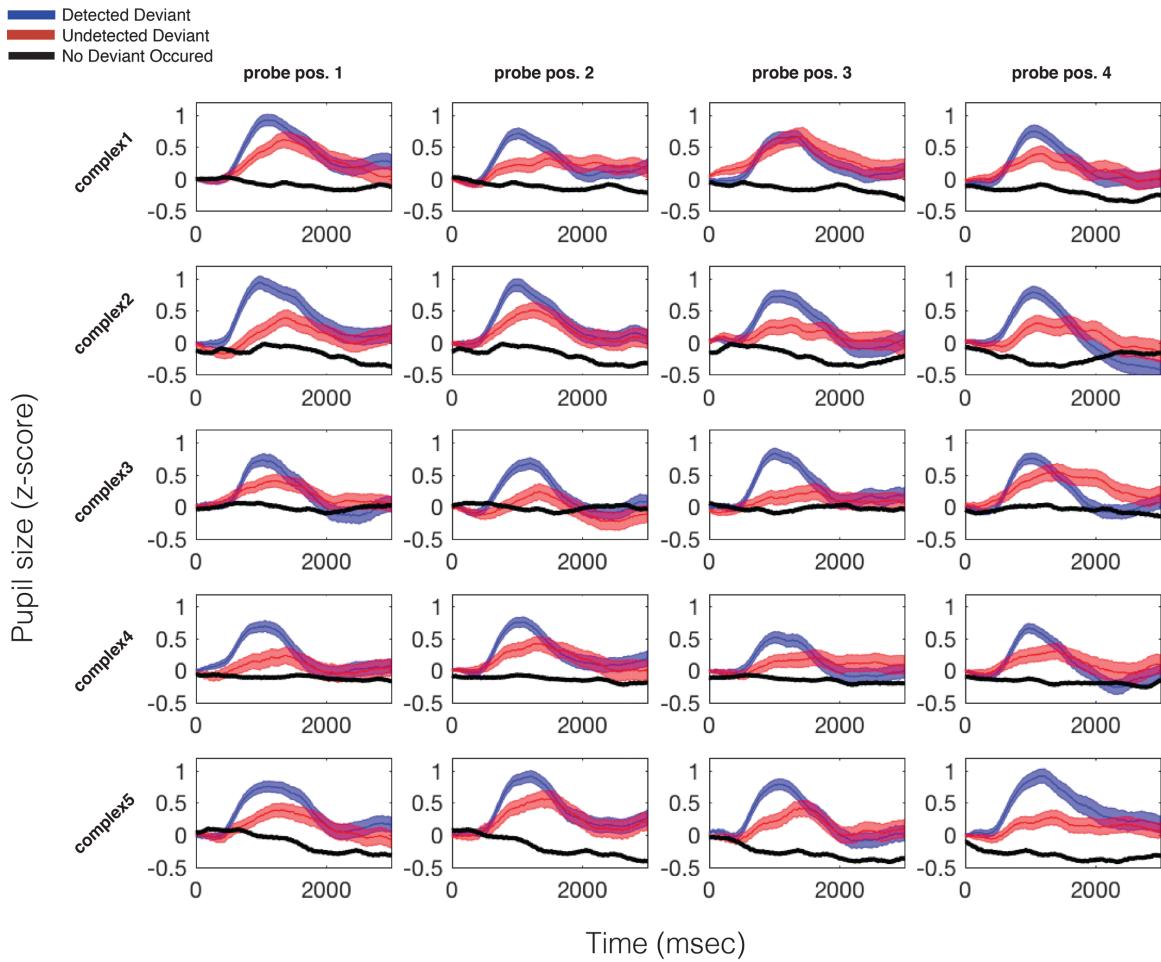


Figure 4. Average pupillary responses, across all participants, to all probed locations in all stimuli. The blue trace indicates the average pupillary response to trials during which a deviant occurred and was detected; the red trace represents trials during which a deviant occurred but was not detected; the black trace indicates the same moments in time on trials in which no deviant occurred. All data are time locked to deviant onset. For no-deviant trials, this refers to the same point in time that the deviant could have occurred (but did not). Width of each trace is the standard error of the mean.

Table 2. Pearson Correlation Coefficients for all predictor and outcome variables on ‘hit’ and ‘miss’ trials.

Variables	1 Hit (Miss)	2	3	4	5	6	7
1. Mean baseline pup	-						
2. Decibel contrast	.004 (0)	-					
3. Resonator level	.012 (0.004)	-.185**(-.338)**	-				
4. Mean evoked pup	-.811** (-.787)**	-.001 (.004)	-.022 (.01)	-			
5. Max evoked pup	-.728** (-.742)**	-.009 (.009)	-.015 (-.01)	.868** (.887**)	-		
6. Max latency	-.443** (-.492)**	.023 (-.019)	.015 (-.026)	.443** (.466**)	.357** (.417**)	-	
7. Pup velocity	-.144** (-.142)**	-.055* (.013)	.008 (.023)	.245** (.230**)	.219** (.196**)	.163** (.168**)	-
8. Reaction time	.045* (-)	.077** (-)	.006 (-)	-.01 (-)	-.039* (-)	.011 (-)	-.003 (-)

Note. ‘Miss’ trial correlation coefficients are in parenthesis; there is no reaction time for a ‘miss’ trial. Pup refers to pupil size. Decibel contrast is the change in dB, relative to baseline volume, of a deviant on any given trial. For the ‘hit’ data, df = 3108, for the ‘miss’ data, df = 2167, *p < .05, **p < .001

Table 3. Generalized linear mixed-effects logistic regression model with dependent variable hit (1) or miss (0).

	95% CI for OR			
	OR	Z stat	Lower	Upper
Max pupil size	1.28 **	8.35	1.21	1.16
Resonator output	1.26 **	3.86	1.12	1.41

Note. 5279 observations. OR = Odds Ratio. CI = Confidence Interval. **p < .001

Continuous pupil signal

While the pupil dilation response to deviants is of interest with regards to indexing change detection, we also wished to examine the continuous pupillary response to our rhythmic stimuli. Specifically, we wanted to assess whether the pupil entrained to the rhythmic patterns and, if so, whether such dynamic changes in pupil size were unique for each stimulus.

As can be seen in Figure 5, there appears to be a stimulus-specific correspondence between the model predicted pupil time series (red) and the observed continuous pupillary signal (black). Note the remarkably similar predict-

ions of the peak-filtered oscillator model (solid red) and the amplitude envelope model (dashed red). This correspondence between the two predictions and the recorded pupil data was also observable in the Power Spectral Density (PSD) estimates for each stimulus (Figure 6). Similar to studies of pupil oscillations in the visual domain (39), we do not see much power in the pupil spectrum beyond about 3 Hz; therefore, we plot frequencies in the range of 0 to 3 Hz (Figure 6). In Figure 6, it is clear that the spectra of the pupil signal and that of the two model predictions overlap. However, this analysis is not sufficient to infer that the pupil is tracking our model output in a stimulus-specific way, though it does indicate pupillary entrainment to the prominent periodicities in the stimuli.

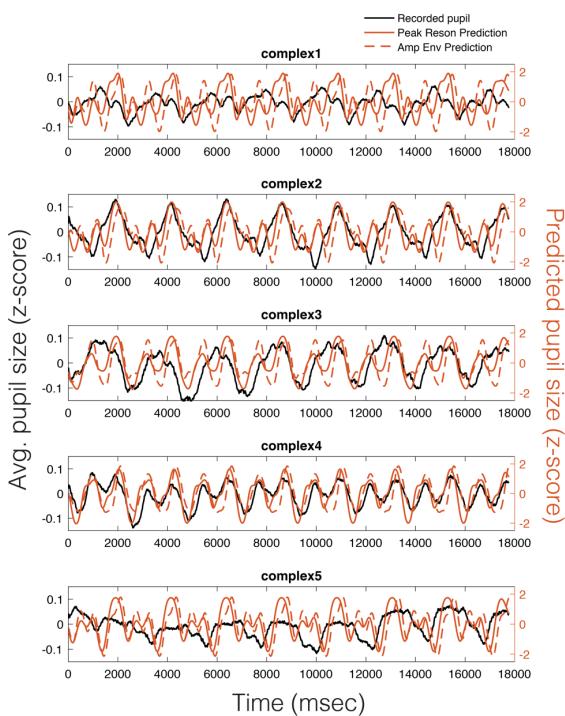


Figure 5. Average pupillary responses, across all participants, to 8-loop epochs of each stimulus (black), compared to the peak reson filter model-predicted pupillary responses (solid red) and the amplitude envelope-predicted pupillary responses (dotted red).

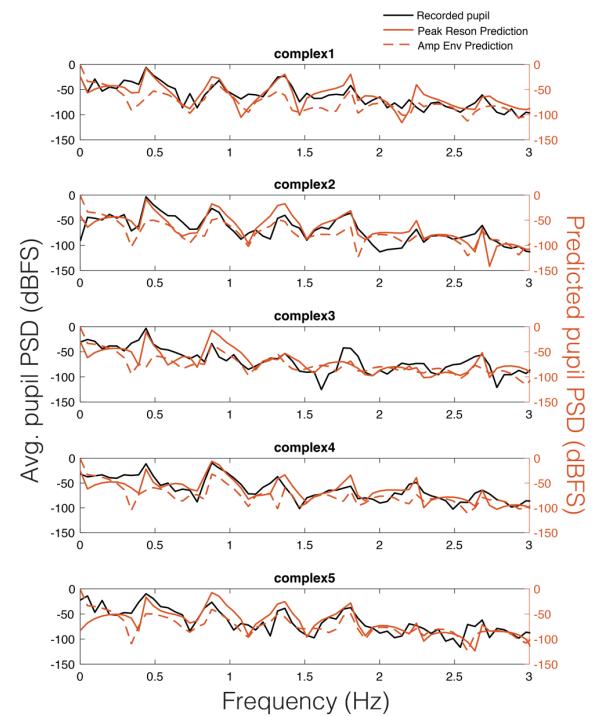


Figure 6. Average pupillary power spectral density (PSD) across participants (black) vs. peak reson filter model-predicted PSD for each stimulus (solid red) and amplitude envelope-predicted PSD (dashed red).

To examine the relationship between each participant's pupil time series and the model prediction for each stimulus, we examined their phase coherence. Because the peak reson and amplitude envelope models are very similar in temporal and spectral realms, we chose to compute phase coherence for only the peak reson model. An additional reason for doing this was because the peak reson model outputs predicted salient metric frequencies at which we can calculate coherence on a theoretical basis, whereas, with the amplitude envelope model, one would have to devise a comparable method to pick peaks in the envelope spectrum and ensure that they are of a similar number, spacing, and theoretical relevance.

Shuffling the stimulus labels allowed us to compute null coherence estimates (see *Time Frequency Analyses* for more details). The average true vs. null coherence value for each participant, across each model-predicted

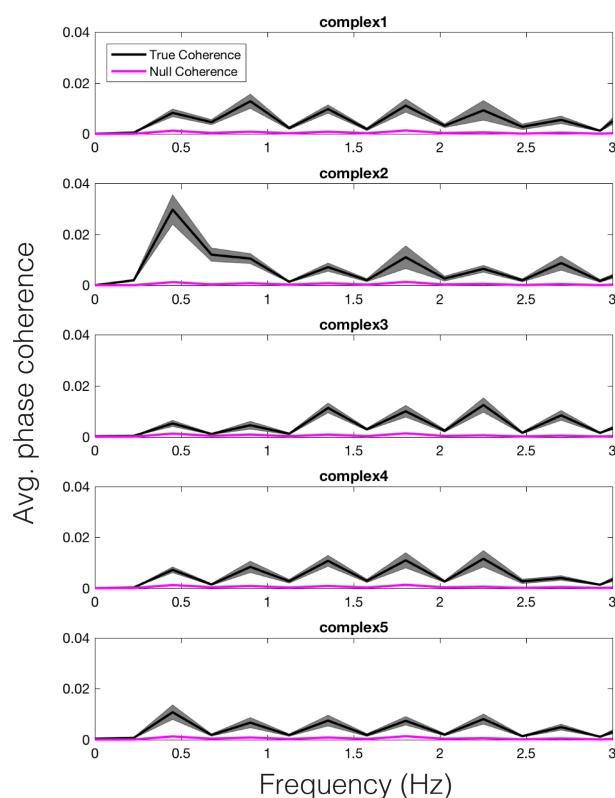


Figure 7. Average phase coherence across participants between the pupil and peak reson filter model, for each stimulus (black). Average null coherence is plotted in magenta. The width of each trace represents standard error of the mean.

peak frequency, was subjected to a paired samples t-test, revealing a significant difference in the true vs. null distributions for the peak reson-filtered model, $t(14) = 16.56, p < .001$. Thus, we can conclude that the changing dynamics of pupil size were entrained to the model predictions for each stimulus in a unique way. For illustration, we have plotted the average coherence, across participants, for each stimulus, in Figure 7 (black). The null coherence estimate is plotted in magenta. We wish to note that similar results would have likely been obtained via calculating the coherence between the amplitude envelope model and the pupil signal.

Discussion

The current experiment used a linear oscillator model to predict both perceptual detection thresholds and the pupil signal during a continuous auditory psychophysical task. During the task, participants listened to repeating percussion loops and detected momentary intensity increments. We hypothesized that the linear oscillator model would predict perceptual thresholds for detecting intensity deviants that were adaptively embedded into our stimuli, as well as the continuous pupillary response to the stimuli.

The linear oscillator model reflects the predictions of Dynamic Attending Theory (DAT), which posits that attention can become entrained by an external (quasi)-periodic stimulus. The model is driven largely by onsets detected in the acoustic envelope of the input signal, which get fed through a bank of linear oscillators (reson filters). From there it is possible to calculate which oscillators are most active, mask the output at those peak frequencies, and average over time. Throughout the paper, we considered this peak-filtered signal the ideal prediction of temporal salience, as in (60); however, for comparison, we also tested how well output from all resonators would predict our data, as well as how well the amplitude envelope alone (without any processing through the oscillator model) would do in predicting both perceptual thresholds and the pupillary signal.

The peak-filtered model was best at predicting perceptual thresholds, providing an important replication and extension of our previous study (60). In the present study we used only complex stimuli and intensity increments but our previous study showed the same predictive effects of the

peak-filtered model for both intensity increments and decrements, as well as simple and complex stimuli (60). We assume that such results imply that the peaks extracted by our model are attentionally relevant periodicities that guide listeners' attention throughout complex auditory scenes. The fact that perceptual thresholds were higher at moments of low predicted salience, i.e. a deviant needed to be louder at that moment in time for participants to hear it and vice versa, indicates that attention is not evenly distributed throughout time, in line with the predictions of DAT. However, we note that the linear oscillator model's temporal salience prediction was strongly correlated with the magnitude of the amplitude envelope of the signal.

Indeed, when it comes to the pupil signal, both the peak-filtered model and the amplitude envelope performed almost identically. This similarity is likely a result of a variety of factors: 1) the rhythms used in the current study are stationary (unchanging over time) and, though there are some moments lacking acoustic energy, overall, the prominent periodicities are present in the acoustic signal. Hence, the Fourier Transform of the amplitude envelope yields roughly identical peak periodicities to that of the model. 2) Convolving both signals with the pupillary response function smears out most subtle differences between the two signals, making them even more similar.

Regardless of the ambiguity regarding which model may be a better predictor, the pupillary results reported in this paper are exciting nonetheless. First and foremost, we show that the pupil can entrain to a rhythmic auditory stimulus. To our knowledge, we are the first to report such a finding, though others have reported pupillary entrainment in the visual domain (39,40). The continuous pupillary signal and the pupil spectrum to each stimulus were both well predicted by the linear oscillator model and the amplitude envelope of the audio signal. That pupil dilation/constriction dynamics, controlled by the smooth dilator and sphincter muscles of the iris, respectively, entrain to auditory stimuli is in line with a large literature on music-evoked entrainment. Though the pupil has never been mentioned in this literature, other areas of the autonomic nervous system have been shown to entrain to music (41). It remains to be tested how pupillary oscillations might relate to cortical neural oscillations, as highlighted in the introduction of this paper. Are pupillary delta oscillations phase-locked to cortical delta? Do cortical steady-state evoked potentials overlap with those of the pupil? Pupillometry is a more mobile and cost-effective method than

EEG, as such, characterizing the relationship between pupillary and cortical responses to music will hopefully allow future studies to use pupillometry in situations that otherwise might have required EEG.

Furthermore, we have shown not only that the pupil entrains to the prominent periodicities present in our stimuli, but also that the oscillatory pupillary response to each stimulus is unique. These results extend those of (55) and speak to the effectiveness of using pupillometry in the context of music cognition studies. Unlike (55), we did not use deconvolution or dynamic time-warping to assess the fit of our pupil data with our stimuli, rather, we took a forward approach to modeling our stimuli, convolving stimulus-specific predictions with a pupillary response function, effectively removing the need for algorithms like dynamic time-warping or fitting optimizations. We hope that this approach will prove beneficial for others, especially given the simplicity in calculating the amplitude envelope of a signal and convolving it with the pupillary response function. With regards to our linear oscillator model, future work will use a wider and more temporally dynamic variety of stimuli to assess the power of our linear oscillator model vs. the amplitude envelope in predicting the time-varying pupil signal across a diverse range of musical cases. Hopefully such a study will shed more light on the issue of whether pupillary oscillations are an evoked response or a reflection of attention and in what contexts one might need to use a more complex model, if any.

Even in the case of oscillations being driven more so by the stimuli than by endogenous attention, we still feel that such oscillations nevertheless shape subsequent input and reflect, in some way, the likely attended features of the auditory, and perhaps visual, input. Because pupil dilation blurs the retinal image and widens the visual receptive field, while pupil constriction sharpens the retinal image, narrowing the visual receptive field (40), oscillations in pupil size, which are driven by auditory stimuli may also have ramifications for visual attention (89) and audiovisual integration. For example, visual attention should be more spatially spread (greater sensitivity) at moments of greater auditory temporal salience (larger pupil dilation). It is possible, however, that such small changes in pupil size elicited by music are negligible with respect to visual sensitivity and/or acuity (see (90) for a discussion). In short, such interactions remain to be empirically tested.

Another important finding of the current study is the

pupil dilation response (PDR) to deviants. Of particular interest is the result that the pupil responds to deviants even when participants do not report hearing them, providing further evidence that pupillary responses are indicators of preconscious processing (91). However, the current results raise an additional important question of whether the PDR might be more akin to the mismatch negativity (MMN) than the P3, which requires conscious attention to be elicited (92). Others have shown a PDR to deviants in the absence of attention (51,54) and here we show a PDR to deviants that did not reach participants' decision thresholds, or possibly conscious awareness, despite their focused attention. Hence, though there is evidence connecting the P3a to the PDR and the LC-NE system (43,93), an important avenue of future research will be to disentangle the relationship between the PDR, P3a, and MMN, which can occur without the involvement of conscious attention (94) or perception (95). While all three of these measures can be used as indices of deviance detection, the P3a has been proposed to reflect comparison of sensory input with a previously formed mental expectation that is distributed across sensory and motor regions (96,97), whereas the MMN and PDR have been interpreted as more sensory-driven, pre-attentive comparison processes.

Though we are enthusiastic about the PDR results, a few considerations remain. Since the present experiment only utilized intensity increments as deviants, it could be argued that the pupil dilation response observed on trials during which a deviant was presented below perceptual threshold does not reflect subthreshold processes but rather a linear response to stimulus amplitude. To this argument, we point to the fact that there was no correlation between the mean or max evoked pupil size on any given trial and the contrast in dB SPL on that trial. However, to further address this possible alternative interpretation, we conducted an experiment involving both intensity increments and decrements. Those data show the same PDR patterns for both increments and decrements, hit vs. missed trials, as reported in the current study (98). In addition, previous studies of rhythmic violations showed a standard PDR to the omission of an event (e.g. (54)), suggesting that the PDR is not specific to intensity increment deviance.

An additional critique might be that the difference between the pupil size on hit vs. missed trials is because hit trials require a button press while miss trials do not. Though it may be the case that the additional motor response required to report deviance detection results in a

larger pupil size, this is unlikely to fully account for the difference in results (52,53). Critically, even if a button press results in a greater pupil dilation, this does not change the fact that a PDR is observed on trials in which a deviant was presented but not reported as heard (uncontaminated by a button press).

In summary, our study contributes to a growing literature emphasizing the benefits of using eye-tracking in musical contexts (for further examples, please see the other articles in this Special Issue on Music & Eye-Tracking). We have shown that the pupil of the eye can reliably index deviance detection, as well as rhythmic entrainment. Considered in conjunction with previous studies from our lab, the linear oscillator model utilized in this paper is a valuable predictor on multiple scales – tapping and large body movements (57–59), perceptual thresholds (60), and pupil dynamics (current study). In general, the model can explain aspects of motor entrainment within a range of .25 – 10 Hz – the typical range of human motor movements. Future work should further compare the strengths and limitations of models of rhythmic attending (e.g. (27,99)), the added benefits of such models over simpler predictors such as the amplitude envelope, and the musical contexts in which one model is more effective than another.

Ethics and Conflict of Interest

The authors declare that the contents of the article are in agreement with the ethics described in <http://biblio.unibe.ch/portale/elibrary/BOP/jemr/ethics.html> and that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

This research was supported in part by LF's Neuroscience Graduate Group fellowship, ARCS Foundation scholarship, and Ling-Lie Chau Student Award for Brain Research, as well as a Templeton Advanced Research Program grant from the Metanexus Institute to PJ.

We wish to thank Dr. Jeff Rector for providing the communication layer between the MAX/MSP software and the eye-tracking computer, and Dr. Peter Keller for constructing the rhythmic patterns used in this study.

References

1. London J. Hearing in Time: Psychological Aspects of Musical Meter. New York, NY: Oxford University Press; 2012.
2. Bergeson TR, Trehub SE. Infants Perception of Rhythmic Patterns. *Music Percept An Interdiscip J.* 2006;23(4):345–60.
3. Yee W, Holleran S, Jones MR. Sensitivity to event timing in regular and irregular sequences: Influences of musical skill. *Percept Psychophys.* 1994;56(4):461–71.
4. Jones M, Johnston H, Puente J. Effects of auditory pattern structure on anticipatory and reactive attending. *Cogn Psychol.* 2006;53(1):59–96.
5. Cason N, Schön D. Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia.* 2012;50(11):2652–2658.
6. Bolger D, Trost W, Schön D. Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychol.* 2013;142(2):238–44.
7. Bolger D, Coull JT, Schön D. Metrical rhythm implicitly orients attention in time as indexed by improved target detection and left inferior parietal activation. *J Cogn Neurosci.* 2014;26(3):593–605.
8. Escoffier N, Sheng DY, Schirmer A. Unattended musical beats enhance visual processing. *Acta Psychol.* 2010;135(1):12–6.
9. Grahn JA. See what I hear? Beat perception in auditory and visual rhythms. *Exp Brain Res.* 2012;220(1):51–61.
10. Grahn JA, Henry MJ, McAuley JD. fMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. *Neuroimage.* 2011;54(2):1231–43.
11. Hove MJ, Fairhurst MT, Kotz SA, Keller PE. Synchronizing with auditory and visual rhythms: an fMRI assessment of modality differences and modality appropriateness. *Neuroimage.* 2013;67:313–21.
12. Miller JE, Carlson LA, McAuley JD. When what you hear influences what you see: listening to an auditory rhythm influences the temporal allocation of visual attention. *Psychol Sci.* 2013;24(1):11–8.
13. Jones MR, Boltz M. Dynamic attending and responses to time. *Psychol Rev.* 1989;96(3):459–91.
14. Large EW, Snyder JS. Pulse and meter as neural resonance. *Ann N Y Acad Sci.* 2009;1169:46–57.
15. Large EW, Palmer C. Perceiving temporal regularity in music. *Cogn Sci.* 2002;26:1–37.
16. Schroeder CE, Lakatos P. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 2009;32(1):9–18.
17. Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science (80-).* 2008;320(5872):110–3.
18. Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P. Dynamics of Active Sensing and perceptual selection. *Curr Opin Neurobiol.* 2010;20(2):172–6.
19. Buzsaki G. Neuronal Oscillations in Cortical Networks. *Science (80-).* 2004;304:1926–1929.
20. Siegel M, Donner TH, Engel AK. Spectral fingerprints of large-scale neuronal interactions. *Nat Rev Neurosci.* 2012;13(2):121–34.
21. Henry MJ, Herrmann B. Low-Frequency Neural Oscillations Support Dynamic Attending in Temporal Context. *Timing Time Percept.* 2014;2(1):62–86.
22. Morillon B, Hackett TA, Kajikawa Y, Schroeder CE. Predictive motor control of sensory dynamics in auditory active sensing. *Curr Opin Neurobiol.* 2015;31:230–8.
23. Repp BH. Sensorimotor synchronization: A review of the tapping literature. *Psychon Bull Rev.* 2005;12(6):969–92.
24. Nozaradan S, Peretz I, Keller PE. Individual Differences in Rhythmic Cortical Entrainment Correlate with Predictive Behavior in Sensorimotor Synchronization. *Sci Rep.* 2016;6:20612.
25. Grahn JA, Rowe JB. Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cereb Cortex.* 2013;23(4):913–21.
26. Iversen JR, Repp BH, Patel AD. Top-down control of rhythm perception modulates early auditory responses. *Ann N Y Acad Sci.* 2009;1169:58–73.
27. Large EW, Herrera JA, Velasco MJ. Neural Networks for Beat Perception in Musical Rhythm. *Front Syst Neurosci.* 2015;9:159.
28. Morillon B, Schroeder CE, Wyart V. Motor contributions to the temporal precision of auditory attention. *Nat Commun.* 2014;5:5255.
29. Teki S, Grube M, Kumar S, Griffiths TD. Distinct neural substrates of duration-based and beat-based auditory timing. *J Neurosci.* 2011;31(10):3805–12.
30. Schaefer KP, Süss KJ, Fiebig E. Acoustic-

- induced eye movements. *Ann N Y Acad Sci.* 1981;374:674–688.
31. Maroti E, Knakker B, Vidnyanszky Z, Weiss B. The effect of beat frequency on eye movements during free viewing. *Vis Res.* 2017;131:57–66.
32. Recanzone G. Auditory Influences on Visual Temporal Rate Perception. *J Neurophysiol.* 2002;89:1078–1093.
33. Sekuler R, Sekuler AB, Lau R. Sound alters visual motion perception. *Nature.* 1997;385:308.
34. Berger CC, Ehrsson HH. Mental imagery changes multisensory perception. *Curr Biol.* 2013;(23):1367–1372.
35. Aston-Jones G, Rajkowsky J, Kubiak P, Alexinsky T. Locus Coeruleus Neurons in Monkey Are Selectively Activated by Attended Cues in a Vigilance Task. *J Neurosci.* 1994;14(7):4460–7.
36. Berridge CW, Waterhouse BD. The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Res Rev.* 2003;42(1):33–84.
37. Rajkowsky J, Kubiak P, Aston-Jones G. Correlations between locus coeruleus (LC) neural activity, pupil diameter, and behavior in monkey support a role of LC in attention. *Soc Neurosci Abstr.* 1993;19(974).
38. Sara SJ. Locus Coeruleus in time with the making of memories. *Curr Opin Neurobiol.* 2015;35:87–94.
39. Naber M, Alvarez GA, Nakayama K. Tracking the allocation of attention using human pupillary oscillations. *Front Psychol.* 2013;4:919.
40. Daniels LB, Nichols DF, Seifert MS, Hock HS. Changes in pupil diameter entrained by cortically initiated changes in attention. *Vis Neurosci.* 2012;29(2):131–42.
41. Trost JW, Labbe C, Grandjean D. Rhythmic entrainment as a musical affect induction mechanism. *Neuropsychologia.* 2017;96:96–110.
42. McGinley MJ, David S V, McCormick DA. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron.* 2015;87:179–92.
43. Murphy PR, Robertson IH, Balsters JH, O’Connell RG. Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology.* 2011;48(11):1532–43.
44. Joshi S, Li Y, Kalwani RM, Gold JI. Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron.* 2016;89(1):221–34.
45. Murphy PR, O’Connell RG, O’Sullivan M, Robertson IH, Balsters JH. Pupil Diameter Covaries With BOLD Activity in Human Locus Coeruleus. *Hum Brain Mapp.* 2014;35:4140–4154.
46. Gingras B, Marin MM, Puig-Waldmuller E, Fitch WT. The Eye is Listening: Music-Induced Arousal and Individual Differences Predict Pupillary Responses. *Front Hum Neurosci.* 2015;9:619.
47. Weiss MW, Trehub SE, Schellenberg EG, Habashi P. Pupils dilate for vocal or familiar music. *J Exp Psychol Hum Percept Perform.* 2016;42(8):1061–5.
48. Lange EB, Zweck F, Sinn P. Microsaccade-rate indicates absorption by music listening. *Conscious Cogn.* 2017;55:59–78.
49. Beatty J. Phasic Not Tonic Pupillary Responses Vary With Auditory Vigilance Performance. *Psychophysiology.* 1982;19(2):167–72.
50. Hong L, Walz JM, Sajda P. Your eyes give you away: prestimulus changes in pupil diameter correlate with poststimulus task-related EEG dynamics. *PLoS One.* 2014;9(3):e91321.
51. Liao HI, Yoneya M, Kidani S, Kashino M, Furukawa S. Human Pupillary Dilation Response to Deviant Auditory Stimuli: Effects of Stimulus Properties and Voluntary Attention. *Front Neurosci.* 2016;10:43.
52. Einhäuser W, Stout J, Koch C, Carter O. Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proc Natl Acad Sci U S A.* 2008;105(5):1704–9.
53. Laeng B, Eidet LM, Sulutvedt U, Panksepp J. Music chills: The eye pupil as a mirror to music’s soul. *Conscious Cogn.* 2016;44:161–78.
54. Damsma A, van Rijn H. Pupillary response indexes the metrical hierarchy of unattended rhythmic violations. *Brain Cogn.* 2017;111:95–103.
55. Kang O, Wheatley T. Pupil dilation patterns reflect the contents of consciousness. *Conscious Cogn.* 2015;35:128–35.
56. Davies MEP, Degara N, Plumley MD. Evaluation Methods for Musical Audio Beat Tracking Algorithms. *Tech Rep C4DM-TR-09-06 8 Oct 2009.* 2009;(October):17.
57. Tomic ST, Janata P. Beyond the beat: modeling metric structure in music and performance. *J Acoust Soc Am.* 2008;124(6):4024–41.
58. Hurley BK, Martens PA, Janata P. Spontaneous sensorimotor coupling with multipart music. *J*

- Exp Psychol Hum Percept Perform. 2014;40(4):1679–96.
59. Janata P, Tomic ST, Haberman JM. Sensorimotor coupling in music and the psychology of the groove. *J Exp Psychol Gen.* 2012;141(1):54–75.
60. Hurley BK, Fink LK, Janata P. Mapping the Dynamic Allocation of Temporal Attention in Musical Patterns. *Musical Patterns. J Exp Psychol Hum Percept Perform.* 2018;Advance On.
61. Leman M, Lesaffre M, Tanghe K. Computer code IPEM Toolbox. Ghent University, Ghent, Belgium; 2001.
62. McCloy DR, Larson ED, Lau B, Lee AKC. Temporal alignment of pupillary response with stimulus events via deconvolution. *J Acoust Soc Am.* 2016;139(3).
63. Hoeks B, Levelt WJM. Pupillary dilation as a measure of attention: A quantitative system analysis. *Behav Res Methods, Instruments, Comput.* 1993;25(1):16–26.
64. Wierda SM, van Rijn H, Taatgend NA, Martens S. Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proc Natl Acad Sci U S A.* 2012;109(22):8456–60.
65. Nozaradan S, Peretz I, Mouraux A. Selective Neuronal Entrainment to the Beat and Meter Embedded in a Musical Rhythm. *J Neurosci.* 2012;32(49):17572–81.
66. Tomic ST, Janata P. Ensemble: A Web-based system for psychology survey and experiment management. *Behav Res Methods.* 2007;39(3):635–50.
67. King-Smith PE, Grigsby SS, Vingrys AJ, Benes SC, Supowitz A. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Res.* 1994;(34):885–912.
68. Marvit P, Florentine M, Buus S. A comparison of psychophysical procedures for level-discrimination thresholds. *J Acoust Soc Am.* 2003;113:3348–61.
69. S.R. Research. Eyelink 1000 User Manual. Ltd. SRR, editor. Vol. 1.5.0. Ontario, Canada: SR Research Ltd.; 2009.
70. MATLAB. Natick, MA: MathWorks, Inc.; 2017.
71. Janata P. The neural architecture of music-evoked autobiographical memories. *Cereb Cortex.* 2009;19(11):2579–94.
72. Wang CA, Boehnke SE, Itti L, Munoz DP. Transient pupil response is modulated by contrast-based saliency. *J Neurosci.* 2014;34(2):408–17.
73. Welch PD. The Use of the Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE® Trans Audio Electroacoust.* 1967;AU-15:70–3.
74. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team the RDC. nlme: Linear and nonlinear mixed effects models (Version 3.1-113). . 2013.
75. R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>.
76. Cohen J. Statistical power analysis for the behavioral sciences. New York, NY: Routledge; 1988.
77. Selya AS, Rose JS, Dierker LC, Hedeker D, Mermelstein RJ. A practical guide to calculating Cohen's f^2 , a measure of local effect size, from PROC MIXED. *Front Psychol.* 2012;3.
78. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 1974;19(6):716–23.
79. Burnham KP, Anderson DR. Multimodel inference: Understanding AIC and BIC in model selection. *Sociol Methods Res.* 2004;33(2):261–304.
80. Schwarz G. Estimating the dimension of a model. *Ann Stat.* 1978;6(2):461–4.
81. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Mem Lang.* Elsevier Inc.; 2013;68(3):255–78.
82. Johnson PCD. Extension of Nakagawa & Schielzeth's R_GLM2 to random slopes models. *Methods Ecol Evol.* 2014;5:94–96.
83. Nakagawa S, Johnson PCD, Schielzeth H. The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J R Soc Interface.* 2017;14(134):20170213.
84. Nakagawa S, Schielzeth H. A general and simple method for obtaining R^2 from Generalized Linear Mixed-effects Models. *Methods Ecol Evol.* 2013;4:133–142.
85. Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015;67(1):1–48.
86. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77.

87. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1998;44(3):837–45.
88. Green D, Swets J. Signal detection theory and psychophysics. New York: John Wiley and Sons; 1996.
89. Mathot S, Van der Stigchel S. New Light on the Mind's Eye: The Pupillary Light Response as Active Vision. *Curr Dir Psychol Sci*. 2015;24(5):374–8.
90. Mathôt S. Pupilometry: Psychology, Physiology, and Function. *J Cogn*. 2018;1(1).
91. Laeng B, Sirois S, Gredeback G. Pupilometry: A Window to the Preconscious? *Perspect Psychol Sci*. 2012;7(1):18–27.
92. Polich J. Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol*. 2007;118(10):2128–48.
93. Nieuwenhuis S, De Geus EJ, Aston-Jones G. The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*. 2011;48(2):162–75.
94. Naatanen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol*. 2007;118(12):2544–90.
95. Allen J, Kraus N, Bradlow A. Neural representation of consciously imperceptible speech sound differences. *Percept Psychophys*. 2000;62(7):1383–93.
96. Navarro Cebrian A, Janata P. Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. *Brain Res*. 2010;1342:39–54.
97. Janata P. Acuity of mental representations of pitch. *Ann N Y Acad Sci*. 2012;1252(1):214–21.
98. Fink L, Hurley B, Geng J, Janata P. Predicting attention to auditory rhythms using a linear oscillator model and pupilometry. In: Proceedings of the Conference on Music & Eye-Tracking. Frankfurt, Germany; 2017.
99. Forth J, Agres K, Purver M, Wiggins GA. Entraining IDyOT: Timing in the Information Dynamics of Thinking. *Front Psychol*. 2016;7:1575.