Predicting the Beat Bin – Beta Oscillations Support Top-Down Prediction of the Temporal Precision of a Rhythmic Event

Sabine Leske[a,b,c,*], Tor Endestad[a,d], Vegard Volehaugen[a,d], Maja D. Foldal[a,d], Alejandro O. Blenkmann[a,d], Anne-Kristin Solbakk[a,c,d,e], Anne Danielsen[a,b]

Author affiliations:

[a] RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion,

University of Oslo, Oslo, Norway

[b] Department of Musicology, University of Oslo, Oslo, Norway

[c] Department of Neuropsychology, Helgeland Hospital, Mosjøen, Norway

[d] Department of Psychology, University of Oslo, Oslo, Norway

[e] Department of Neurosurgery, Oslo University Hospital, Rikshospitalet, Oslo, Norway

*Correspondence to: Sabine Leske

Email: sabine.jatzev@gmail.com

Keywords: Temporal Prediction, Temporal Precision, Beat Perception, Beta Oscillations, P-Center

Abstract

The brain exploits acoustic temporal regularities in the environment to enable prediction of the timing of future events and to optimize sensory processing and behavior. Periodic sensory inputs have been shown to entrain oscillatory brain activity. But to what extent is this entrainment mechanism instrumental to temporal prediction and under top-down control?

The current study addresses these questions by investigating whether the brain can use a cue to predict the precision of a beat, thereby exerting flexible endogenous control via neural entrainment according to the precision needed in the given sensory context. Using sound cues, we informed participants about the acoustic shape (sharp or smooth) of the upcoming target sound, thus inducing two alternative levels of predicted temporal precision: high or low. Participants were asked to detect the delay of a target sound relative to a series of preceding isochronous sounds that were physically constant across all conditions.

The results show that the information of the cue modulated the prediction of beat precision. Pre-target beta (15–25 Hz) power, as well as beta and alpha (8–12 Hz) phase locking, was enhanced if the listener predicted a target sound with a high compared to a low temporal precision. Interestingly, pre-target beta power correlated positively with behavioral performance in the delay-detection task but negatively with the variability of the perceptual center (P-center) of the sounds, demonstrating the tight relation between beta oscillations and perceptual temporal precision. This study provides new insight into the predictive processes underlying entrainment to different levels of temporal precision during beat perception and shows for the first time that these processes are in fact under top-down control.

Significance Statement

Neural entrainment offers a parsimonious explanation for the temporal prediction of strictly periodic events, but most temporal regularities are quasi-isochronous. How does the brain allow for perceptual flexibility in the anticipation of the temporal precision of an auditory rhythm, and to what extent is this mechanism under top-down control? By employing human electroencephalography and auditory cues in a delay-detection task, we investigated the neural mechanisms supporting the prediction of beat precision. We found that neural entrainment via beta activity is under top-down control and enables the prediction of temporal precision. Critically, behavior was influenced by the anticipated temporal precision and correlated with beta power. These results provide evidence that beta oscillations—an important neural substrate for beat entrainment—are instrumental to the top-down prediction of beat precision.

Introduction

Humans, among some other species, show a strong tendency to synchronize their movements to periodic events that express a regular pulse or beat (1), and their oscillatory brain activity synchronizes with rhythmic sounds (2). This aspect of beat perception has been referred to as neural entrainment (2–4) and there is an active debate if it is instrumental to temporal prediction and under active top-down control (2, 3, 5).

However, high precision in the temporal processing of auditory events is not needed in all contexts. In fact, both the temporal regularity and distinctness of rhythmic acoustic stimuli can vary substantially. Most environmental rhythms and patterns in human behavior, such as walking, dancing and speech, *do not* show strict isochrony (3). Moreover, if the crucial content of, for example, a speech syllable or a musical note is stretched out in time, it might be beneficial to lower the perceptual temporal precision and rather widen the time window for perceptual processing. Hence, dynamically adjusting the temporal precision of sensory predictions in accordance with context- and task-specific demands is fundamental. How the brain prepares for sounds with varying temporal precision has not been investigated systematically with respect to neural entrainment, despite precision being an important aspect of temporal prediction. The goal of the current study was to reveal neural oscillatory mechanisms underlying predictions about the temporal precision of periodic auditory events, and how the involved processes can flexibly adapt in a top-down manner.

Research into human perception of musical rhythm has demonstrated that the perceived precision of a beat depends on the acoustic shape of the sounds or other input. For example, a sharp sound with a short attack phase is perceived with a high temporal precision whereas a smooth sound with a long attack phase, like a violin note, gives rise to a perceived pulse that allows for some flexibility and has lower temporal precision (6). The perceived temporal precision of a sound can be measured via the perceptual center location (P-center) and its variability (6–8). The P-center reflects the subjectively perceived time of occurrence of a sound (9, 10). This can vary substantially for different sound types, leading e.g., to high (for a sharp sound) or low precision (for a smooth sound) in the perception of the time of occurrence across instances (6–8, 10–13). In music cognition studies, this has been referred to as the beat bin (14), which denotes the concept that a beat is not defined as a definite point in time but rather by a probability distribution. Hence the beat bin reflects rhythmic tolerance or the temporal window within which sonic events can occur and still be perceived as (part of) the beat (6, 8, 15). The temporal precision of perceiving a beat seems to vary systematically with the precision needed in the given acoustic context (6). Accordingly, we will use the term precision when referring to the width of a beat bin throughout the rest of the paper. In the same vein, it has been shown for speech processing, that even if the acoustic sequence is not perfectly isochronous, neural entrainment is tolerant to timing variability ((5, 16–18) but see (19)). In perceptual neuroscience, a systematic investigation of the effect of the auditory stimulus shape on neural entrainment is missing (but see (20, 21)), except for studies investigating neural tracking of the acoustic envelope during speech perception (22). It is known that the input shape can have a big impact on neural synchronization, considering the fact that oscillators are not shape invariant (23), but the nature of this relation is still unclear.

The phase reset of ongoing spontaneous slow (<10 Hz) oscillations is typically considered to be the mechanism underlying neural entrainment to external rhythms (5, 24–26). However, it might also manifest as the alignment of rhythmically generated cortical oscillatory bursts entailing higher (e.g., beta-band: 13-30 Hz) frequency ranges (5). While neural entrainment can arise from environmental inputs directly, these inputs can also be linked with rhythmic motor sampling patterns that reset and entrain neural activity in sensory areas (1, 5). This neural process has been referred to as active sensing (4, 5, 27).

In support of this framework, (28, 29) revealed that rhythmic movement enhances acoustic temporal perception. Beta activity originating in the sensorimotor cortex was time-locked to the expected auditory targets and directed toward auditory brain regions, while being coupled to the

phase of delta oscillations (1-3 Hz) at the beat rate of the stimuli (29). Moreover, Fujioka et al. (30, 31) reported that oscillatory beta activity of interacting motor and auditory regions in humans, predicts the timing of an auditory beat, peaking already before its onset. Beta oscillations might thus represent the internalization of predictable intervals of a beat and reflect the efference copy signal from the motor system generated at the tempo of the beat (1, 30, 31). If beta oscillations play a functional role in auditory beat prediction, they should be modulated in a top-down manner by predictions about the temporal precision of beat. In the current study, we investigated if pretarget beta oscillations also reflect neural expectations about the temporal precision of a sound embedded in a beat sequence. Indeed, it has been proposed that a role of beta bursts may be to phase reset slower oscillations, aligning excitability phases with expected stimuli, and thus optimizing perception at a low energetic cost (32, 33). However, mechanisms underlying this type of active sensing, and the respective neural pathways controlling the gain of beat perception, are not fully known (34). Importantly, it is still actively debated whether the oscillations of the auditory cortex passively entrain to sensory input, or whether neural entrainment is under active top-down control (2, 4, 35–39).

In the present study, we examined whether neural processes supporting the prediction of different levels of temporal precision in beat perception can be under active top-down control. Participants heard a sequence of isochronous sounds and had to detect the delay of a target sound at the end of the sequence (Figure 1). We built on the findings from human rhythm perception research reported above, employing acoustic features (e.g., short or long attack phase) that create stimulus shapes shown to induce high or low temporal precision in beat perception (6–8, 40). By manipulating the listener's expectation about the acoustic shape (sharp or smooth) of target sounds via sound cues, we induced different predictions regarding the temporal precision (high or low) of a target sound following an isochronous beat sequence (Figure 1). We kept the physical stimulus features of the entraining sounds succeeding the sound cue (and were part of the isochronous beat sequence) constant across all conditions. This ensured that condition differences can be solely explained by differences in the neural prediction of precision. The sound cue was valid in the majority of the trials to enable prediction of the acoustic shape of the target sound and thereby its perceptual temporal precision. Participants performed a delay detection task, which provides an indication of whether the cue information was behaviorally relevant for the anticipation of temporal precision and if it was utilized by the temporal prediction processes of the brain.

In accordance with studies investigating temporal prediction, beat perception and entrainment (1, 4, 29–31, 41, 42), we hypothesized that the delay detection performance would be affected by cue validity (43–46), particularly for the high-precision cue. Furthermore, we anticipated that pre-target beta power and phase locking would be enhanced when participants expected high temporal precision, subserving the prediction of beat precision in a top-down manner.

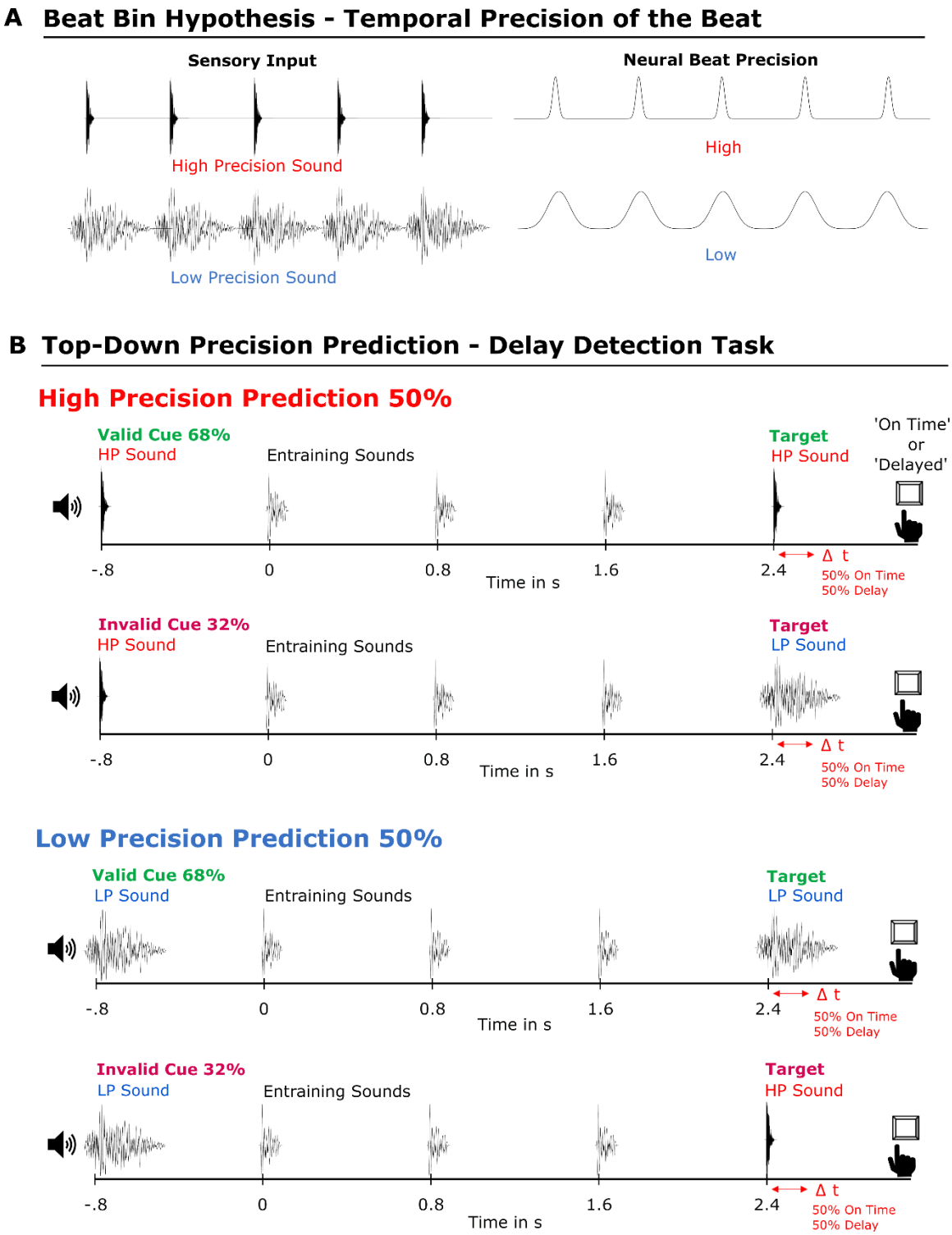

Figure 1. Hypothesis and Experimental Design. Stimulus plots are a visualization of the sounds used in the study: The HP sound (High-Precision, red), the LP sound (Low-Precision, blue) and the

entraining sound (black). A) Exemplary illustration of the Beat Bin Hypothesis. The acoustic shape of the sound modulates the temporal precision of the perceived (neural) beat. The acoustic shape of the HP sound (red) induces a high-precision beat percept. The acoustic shape of the LP sound (blue) induces a low precision beat percept, which allows for enhanced flexibility with respect to perceived beat timing. B) Experimental Design: two-alternative forced choice delay detection task incorporating a cueing paradigm. Shown are all factor combinations: Precision Prediction, Cue Validity, and Delay (2x2x2). Participants were asked to judge if the target sound was delayed or on time relative to an isochronous sequence of three entraining sounds. Target sounds were on time or delayed with 50/50% probability. The HP sound Cue indicated an HP sound target, inducing a High-Precision Prediction (red). The LP sound Cue indicated a LP sound target, inducing a Low-Precision Prediction (blue). The cue validity was 68%. In the case of an invalid cue (32%), the target sound with low perceptual temporal precision was presented for High-Precision Prediction and vice versa. The slightly earlier onsets of the entraining and LP sounds schematically illustrate the P-center alignment in the study.

Materials and Methods

Participants. Participants were recruited at the University of Oslo. In total, 28 healthy volunteers participated in the study after giving written informed consent. They were compensated with gift cards worth 300 NOK. Beforehand, the Department of Psychology's internal research ethics committee had approved the study, which was conducted in agreement with the Declaration of Helsinki. Nine participants were excluded because their performance on the delay detection task was not significantly higher than chance level (see "Statistical analysis" below). The final sample included 19 subjects (13 female, all right-handed, mean age = 25.79 years, std =3.49 years). All reported normal vision and hearing, no central nervous system damage or disorder, and no cognitive difficulties. They also reported that they did not receive any psychiatric treatment and no medication for illness at the time of participation.

Experimental Task and Stimuli. Using a cueing paradigm, we manipulated the prediction of the acoustic shape (sharp vs. smooth) of the target sound, thereby inducing different predictions regarding the perceptual temporal precision of the target (High- vs. Low-Precision Prediction). Both the cue and the target sound were embedded in an isochronous sound sequence (Figure 1). The cue was either a High- (50%) or Low-Precision (50%) sound (further described below), presented at the beginning of each trial. The cue indicated the acoustic features of the target sound (i.e., whether the target to expect was a High- or Low-Precision sound). In case of a Valid Cue (68% of trials), the target sound was identical to the cue, whereas in the invalid condition it was not identical (32% of trials). The target sounds followed three isochronous entraining sounds, which were physically constant across all conditions. Participants had to detect whether the target sound was on time or delayed in a two-alternative forced choice task. The cue, entraining sounds and target sound were presented with an inter-stimulus interval (ISI) of 800 ms. This resulted in a beat frequency of 1.25 Hz (Figure 1).

To induce high or low perceptual temporal precision, we utilized sounds with two different acoustic shapes that induce a high or low P-center variability, respectively (6, 40). We refer to these sounds as High- (HP) or Low-Precision (LP) sounds. These differently shaped sounds were originally designed to investigate acoustic factors influencing the P-center location and variability (6, 40). P-center locations are reported in ms relative to the physical onset of the sound. The noise sound with the highest P-center variability results (group mean std = 28 ms) was selected as a sound inducing low perceptual temporal precision, referred to as Low-Precision sound. This sound had a duration of 400 ms, 50 ms rise time, and a center frequency of 100 Hz (6). The sound with the lowest P-center variability (group mean P-center std = 2 ms) was selected as a sound inducing high perceptual temporal precision, referred to as High-Precision sound, with a duration of 40 ms,

2 ms rise time, and a spectral centroid of 2370 Hz (40). The three entraining sounds following the cue sound, had a duration of 100 ms, 3 ms rise time and a center frequency of 100 Hz, and had been shown to induce a P-center variability of 7.78 ms (6).

Except for the High-Precision sound, where we assumed a P-center location of 0 ms, based on previous studies (40), all sound types were aligned to participant-specific P-center locations (see the procedure for the P-center estimation below), to ensure that the sound series were perceived as isochronous. For example, this means that for a particular participant having a P-center location of 15 ms for the entraining sound, the onset of the sound was accordingly adapted to be 15 ms earlier. This particular case would then result in an ISI of 785 ms between, for example, the HP sound cue (P-Center of 0 ms) and the following entraining sound, but would induce a perceived isochrony of 800 ms. Given its P-center location of 0 ms (corresponding to the physical onset of the sound), the HP target sound occurs approximately at 2.4 s for the On Time condition. The target sound was also P-center aligned relative to the entraining sound, resulting in possible earlier or later target onsets. The ISI between all three entraining sounds was always 800 ms.

Overall, and for each factor combination, 50% of the target sounds were “on time“ and 50% were “delayed”, resulting in a 2 x 2 x 2 factorial design, with factors Precision Prediction (High/Low) x Cue Validity (Valid/Invalid) x Delay (On Time/Delayed) and eight conditions in total (Figure 1). The sound cues were presented in a pseudorandom order, preventing repetition in successive trials.

Participants indicated if the target sound was “on time” (isochronous to the preceding entraining sounds) by pressing the left green button of a response box (Cedrus Response Pad, Model RB-740, www.cedrus.com) with the index finger of their right hand, or “delayed”, by pressing the right red button with the middle finger of their right hand. Participants were given 1.3 s to respond before the next trial was automatically initiated. They were asked to fixate a central black cross on a gray background that remained on the screen during the entire experimental block.

The target delay threshold was determined individually for each participant after one initial training block consisting of 100 trials. The training block started with a delay of 90 ms for the High-Precision target sound and 100 ms for the Low-Precision target sound. After the training block, the threshold was automatically adapted. A binomial test was used to determine if the delay detection performance was better than chance level (with respect to accuracy), and the target delay was increased if the performance was too low (by either 30 ms if the accuracy was below 30%, by 20 ms if the accuracy was below 50% or by 10 ms if the accuracy was above 50% but not significantly better than chance level). If the performance was too high (approaching 100% accuracy), then the delay was reduced (by 20 ms if the accuracy was above 90% or by 10 ms if the accuracy was above 80%) to avoid behavioral ceiling effects (see Supplementary Material for the algorithm used). Individually adjusted temporal delays were held constant across trials and blocks for the main experimental task.

The main experimental task consisted of three blocks of 100 trials each, resulting in 300 trials. The inter-trial intervals (ITI) were uniformly distributed between 1.2 and 2.2 s, with a mean length of 1.7 s.

Estimation of the Loudness of Stimuli. In a pilot experiment, we asked 10 participants to judge if the loudness level of the HP and LP sounds were perceptually similar to the entraining sound. Participants were allowed to increase or decrease the loudness level of the HP/LP until it was perceptually similar to the loudness of the entraining sounds. The mean of the amplitude factors for the HP and LP sound across all 10 participants was used as the new amplitude factor for the main experiment (please see stimuli and scripts in the online repository). The sound loudness was at a comfortable level and was constant across participants and blocks.

P-center estimation. Before the main task, individual P-Centers were estimated via a click alignment task (see (6, 7) for details of the procedure), which is a computerized task to estimate the P-Center location and variation (measured in standard deviation). Danielsen et al. (6) used only four trials for the estimation of the P-center location and variability, since pilot tests showed that this was sufficient to provide a good estimate. For the current study, we increased the number of test trials to get reliable estimates, especially for the LP sound, since this sound creates the highest perceptual variability within a subject. We used 9 trials for the LP sound, 6 trials for the entraining sound, and 3 trials for the HP sound. We used only 3 trials for the HP sound, since previous studies had found that 3 to 4 trials are sufficient (6, 40) and we did not intend to use the result from the click alignment task for the P-Center alignment of the experimental stimuli. A previous study (40) had already shown (using a click alignment task) that the HP sound had a grand average P-center location of 0 ms and a standard deviation of 2 ms (mean over 3 trials). Accordingly, we used the respective P-center location for the stimulus alignment in the current study. Please see Figure 2 and Supplementary Table 1 for the individual P-center locations and standard deviations.

Post-Experiment Self Report. All participants completed a self-report (see Supplementary Figure 1) after finishing the experimental task. The questionnaire included six open questions regarding their behavioral strategies and their subjectively perceived use of the cue while performing the experimental task.

Goldsmith Musical Sophistication Index. We collected data for the Goldsmith Musical Sophistication Index (Gold-SMI) questionnaire (65), which was not used for the current study (see Supplementary Material).

Procedure. The session started with participants reading and signing the consent form before setting up the EEG-recording and the experimental tasks. Auditory stimuli were played at a comfortable volume from two Genelec speakers (model 8030 W) flanking both sides of the screen.

After performing the click-alignment task to estimate the individual P-center location and variability (approximately 5 minutes), participants did a training block to estimate the delay threshold (approximately 10 minutes). After a short break, participants performed the three experimental blocks of the main experiment (approximately 12 minutes each). Participants decided if they would like to take short breaks in between blocks. The entire session, including EEG preparation, experimental tasks, breaks, and administration of questionnaires, lasted 2-3 hours.

EEG recording. EEG and electro-oculography (EOG) data were recorded using a BioSemi Active Two 64 Ag-AgCl electrode system (BioSemi, Amsterdam, Netherlands). A BioSemi headcap was used to position the 64 electrodes on the scalp of the head, according to the international 10–20 system for electrode placement. Data were sampled at 1024 Hz during online recording. Four external electrodes measured vertical and horizontal eye movements (EOG) and were positioned above and below the right eye, and lateral to the participant's left and right eyes. Two additional electrodes were positioned on the left and right earlobes for later EEG re-referencing.

Behavioral analysis. We assessed the delay detection performance using a binomial test to ensure that only participants performing significantly better than chance level were included in the analysis. Due to this criterion, 9 subjects performing significantly below chance level were excluded, resulting in a final sample of 19 participants.

To quantify the behavioral performance of the delay detection task, we calculated the d-prime (66) for each condition based on correct and incorrect responses (On Time or Delay), which is a performance measure for sensitivity based on signal detection theory. We applied a log-linear rule, adding 0.5 to Hits and False Alarms (FA) and 1 to the total number of On Time and Delay trials. By taking the inverse of the normal cumulative distribution function (Zstd) for Hits and FA, the d-prime was calculated by: Zstd(Hits) - Zstd (FA). This framework also allows for estimating the response bias separately from sensitivity, which is calculated by: -0.5 *Zstd(Hits) + Zstd(FA).

Since we were mainly interested in evaluating if the Cue was behaviorally relevant for the perception of the target and the delay detection task, we pooled over the On-Time and Delayed conditions, and calculated the d-prime and response bias for the factor combinations: High-Precision-Valid (High-Precision Cue and High-Precision Target), High-Precision-Invalid (High-Precision Cue and Low-Precision Target), Low-Precision-Valid (Low-Precision Cue and Low-Precision Target), and Low-Precision-Invalid (Low-Precision Cue and High-Precision Target), resulting in a 2 x 2 (Precision Prediction x Validity) factorial design.

Reaction times were analyzed only for the correct responses for each of the four conditions and in addition for the Valid versus Invalid conditions (pooled over the levels of the factors Precision Prediction and Delay).

EEG analysis. *Data Preprocessing.* The focus of the analysis of the neurophysiological data was to investigate if the cue signaling the temporal precision of the target (High/Low) affected the neural processing of the entraining sounds, and if the pre-target oscillatory activity mirrors predictive processes optimizing temporal perception. Therefore, we concentrated on the comparison of the High- and Low-Precision Prediction condition, pooling over the two factors Delay and Validity (both are unknown to the participant before the target arrives), and including only trials with correct responses, to reduce noise (e.g., attention not on the task etc.).

EEG data were analyzed offline using custom-written scripts and the Fieldtrip toolbox (67) in MATLAB (R2020a, Mathworks Inc., Natick, MA, USA). Continuous EEG data were high-pass filtered with a zero-phase Butterworth IIR filter with a half-amplitude cut-off at 0.1 Hz (18dB/oct roll-off) to remove slow drifts in the data and down-sampled to 1000 Hz. The down-sampling was performed via the Fieldtrip toolbox, which in turn makes use of the Matlab function “decimate”, incorporating a 400 Hz low-pass Chebyshev Type I impulse response (IIR) filter of order 8 that is applied to the data before down-sampling, to prevent anti-aliasing. The data were offline referenced to linked earlobes. Noisy segments and bad channels in the continuous EEG data were defined by visual inspection (i.e., muscle artifacts). Following removal of bad channels, an independent component analysis (ICA) was used to identify and remove blinks and horizontal eye movements in the continuous EEG data. Removed ocular components were identified by visual inspection.

The data were referenced to a common average, segmented into epochs of -0.8 to 3.2 s, time locked relative to the onset of the first entraining sound and epochs containing artifacts were removed. Since the different stimuli types were P-center aligned, the alignment to the first entraining sound ensured that trials could be compared across participants. Please note that the three entraining sounds, following the cue and preceding the target sound, were presented at a constant rate of 1.25 Hz (SOA 0.8 s) and were physically constant across all conditions. This ensures that condition differences during the entrainment period can be attributed to different predictions of temporal precision. Epoched data were demeaned and detrended.

Frequency Analysis. Time-frequency resolved power was calculated via wavelet transform (Morlet wavelets of 3 cycles) for the entire epoch (-800 to 3200 ms) and the beta frequency range (15-25 Hz) in steps of 1 Hz and with a sliding window in steps of 5 ms. Single-trial power values were baseline corrected relative to the power mean across the entire epoch (0 – 2.4 s) for each frequency bin and expressed in percentage change. Power values were averaged across frequencies of the beta-band (15-25 Hz) to retain beta-power time series. Beta-power time series were smoothed with a moving average of 50 samples (equivalent to 250 ms, using the Matlab function ‘smoothdata’). The average beta-power time series across trials per participant was calculated as input for the statistical analysis. The time window of interest was the interval between the last (third) entraining sound and the target sound onset, lasting from 1.6 to approximately 2.4 s.

The onset of the target sound varied across conditions and participants. Depending on the individual P-center locations of the LP and the entraining sound, the earliest possible target onset was at 2.28 s. Please see Supplementary Table 1 for individual P-center locations and standard deviations. We therefore restricted the statistical group analysis to this time window (1.6 - 2.28 s)

to avoid contamination of activity in the pre-target time period via the evoked response of target sounds in some participants.

In order to identify spectral peaks in the power spectrum indicative of oscillatory activity, the power spectrum of the last entrainment period before target onset was calculated by applying a Fourier transform to the respective time window (1.6 – 2.4 s) using a Hanning window and subsequently averaging power spectra across trials for each task condition and participant.

In addition, data were transformed into complex wavelet coefficients via time-frequency wavelet transform, using Morlet wavelets (3 cycles), for the frequency range of 8 to 30 Hz in steps of 0.5 Hz and with a sliding window in steps of 10 ms for the pre-target time window 1.8 – 2.28 s. This time window was chosen according to the results of the statistical beta-power time series contrast, showing significant modulation by the factor Precision Prediction (see the Results section about beta band power).

Wavelet coefficients were used to compute time-frequency resolved single-trial power (squared absolute value) and inter-trial phase coherence (ITPC) to estimate phase consistency across trials (68) for each cue condition (High- and Low-Precision Prediction) and participant. Rayleigh's Z-transformation was applied to the ITPC values to take into account differences in the number of trials between conditions (68), which were small but had to be considered, since it can bias the ITPC measure. Rayleigh's Z-transformation was calculated by: ITPCz = N*ITPC^2, with N representing the number of trials per condition.

Because phase coherence can be biased by power differences (69), we excluded electrodes and time points showing significant beta-power differences between *High-* and *Low-Precision Prediction.* Approximately 90% of the electrodes showed no significant beta power differences for the pre-target time window between 2.215 and 2.28 s, right before target onset. This spatio-temporal selection (2.215 - 2.28s, all 64 sensors excluding 6 parietal electrodes with significant power modulation: TP8, CP6, P6, P8, P10, PO8) of ITPCz values was submitted to the statistical group comparison.

Statistical analysis. For the behavioral measures (d-prime and response bias), we conducted two repeated measures ANOVAs on the factors Precision Prediction and Validity (2 x 2 factorial design), pooled over the conditions On-Time and Delay. Assumptions of normality with respect to the residuals were tested via Lilliefors goodness-of-fit test of composite normality. Dependent-samples t-tests were used for the simple contrasts for all conditions of the 2 x 2 factorial design. Simple contrasts for reaction times were also computed for the conditions Valid versus Invalid (pooled over the levels of the factors Precision Prediction and Delay). P-center locations and standard deviations were compared via non-parametric signed-rank tests for all four conditions of the 2 x 2 (Precision Prediction and Validity) factorial design (pooled over the condition levels of Delay), since assumptions of normality were violated.

Statistical group analysis of beta power time-series, spectral modulation of beta power time series, and ITPCz for the two conditions of interest (High- versus Low-Precision Prediction) was conducted via one-tailed non-parametric cluster-based permutation dependent-samples t-tests, correcting for multiple comparisons (70), applying 10000 permutations. Based on previous studies showing that beta power correlates positively with temporal accuracy and prediction (1, 29, 30, 41, 42, 48), we had a clear hypothesis about the directionality of the effect. A beta power and phase increase go along with enhanced temporal precision.

Single trial time-frequency resolved power values were submitted to two-sided independent-samples t-tests contrasting High- versus Low-Precision Prediction, to retain individual t-values for the group correlation between neural and behavioral data. The respective correlations of beta-power t-values (t-contrast of High- versus Low-Precision Prediction) with behavioral measures (d-prime and P-center variability) were investigated by performing one-tailed non-parametric Spearman's rank correlations between d-prime values for the Valid High-Precision Prediction

condition and the individual t-values (signal power contrast of High- versus Low-Precision Prediction) using non-parametric cluster-based permutation tests to correct for multiple comparisons, performing 10000 permutations.

To estimate the effect size for cluster-based permutation statistics, we computed average (power or phase) values for each participant and condition across the cluster extent (channels, time bins and frequency bins if applicable) and calculated Cohen's d effect size for within-subjects designs (71). For correlational analysis, the average Spearman's rho across the cluster (channels, time bins and frequency bins) is reported as effect size.

Results

We reasoned that if neural entrainment to a regular beat is under top-down control, then a cue indicating the perceptual temporal precision of a target sound following regular beats should affect the detection of a delay of the respective sound (i.e., if a sound is on time or delayed relative to the beat). We used a two-alternative forced choice delay detection task to test if cue validity affects detection performance, while measuring EEG. Participants (n = 19) had to detect the temporal delay of a target sound and respond via button press (indicating Delayed or On Time). Importantly, the target sound followed three entraining sounds that were physically constant across conditions and presented at a rate of 1.25 Hz (see Figure 1).

Trials started with a sound cue indicating the acoustic features of the upcoming target sound. By this, we aimed to induce expectations regarding the beat precision at two different levels: *High*- (50%) and *Low-Precision* (50%) *Prediction*. The cue preceded three entraining beats by approximately 800 ms and was either a High-Precision (HP) sound (Valid Cue, High-Precision Prediction) indicating the same HP sound as a target, or a Low-Precision (LP) sound (Valid Cue, Low-Precision Prediction) indicating the respective LP sound as a target (68% *Cue Validity*) after the three entraining sounds. In case of an invalid cue (32%), a target with the acoustic features of the other condition was presented.

Behavior is affected by the predicted temporal precision. To ensure an isochronous beat perception of 1.25 Hz between different sound types, the sounds were aligned according to individual P-center locations measured via a click alignment task at the beginning of the experiment (see Methods). First, we tested if the stimuli showed the expected perceptual effect with respect to temporal precision, which is reflected by the P-Center variability measured via the standard deviation (std) across trials (Figure 2). We replicated the directionality of the results by Danielsen et al. (6) and Câmara et al. (40), showing that the HP sound had the earliest P-center location relative to sound onset (group mean: 6.84 ms) and smallest variability (group std: 29.48 ms), followed by the entraining sound (group mean: 18.28 ms, group std: 42.77 ms), and the LP sound with the highest mean values for the P-center location (group mean: 50.99 ms) and variability (group std: 62.41 ms). Signed-rank tests showed that the two sound types (HP and LP sound) differed significantly from each other in their P-center location ($Z = 4.04$, $P = 0.00003$, Cohen's $d = 1.6319$ (very large effect size)) and variability ($Z = 3.18$, $P = 0.0007$, Cohen's $d = 0.8271$ (large effect size), Figure 2). This result is important since it ensured that the acoustic features used for the two cue conditions (High- versus Low-Precision Prediction) had the intended effect on the perception of temporal precision (inverse of variability).

The HP sound did not differ significantly from the entraining sound in P-center location ($Z = 0.93$, $P = 0.175$), but in P-center variability ($Z = 1.927$, $P = 0.027$, Cohen's $d = 0.4467$ (small effect size)). In contrast, the LP sound differed significantly from the entraining sound in P-center location ($Z = 3.24$, $P = 0.0006$, Cohen's $d = 1.2499$ (very large effect size)), but not in P-center variability, showing only a trend effect ($Z = 1.61$, $P = 0.0542$). This confirmed that the perceptual effect (P-

center location and variability) of the acoustic features of the entraining sound was somewhere in between the HP and LP sounds. The target delay thresholds, estimated from the training block to ensure above chance level performance in the delay detection task, also differed significantly between the HP (threshold group mean: 100 ms) and the LP sounds (threshold group mean: 110 ms), with a higher mean for the LP sound (Z = 3.33, P = 0.0004, Cohen's d = 1 (large effect size), Figure 2).

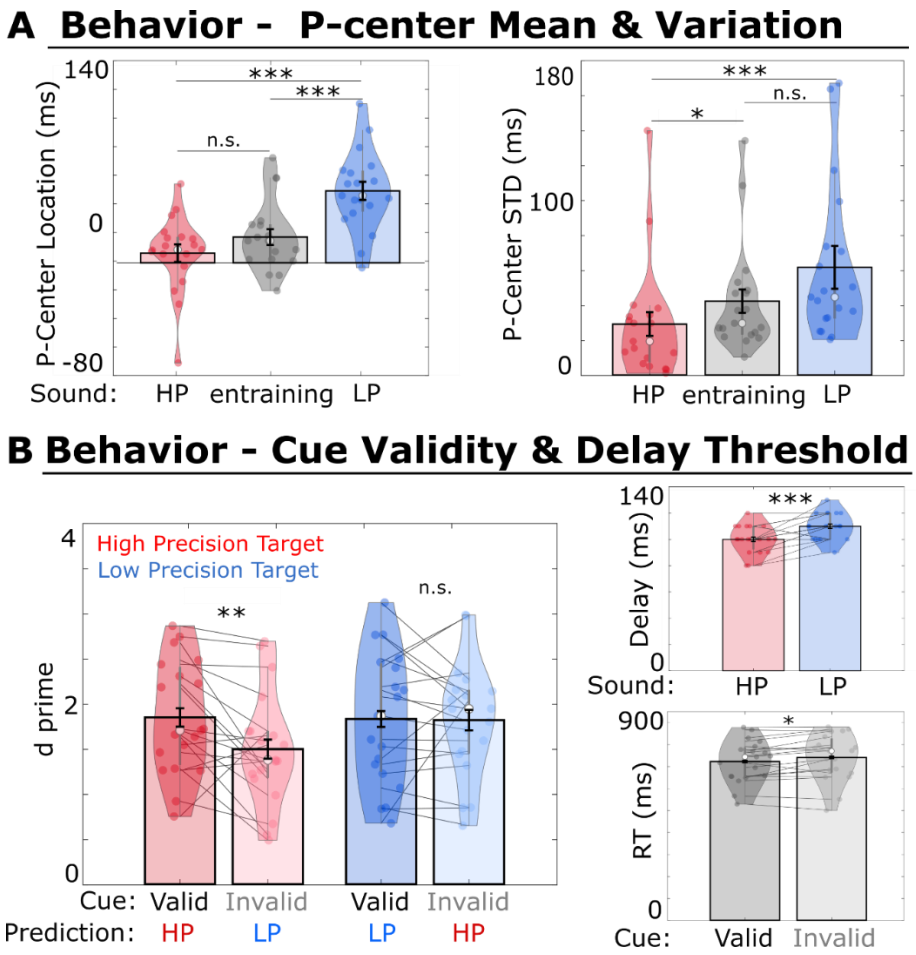

Figure 2. Behavioral Results. Bar plots represent the mean and black error bars represent the adjusted standard error (SE) across participants for a within-subject design, according to the Cousineau-Morey correction (55). Violin plots show kernel density estimates and data points (56). Asterisks refer to significant differences between conditions (*<0.05, **<0.01,***<0.001). A) P-center location and variability (std) for all stimuli, HP (High-Precision, red), entraining (gray) and LP (Low-Precision, blue) sound from the click alignment task. B) Behavioral results for the temporal delay detection task. Left: Valid cue conditions are depicted in darker and invalid cue conditions are depicted in light colors. Sensitivity (d-prime) was relatively increased for the Valid - High-Precision Prediction Condition (dark red) compared to the Invalid - Low-Precision Prediction Condition (light red), when participants were judging the temporal delay of a high-precision target (two left bars, red). Sensitivity was not affected by the validity of the cue for low-precision targets

(Valid - Low-Precision Prediction (dark blue) versus Invalid - High-Precision Prediction (light blue)). Upper Right: Individual delay threshold results from the training block preceding the main experiment. These thresholds were used for the main experimental blocks. Lower Right: Reaction times for the Valid versus Invalid conditions (pooled over the factor levels Precision Prediction and Delay).

We calculated the d-prime measure to estimate delay detection performance and response bias (i.e., tendency to respond “On-time” or “Delayed”) for the factors Precision Prediction (High/ Low) and Validity (Valid/ Invalid), pooled over the two levels of the factor Delay (On-Time/ Delayed). A repeated measures ANOVA with the factors Precision Prediction and Validity showed a significant main effect of cue Validity ($F(1,18)$ 7.34, $P = 0.014$), but not Precision Prediction ($F(1,18) = 1.51$, $P = 0.236$), nor a significant interaction between Validity and Precision ($F(1,18) = 2.461$, $P = 0.13$). Dependent samples t-test showed that perceptual sensitivity (d-prime) was significantly increased for validly cued HP-targets (HP Prediction Condition) compared to invalidly cued HP-targets (LP Prediction Condition, $t(18) = 2.7107$, $P = 0.0072$, Cohen’s $d = 0.6219$ (medium effect size)). This reveals that the delay detection performance deteriorated when the participants expected a target with a low perceptual temporal precision but received a target with a high perceptual temporal precision (Figure 2). In contrast, there was no effect of Validity for trials with an LP-target sound, as there was no significant difference between validly cued LP-targets (LP-prediction condition) and invalidly cued LP-targets (HP-prediction condition) with respect to the d-prime measures ($t(18) = 0.0961$, $P = 0.4622$). Moreover, there were no significant differences for the response bias for any of the conditions. Altogether, these results indicate that cueing the temporal precision of the target sound was behaviorally relevant for evaluating the timing and detecting delays of the target with a high perceptual temporal precision (Figure 2). Since delay detection performance was only deteriorated for incorrectly cued HP-targets, this suggests involvement of prediction in the processing of perceptual temporal precision. An ANOVA showed a significant main effect of Validity on reaction times ($F(1,18) = 5.7866$, $P = 0.027$). Dependent samples t-tests revealed that reaction times were significantly slower ($t(18) = -2.4$, $P = 0.0136$, Cohen’s $d = -0.5519$ (medium effect size)) for Invalid trials compared to Valid trials (pooled over Precision Prediction and Delay, Figure 2).

Interestingly, the majority of the participants (79%) stated in the (post-experiment) self-report that they did not consciously use the information of the cue to solve the task. Participants were instructed not to move during the experiment and 47% reported that they did not move to the beat at all, 37% reported that they moved to the beat on some occasions, and 13% reported that they moved to the beat during the experiment. The majority (68%) of the participants indicated that they had the impression that it was easier to judge the temporal delay of the HP target sound. See Supplementary information for the other results of the questionnaire.

Beta band power is modulated by the predicted perceptual temporal precision. To reveal neurophysiological processes underlying the prediction of beat precision, we focused on pre-target beta-oscillations during the entrainment interval, aligned to the first entraining sound. The presentation of sound cues was assumed to induce different levels of prediction of the perceptual temporal precision of the target sound that was embedded in the beat sequence. Accordingly, we focused on the main factor Precision Prediction (High versus Low), comparing trials starting with a HP sound cue with trials starting with an LP sound cue.

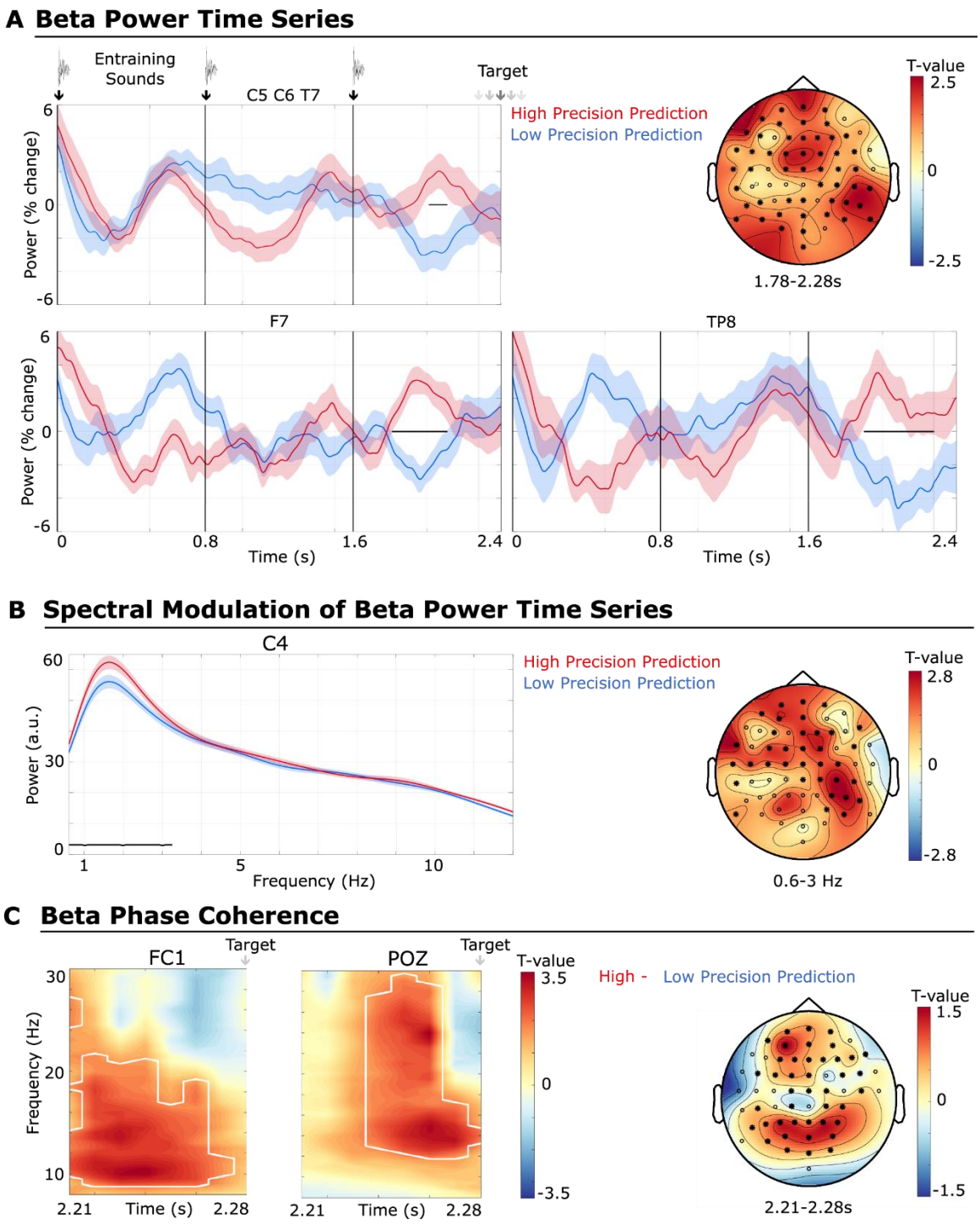

Figure 3. Beta power and phase effect for High- versus Low-Precision Prediction. Shaded areas represent the adjusted standard error (SE) for a within-subject design, according to the Cousineau-Morey correction (55). Black lines indicate significant differences in the temporal dimension. White lines enclose significant differences in the time-frequency dimensions. The spatial extent of the significant clusters is depicted by bold stars in the topographical plots. The small gray error symbols

for the target symbolize multiple possible onsets for the target according to individual P-Centers. A) Pre-target Beta-power time series (percentage change) for High- (red) and Low-Precision (blue) Prediction for exemplary sensors over centro-temporal (T7, C5, C6), frontal (F7) and parietal (TP8) regions. Pre-target beta power is significantly enhanced for the High-Precision Prediction. B) Spectral modulation of beta-power time series for the High- (red) and Low-Precision (blue) Prediction conditions. The spectral modulation of the pre-target beta power time series is significantly enhanced for the High-Precision Prediction condition within the delta (1-3 Hz) band (topographical plot shown on the right), theta (2-7 Hz), and alpha (8-12 Hz) frequency range (data not shown). C) Intertrial Phase Coherence (ITPC). Pre-target phase locking is significantly enhanced for the High-Precision Prediction condition in the alpha and beta band range.

Grand average spectra across participants were calculated for the entire entrainment period (0 - 2.4 s). The spectra revealed narrowband peaks of power at the theta (4-6 Hz), alpha (8-12 Hz), and beta (15-25 Hz) frequency ranges, indicating the involvement of oscillatory activity (Supplementary Figure 1).

The temporal modulation of the beta-power time series revealed rhythmic alignment of beta power to the entraining sounds for the High-Precision Prediction condition at certain electrodes. For example, beta-band power at left (C5, T7) and right (C6) centro-temporal sensors evinced an Event-Related Depression (ERD) following the entraining sound, and a beta power increase (Event-Related Synchronization, ERS) shortly before the next entraining sound (Figure 3). Beta-power time series showed different dynamics depending on the cue (High- vs. Low-Precision Prediction), already evident after the first entraining sound and throughout different time segments of the entire entrainment period (Figure 3).

For the following analysis, we concentrated on the time period shortly before target onset, starting with the last entraining (third) sound and ending with the first possible target sound onset (see Methods for details). This corresponds to the time period of 1.6 - 2.28 s relative to the first entraining sound. Based on previous studies investigating beat perception and temporal prediction (e.g., (31, 41, 47, 48), which reported enhanced beta power, we also expected a beta power increase for the High-Precision Prediction condition in our data. The statistical contrast (one-sided cluster-based permutation test, n = 19) between High- and Low-Precision Prediction revealed that beta power was significantly enhanced for the cue condition predicting High-Perceptual Precision. This relative increase in beta power started earliest approximately 500 ms prior to- and lasted maximally until the earliest possible occurrence of the target sound (range = 1.78 - 2.28 s), entailing extended frontal, central, temporal, and parieto-occipital sites (Figure 3, cluster t-sum = 4823.1, cluster-P = 0.0246). The effect size for the average power values across the cluster (channels and time points) was Cohen's d = 0.81 (large effect).

The beta power time series preceding the target by approximately 800 ms (1.48 - 2.28 s, to avoid evoked responses by early target onsets in some participants, see Methods) were submitted to a spectral decomposition to investigate frequency modulations of beta power over time. The power spectra of the beta power time series showed a peak at approximately 1.6 Hz for both conditions, revealing that the strongest modulation of the beta-power time series was in the delta range (1-3 Hz), and entailing the beat rate (1.25 Hz) of the entraining sound (Figure 3B). Beta power dynamics showed most prominently a significantly stronger modulation at the delta (1-3 Hz) frequency range (Figure 3B) for the High-Precision Prediction Condition (one-sided cluster-based permutation test, n = 19, cluster t-sum = 5765.0, P = 0.0238). However, this modulation was also apparent for the alpha (8-12 Hz) frequency and on some electrodes for the theta (4-7 Hz) frequency range (data not shown). These frequency modulations involved central, frontal, and parietal sites (Figure 3B). The effect size for the average power across the cluster (channels, frequency bins) was Cohen's d = 0.97 (large effect).

Beta band inter-trial phase locking is modulated by temporal precision. Oscillatory phase is an important means for the brain to enable communication between different regions and to modulate excitability of neuronal ensembles to be ready for incoming input (5). Therefore, we were interested in examining if the brain exploits this mechanism for prediction of the perceptual temporal precision of a sound following an entraining beat.

We calculated inter-trial phase coherence (ITPC) for each condition (High- and Low-Precision Prediction) and participant. The statistical comparison was restricted to the time period (2.215 - 2.28 s) and sensors (all 64 sensors except 6 parietal sensors, see Figure 3) that did not show significant beta power differences (see Methods for details). A one-sided dependent-samples cluster-based permutation test on group level revealed significantly enhanced pre-target phase locking shortly before target onset in HP prediction trials, meaning that a HP target was expected (Figure 3). The ITPC effect occurred earliest at 2.22 s and latest at 2.28 s, right before the earliest possible target onset across participants (see Methods) and encompassed the alpha (8-12 Hz) and the beta band (13-30 Hz) ranges (cluster t-sum = 4568.2, P = 0.0128). The effect size for the average power across the cluster (channels, time and frequency bins) was Cohen's d = 1.0115 (large effect). The topography entailed predominantly frontal sites for the alpha and low beta range, and parietal sites for the beta band. Increased phase locking was also visible at central electrodes for the beta band (see Figure 3). Importantly, the ITPC effect vanished towards the end of the trial (Figure 3C), indicating that it was not caused by leakage from evoked responses to the target sounds.

Beta-band power correlates with behavior. We tested neural-behavioral correlations to reveal the behavioral relevance of the identified neural pre-target beta power modulation effects, i.e., the enhancement of beta power for the prediction of high perceptual temporal precision. If the identified beta power modulations reflect underlying neural processes contributing to the prediction of perceptual temporal precision, we expected a positive correlation between temporal deviance detection performance and beta-power t-values obtained from the High- vs. Low-Precision Prediction contrasts at the individual level.

As a first behavioral measure, we selected the sensitivity measure (d-prime) that reflects the performance of the delay detection task in the Valid High-Precision Prediction condition (pooled over On-Time and Delayed). We investigated its relation to the individual beta-power t-values (High- versus Low-Precision Prediction condition) via Spearman correlations, restricting the analysis to the time window that showed a significant group effect for the pre-target beta power (1.8 - 2.28 s, see results above). The group-level test was performed with non-parametric cluster-based permutation tests. It showed a significant correlation (cluster t-sum = 33321.1388, Cluster-P = 0.0407) between the d-prime (Valid High-Precision Prediction) and the beta power t-contrast, starting earliest at 1.83 s and lasting maximally up to 2.28 s, that is, shortly before target onset, with an extended topography over frontal, temporal, parietal, and occipital sites (see Figure 4). The average effect size across the cluster (channels, time and frequency bins) was rho = 0.4982 (medium association). Except for two electrodes, all showed a significant positive correlation between the power t-value (High- versus Low-Precision Prediction contrast) and detection performance (d-prime for the High-Precision Prediction condition). Correlations were predominantly in the beta-band, but also in the alpha-band for some parietal and central electrodes (see Figure 4).

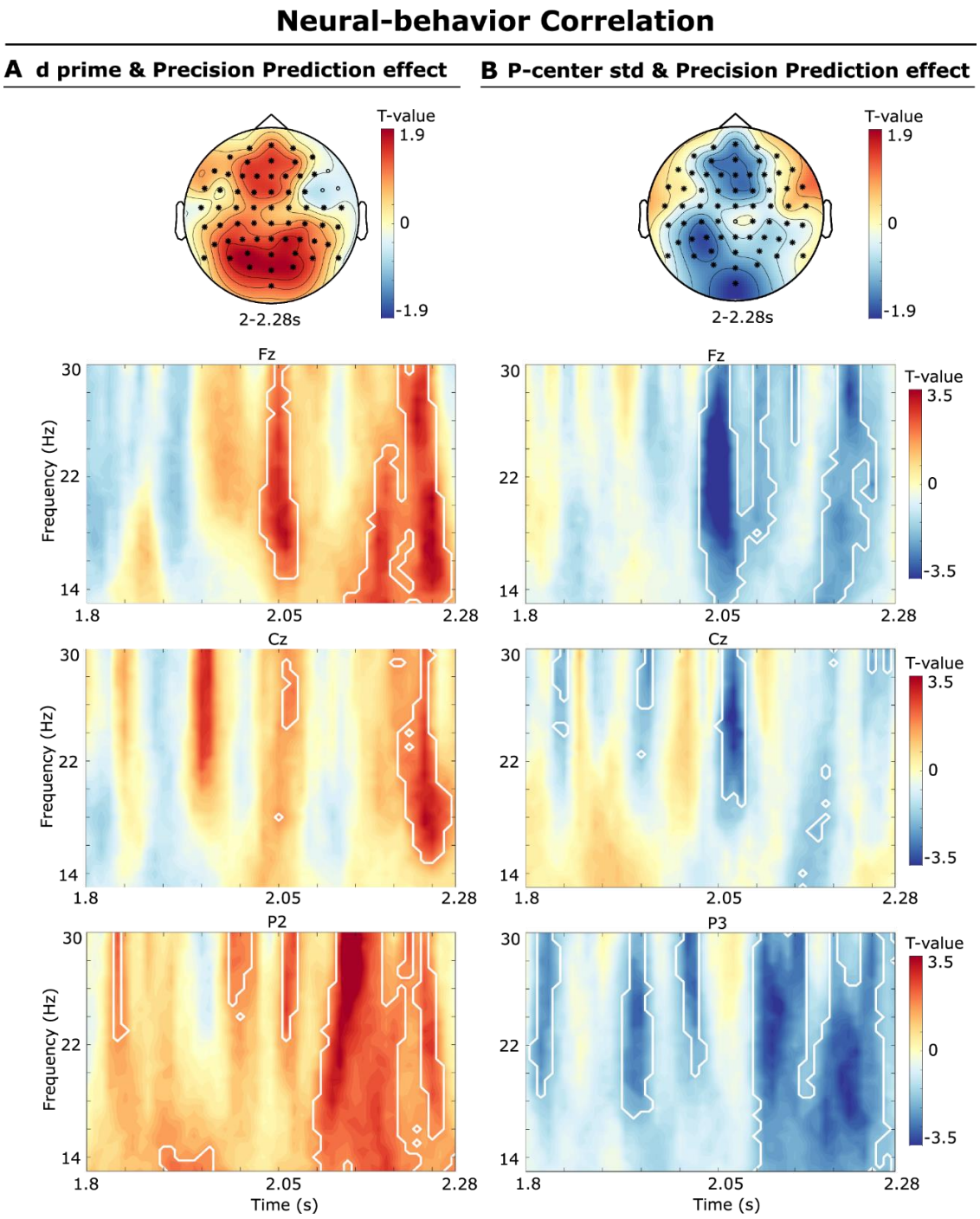

Figure 4. Neural-Behavior Correlation. A) Correlation between the behavioral d-prime measure for the Valid High-Precision Prediction condition and the beta-band contrast between High- and Low-Precision Prediction. There was a positive correlation between the beta-power t-values for the Precision Prediction effect and d-prime for the delay detection task, supporting the link between the neural mechanisms and the behavioral implications of accurately predicting the required perceptual temporal precision. B) Correlation between the behavioral P-center variability measure

for the LP sound (Low-Precision) and the individual t-values for the High- versus Low-Precision Prediction beta power contrast.

As a second behavioral measure, we were interested in the P-center variability, which increases with decreasing precision in the perception of the temporal location of a sound. We expected a negative correlation between P-center variability and the beta-band power t-contrast (High- versus Low-Precision Prediction). As expected, individual differences in P-center variability were most pronounced for the LP sound (see Behavioral Results section). Therefore, we calculated non-parametric Spearman correlations between the individual P-center standard deviation for the LP sound, a measure of individual precision ability, and the individual beta-power t-values for High- versus Low-Precision Prediction. Non-parametric cluster-based permutation tests on the group level revealed significant negative correlations (cluster t-sum = -33901.16, Cluster-P = 0.0338) starting already at 1.9 s and lasting up to 2.3 s, with an extended topography entailing the entire sensor space (Figure 4). The average effect size across the cluster (channels, time and frequency bins) was rho = -0.4858 (medium association). The most pronounced effects were found at posterior- followed by anterior sites.

These results show that the participants who benefitted the most from the HP cue (i.e., larger beta-band power t-contrast values) both performed better on the delay detection task (positive correlation) and were more consistent in estimating the P-center of the LP sound (i.e., lower P-center variability; negative correlation).

Discussion

We investigated neural processes underlying the prediction of the temporal precision of auditory events during beat perception. Flexibility in perceptual temporal precision allows us to perceive regularity in acoustic information that does not show a strictly periodic pattern or has a shape that makes its perceived temporal position indistinct, as can be the case in language (16, 17, 19, 22) or music (6, 14). Although neural entrainment is often associated with strictly isochronous stimuli, the underlying neural mechanisms seem to tolerate a certain amount of temporal variability (5). Such tolerance for temporal variability is also more adaptive, given that most environmental events do not follow strict isochrony. Nevertheless, studies of neural entrainment that directly consider the variability in perceptual temporal precision caused by different acoustic features are lacking. With the current study, we intended to close this gap.

We hypothesized that beta activity entrains with the isochronous beat, encodes the expected temporal precision of upcoming auditory events, and, importantly, that this prediction process is under top-down control. We designed a delay detection task, where we utilized cues to manipulate expectations about the acoustic shape and, consequently, the temporal precision of an upcoming target sound following a short beat sequence. The results showed that performance was affected by the cue, giving clear evidence that the perceptual temporal precision indicated by the cue was behaviorally relevant. At the neural level, the results replicated previous findings that beta-band oscillations entrain with a regular beat (1, 30, 31) and encode temporal predictions (29, 41, 49). Importantly, we revealed that this entrainment process is modulated by active top-down control, enabling the neural system to flexibly adapt the temporal precision of sensory predictions in accordance with context- and task-specific demands, thereby enhancing sensory selection.

We used acoustic features to manipulate the perceptual temporal precision of a sound. To ensure that the acoustic features had the intended effect on temporal perception, we measured the P-Center location and variability, which reflect how precisely a listener locates a sound in time. The

results of previous studies were replicated (6, 40) as the sound inducing a high temporal precision (HP sound with short rise time, short duration, and high frequency) had a significantly lower perceptual variability compared to the sound inducing a low temporal precision (LP sound with long rise time, long duration, and low frequency). These results were important to verify that the target sounds affected perceptual temporal precision in the intended direction.

Furthermore, the relevance of the pre-target cue for task performance was a prerequisite for our argument that the observed neural entrainment is under top-down control. Temporal deviance detection was modulated by cue validity, but only for target sounds with acoustic features inducing a high perceptual temporal precision. This means that expecting a target sound with low temporal precision (LP sound) was detrimental for delay detection performance when the actual target required a high perceptual temporal precision (HP sound). In contrast, cue validity did not significantly affect behavioral performance for target sounds with low perceptual temporal precision. That is, a cue that erroneously signaled a target with high temporal precision was not detrimental for the subsequent temporal processing of a target sound with low temporal precision. These results support the assumption that the crucial element for the top-down prediction was different levels of perceptual temporal precision and not, for example, different acoustic features of the target sounds (e.g., frequency, duration) or different onset times (see also Limitations and Future Research).

Altogether, our results show that the neural system exploited the information conveyed by the cue to predict the temporal precision of sounds in a top-down manner while entraining to a regular beat. Interestingly, the post-experiment self-report (open questionnaire) contradicted the behavioral and neurophysiological results. The majority of the participants (79%) stated that they did not use the cue to solve the delay detection task. This indicates that most participants were not consciously aware that the perceptual system uses the cue information to predict the acoustic shape of the target and the needed perceptual temporal precision for performing the task. On the other hand, most people (68%) reported that it was easier to judge the delay of the HP target sound than the LP target sound. This is interesting, since it coincides with the behavioral advantage of the cue, which was only present for the high-precision target.

The performance enhancement by the valid cue is a pivotal finding, because the ongoing debate about whether oscillatory entrainment is instrumental to temporal prediction revolves around the central question of whether it can be actively modulated in a top-down manner. Growing evidence shows that rhythmic stimuli with a regular structure facilitates temporal processing and enables behavioral benefits due to temporal anticipation (2, 3, 5, 25, 26, 37, 50). Neural oscillations reflect rhythmic fluctuations of neural excitability and are, therefore, excellent candidates for a mechanism underlying temporal prediction via the phase alignment of slow oscillations to external events (2, 3, 5, 25, 26, 37). This also makes oscillatory entrainment an elegant neurophysiological implementation of the dynamic attending theory (5, 35, 50). However, phase-synchronization could be caused via bottom-up evoked responses to the regular stimuli, simply reflecting the periodicity of the events (16, 37). This renders it difficult to determine if neural entrainment is functionally relevant to sensory *selection*. Therefore, a key question is whether top-down factors can actively control neural entrainment (35, 37).

In an effort to synthesize current debates on temporal predictions, Rimmele et al. (4) argued for a neuronal representation that enables a simultaneous bottom-up (stimulus-driven) and top-down (motor- or language-driven) phase reset of low-frequency oscillations in the auditory cortex, which is carried by beta oscillations. Similarly, in light of the active sensing framework, it has been argued that a top-down predictive phase-reset of low-frequency (e.g., delta) oscillations, i.e., an anticipatory phase reset, could be implemented via the motor system encoding temporal predictions in beta-band activity (1, 4, 5, 32). In the same vein, Cannon and Patel (51) proposed a causal role of the motor system in beat-based temporal predictions and suggested that the prediction of beat timing is communicated through the dorsal auditory pathway connecting premotor, parietal, and temporal cortices.

The results of the current study provide support for the active sensing framework, since there is clear evidence that temporal predictions related to beat precision are transmitted via beta activity in a top-down manner. Beta power was relatively enhanced if the participant expected a target sound with high compared to low temporal precision, already several hundred milliseconds before the target onset. This is in line with studies indicating that temporal predictions are carried out by beta oscillations (4, 29, 31, 47–49, 49). Relatedly, Arnal et al. (41) reported that pre-target beta power was enhanced for correctly detected targets in a temporal delay detection task, along with a significantly enhanced cross-frequency coupling between pre-target delta phase and beta power.

In addition, brain-behavior correlations indicated that the participants who showed a higher beta power modulation for the prediction of the temporal precision (i.e., larger beta power t-contrast values for HP versus LP cue trials) both performed better on the delay detection task (i.e., larger d-prime; positive correlation) and were more precise at estimating the P-center of the LP sound (i.e., lower P-center variability; negative correlation). This provides important support for the behavioral relevance of beta power for predicting beat timing and temporal precision of sounds in a beat sequence.

Moreover, beta-band activity was modulated strongest at the delta range (1-3 Hz), which entails the frequency of the entraining isochronous sounds (1.25 Hz) presented prior to the target. Importantly, this spectral modulation was relatively enhanced for the High- compared to the Low-Precision Prediction condition in the delta and theta (4-7 Hz) frequency ranges during the last beat period before the target. These results are in line with the comodulation of delta and beta oscillations that carry temporal predictions found by Morrillon et al. (29). Morillon et al. (28, 29) showed that top-down influences via the motor system sharpen sensory processing by encoding temporal predictions via beta oscillations. Beta oscillations were directed towards sensory regions and coupled to delta oscillations. The authors argued for a central role of beta oscillations in representing temporal information and encoding top-down predictions, realized via a covert form of active sensing with motor areas providing contextual information to sensory regions (29). In the present study, we demonstrated that beta oscillations also convey information about the predicted temporal precision of a sound in a beat sequence in a top-down manner.

In support of the active sensing model, beat perception has been shown to engage the motor cortico-basal ganglia-thalamo-cortical (mCBGT) circuit, with the supplementary motor area (SMA), dorsal striatum and the putamen as important nodes (51). This is complemented by studies examining single-cell and microcircuit activity in macaques (1). The idea of the involvement of the motor system fits nicely with the topography of the beta power results in the current study, which encompassed, next to fronto-central, predominantly central and parietal regions of the scalp. However, since we did not source localize the effect, the interpretation of the scalp topography remains speculative. Nevertheless, beta oscillations are prominent in the motor system (52), and our findings concur with the theoretical framework of active perceptual inference. In addition, the post-experiment self-report also demonstrated an urge to move to the beat to solve the task in at least half of the participants.

Studies investigating rhythm-based and cue-based temporal expectations have shown functional dissociations in behavior and provided evidence that both processes might rely on distinct brain circuits (46, 53, 54). In the current study, we did not intend to dissociate between cue-based and rhythm-based temporal predictions. Accordingly, the task design does not allow for an investigation of a possible dissociation between these cognitive mechanisms. Our goal was to modulate the beat-based entrainment process via a cue in a top-down fashion. Still, multiple distinct mechanisms might have been at play and acted in synergy creating the behavioral benefit of the cue in a task that involves rhythmic expectations.

The results could also be interpreted in the light of the Dynamic Attending Theory (24, 50), suggesting that attentional processes are dynamic and align to the temporal patterns of external events. Large and Jones (24) showed how external rhythms might entrain neural slow oscillations by aligning the phase to optimal time points to direct selective attention. In their computational

model, the attentional focus increases with certainty and decreases or widens with more variability (uncertainty) in an external rhythm. Jones suggested that in addition to phase, the *amplitude* of induced beta is coupled to the phase of the slow oscillations, representing attending pulses that precede the events (55). Similarly, we found a modulation of beta activity depending on the perceptual temporal precision of the target sound. However, since our paradigm precludes investigation of the phase of slow entrained oscillations due to the evoked responses at the entrainment rate, we cannot directly test predictions by DAT related to phase synchronization of delta oscillations.

We could, however, investigate if the phase-locking of the beta band serves top-down predictions about the perceptual temporal precision. The anticipation of a high-precision target sound was accompanied by enhanced phase coupling in the alpha band (8-12 Hz) and beta band shortly before target onset. This is in line with a study by Grabot et al. (56), which found increased alpha-beta phase-amplitude coupling for increased precision in a temporal production task. Since phase coherence (57) and beta oscillations (58) enable communication between different brain regions, these mechanisms might be involved in the prediction of beat timing and the transfer of beat timing information between brain areas.

In a similar vein, Large and Snyder (59) proposed that beta bursts that entrain to a rhythm might reflect the information transfer between motor and auditory cortices during rhythm perception. A proposed scenario is a phase reset being implemented by the corollary discharge from the motor command that generates a sensory sampling action, informing the sensory region about the temporal arrival of the sensory input (5). The phase to which the oscillations are reset/entrained is likely conveying predictions about anticipated features of the external object, similar to entrainment during selective attention (36, 37). This is in line with the active sensing framework and neural entrainment being instantiated as oscillatory bursts in the beta band via rhythmic sampling in the motor system (1, 5, 33, 59).

Limitations and Future Research. The results of the current study add to the evidence provided by studies of beat perception (1, 3, 29–31, 34), supporting that beta oscillations are involved in the neural entrainment to a regular beat, carrying temporal predictions about the next event in a beat sequence. Fujioka et al. (30) source localized beta activity to the auditory cortex and found that the beta event-related synchronization (ERS, power increase shortly before the sound onset) represented the predicted timing of the next sound in a beat sequence.

Our results support this view by showing that the predicted perceptual temporal precision (or shape) of the target sound is encoded by a beta power increase shortly before the expected target sound in a beat sequence. We did not perform a source localization of the beta activity, the analysis of the ERS and event-related desynchronisation (ERD) in relation to the entraining sounds (0 - 1.16 s), as this was beyond the scope of this paper. If the listener expected a high-precision target, then beta-power time series seemed to be aligned to the entraining sounds for some electrodes, with a power increase before the next entraining sounds, and a power decrease after the sound, resembling the ERS and ERD described by Fujioka et al. (30, 31). However, for the condition inducing the prediction of a low-precision target, the alignment of the beta activity was quite different, with a beta power peak halfway between the first and the second, and between the second and the third entraining sound (around 0.4 s and 1.2 s) for several anterior sensors. This is a highly interesting observation and should be further investigated in a future study conducting EEG source analysis.

Another limitation of the study might be that targets not only differed in P-center variability but also in P-center location and duration. The target sounds were P-center aligned to the entraining sounds. The later the P-center location, the earlier the presentation of the sound in order to P-center align it with the previous entraining sounds.

Hence, the neural differences observed for various levels of predicted temporal precision, might be confounded by different expectations regarding the sound onset and duration of the sound. However, pre-target beta power was increased for the High-Precision Prediction, which generally had later sound onsets than the Low-Precision Condition, due to the P-center alignment (LP sound had to be presented earlier to compensate for the relatively late P-center location). Therefore, this explanation seems unlikely. Also, the behavioral effect indicating detrimental effects only for the target sound with a high temporal precision argues against this explanation. We thus believe that the pre-target modulations in beta activity were caused by different predictions of temporal precision of the target sound. In a future study, one could possibly control for the duration of the stimuli, but only change the rise time of the sounds in order to manipulate P-center variability, to rule out this as a possible confounding factor.

Conclusions

Altogether, the results of the current study revealed that beta-band oscillations subserve neural mechanisms responsible for top-down predictions regarding the temporal precision of upcoming auditory events.

Employing a cued delay detection task, we found that beta-band power and phase consistency were relatively enhanced if a target sound with a high (versus a low) perceptual temporal precision was expected. In addition, the spectral modulation of beta-band activity was enhanced in the delta range during the last beat period, entailing the beat frequency (1.25 Hz) of the entraining isochronous sounds. The magnitude of the beta power difference between high versus low perceptual temporal precision correlated positively with task performance. Moreover, the same measure of beta activity correlated negatively with the perceptual variability (i.e., more consistent P-center estimation) of the perceived temporal occurrence of a sound.

Importantly, the results revealed that beta activity is not only involved in the processing of strictly isochronous events, but also quasi-periodic patterns, which predominate in language, and many musical styles (60–64). The study provides unprecedented insight into how neural entrainment via beta bursts can be under active top-down control. Our findings are in line with the active sensing framework (4, 5, 34) and studies relating beta activity with beat perception and temporal predictions (1, 31, 41).

Acknowledgments. We thank Sandra Solli and Olga Asko for fruitful discussions and the volunteers for participating in this study.

Funding. This work was partially supported by the Research Council of Norway, project numbers 240389 and 314925, through its Centres of Excellence scheme (project number 262762) and the TIME project (Grant 249817).

Data availability. Data and all analysis algorithms data will be made available in an online repository upon publication of the paper.

Competing Interest Statement: The authors declare no conflict of interest.

References

1. H. Merchant, J. Grahn, L. Trainor, M. Rohrmeier, W. T. Fitch, Finding the beat: a neural perspective across humans and non-human primates. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140093 (2015).
2. J. Obleser, C. Kayser, Neural Entrainment and Attentional Selection in the Listening Brain. *Trends Cogn. Sci.* **23**, 913–926 (2019).
3. K. Doelling, S. Herbst, L. Arnal, V. van Wassenhove, Psychological and neuroscientific foundations of rhythms and timing (2023) https:/doi.org/10.31234/osf.io/hkgdr (June 19, 2023).
4. J. M. Rimmele, B. Morillon, D. Poeppel, L. H. Arnal, Proactive Sensing of Periodic and Aperiodic Auditory Patterns. *Trends Cogn. Sci.* **22**, 870–882 (2018).
5. P. Lakatos, J. Gross, G. Thut, A New Unifying Account of the Roles of Neuronal Entrainment. *Curr. Biol.* **29**, R890–R905 (2019).
6. A. Danielsen, *et al.*, Where is the beat in that note? Effects of attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds. *J. Exp. Psychol. Hum. Percept. Perform.* **45**, 402–418 (2019).
7. J. London, *et al.*, A comparison of methods for investigating the perceptual center of musical sounds. *Atten. Percept. Psychophys.* **81**, 2088–2101 (2019).
8. A. Danielsen, *et al.*, Sounds familiar(?): Expertise with specific musical genres modulates timing perception and micro-level synchronization to auditory stimuli. *Atten. Percept. Psychophys.* **84**, 599–615 (2022).
9. J. Morton, S. Marcus, C. Frankish, Perceptual centers (P-centers). *Psychol. Rev.* **83**, 405–408 (1976).
10. J. Vos, R. Rasch, The perceptual onset of musical tones. *Percept. Psychophys.* **29**, 323–335 (1981).
11. J. W. Gordon, The perceptual attack time of musical tones. *J. Acoust. Soc. Am.* **82**, 88–105 (1987).
12. R. Villing, “Hearing the Moment: Measures and Models of the Perceptual Centre,” National University of Ireland Maynooth. (2010) (July 10, 2023).
13. A. Danielsen, “Pulse as Dynamic Attending: Analysing Beat Bin Metre in Neo Soul Grooves” in *The Routledge Companion to Popular Music Analysis*, (Routledge, 2018).
14. A. Danielsen, *Musical Rhythm in the Age of Digital Reproduction* (Ashgate Publishing, Ltd., 2010).
15. M. S. Johansson, Making Sense of Genre and Style in the Age of Transcultural Reproduction. *Int. Rev. Aesthet. Sociol. Music* **47**, 45–62 (2016).
16. B. Zoefel, R. VanRullen, EEG oscillations entrain their phase to high-level features of speech sound. *NeuroImage* **124**, 16–23 (2016).
17. J. Gross, *et al.*, Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLOS Biol.* **11**, e1001752 (2013).
18. G. Buzsáki, M. Vöröslakos, Brain rhythms have come of age. *Neuron* **111**, 922–926 (2023).
19. F. Schmidt, *et al.*, Neural speech tracking shifts from the syllabic to the modulation rate of speech as intelligibility decreases. *Psychophysiology* **n/a**, e14362.
20. K. B. Doelling, M. F. Assaneo, D. Bevilacqua, B. Pesaran, D. Poeppel, An oscillator model better predicts cortical entrainment to music. *Proc. Natl. Acad. Sci.* **116**, 10113–10121 (2019).
21. K. B. Doelling, L. H. Arnal, O. Ghitza, D. Poeppel, Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage* **85**, 761–768 (2014).
22. D. Poeppel, M. F. Assaneo, Speech rhythms and their neural foundations. *Nat. Rev. Neurosci.* **21**, 322–334 (2020).
23. K. B. Doelling, M. F. Assaneo, Neural oscillations are a start toward understanding brain activity rather than the end. *PLOS Biol.* **19**, e3001234 (2021).

24. E. W. Large, M. R. Jones, The dynamics of attending: How people track time-varying events. *Psychol. Rev.* **106**, 119–159 (1999).
25. M. J. Henry, J. Obleser, Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci.* **109**, 20095–20100 (2012).
26. M. J. Henry, B. Herrmann, J. Obleser, Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc. Natl. Acad. Sci.* **111**, 14935–14940 (2014).
27. C. E. Schroeder, D. A. Wilson, T. Radman, H. Scharfman, P. Lakatos, Dynamics of Active Sensing and perceptual selection. *Curr. Opin. Neurobiol.* **20**, 172–176 (2010).
28. B. Morillon, C. E. Schroeder, V. Wyart, Motor contributions to the temporal precision of auditory attention. *Nat. Commun.* **5**, 5255 (2014).
29. B. Morillon, S. Baillet, Motor origin of temporal predictions in auditory attention. *Proc. Natl. Acad. Sci.* **114**, E8913–E8921 (2017).
30. T. Fujioka, L. J. Trainor, E. W. Large, B. Ross, Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations. *J. Neurosci.* **32**, 1791–1802 (2012).
31. T. Fujioka, B. Ross, L. J. Trainor, Beta-Band Oscillations Represent Auditory Beat and Its Metrical Hierarchy in Perception and Imagery. *J. Neurosci.* **35**, 15187–15198 (2015).
32. C. C. Canavier, Phase-resetting as a tool of information transmission. *Curr. Opin. Neurobiol.* **31**, 206–213 (2015).
33. S. Sadaghiani, M. J. Brookes, S. Baillet, Connectomics of human electrophysiology. *NeuroImage* **247**, 118788 (2022).
34. B. Morillon, L. H. Arnal, C. E. Schroeder, A. Keitel, Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neurosci. Biobehav. Rev.* **107**, 136–142 (2019).
35. S. Haegens, E. Zion Golumbic, Rhythmic facilitation of sensory processing: A critical review. *Neurosci. Biobehav. Rev.* **86**, 150–165 (2018).
36. P. Lakatos, *et al.*, The spectrotemporal filter mechanism of auditory selective attention. *Neuron* **77**, 750–761 (2013).
37. N. Suess, T. Hartmann, N. Weisz, Differential attention-dependent adjustment of frequency, power and phase in primary sensory and frontoparietal areas. *Cortex* **137**, 179–193 (2021).
38. F. L. Bouwer, Neural Entrainment to Auditory Rhythms: Automatic or Top-Down Driven? *J. Neurosci.* **42**, 2146–2148 (2022).
39. L. Meyer, Y. Sun, A. E. Martin, Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. *Lang. Cogn. Neurosci.* **35**, 1089–1099 (2020).
40. G. S. Câmara, K. Nymoen, O. Lartillot, A. Danielsen, Effects of instructed timing on electric guitar and bass sound in groove performance. *J. Acoust. Soc. Am.* **147**, 1028–1041 (2020).
41. L. H. Arnal, K. B. Doelling, D. Poeppel, Delta–Beta Coupled Oscillations Underlie Temporal Prediction Accuracy. *Cereb. Cortex* **25**, 3077–3085 (2015).
42. S. Kulashekhar, J. Pekkola, J. M. Palva, S. Palva, The role of cortical beta oscillations in time estimation. *Hum. Brain Mapp.* **37**, 3262–3281 (2016).
43. M. I. Posner, C. R. Snyder, B. J. Davidson, Attention and the detection of signals. *J. Exp. Psychol. Gen.* **109**, 160–174 (1980).
44. E. J. Golob, H. Pratt, A. Starr, Preparatory slow potentials and event-related potentials in an auditory cued attention task. *Clin. Neurophysiol.* **113**, 1544–1557 (2002).
45. F. van Ede, F. P. de Lange, E. Maris, Attentional Cues Affect Accuracy and Reaction Time via Different Cognitive and Neural Processes. *J. Neurosci.* **32**, 10408–10412 (2012).
46. A. C. Nobre, F. van Ede, Attention in flux. *Neuron* **111**, 971–986 (2023).
47. D. Müllensiefen, B. Gingras, J. Musil, L. Stewart, The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE* **9**, e89642 (2014).
48. H. Stanislaw, N. Todorov, Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* **31**, 137–149 (1999).

49. R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 1:1-1:9 (2011).
50. M. X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice* (MIT Press, 2014).
51. R. M. van Diepen, A. Mazaheri, The Caveats of observing Inter-Trial Phase-Coherence in Cognitive Neuroscience. *Sci. Rep.* **8**, 2990 (2018).
52. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
53. M. Saleh, J. Reimer, R. Penn, C. L. Ojakangas, N. G. Hatsopoulos, Fast and Slow Oscillations in Human Primary Motor Cortex Predict Oncoming Behaviorally Relevant Cues. *Neuron* **65**, 461–471 (2010).
54. D. Lakens, Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* **4** (2013).
55. F. O'Brien, D. Cousineau, Representing Error bars in within-subject designs in typical software packages. *Quant. Methods Psychol.* **10**, 56–67 (2014).
56. B. Bechtold, Violin Plots for Matlab (2023) (July 20, 2023).
57. J. R. Iversen, B. H. Repp, A. D. Patel, Top-Down Control of Rhythm Perception Modulates Early Auditory Responses. *Ann. N. Y. Acad. Sci.* **1169**, 58–73 (2009).
58. M. D. Foldal, S. Leske, A. O. Blenkmann, T. Endestad, A.-K. Solbakk, Attentional modulation of beta-power aligns with the timing of behaviorally relevant rhythmic sounds. *Cereb. Cortex* **33**, 1876–1894 (2023).
59. M. R. Jones, Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychol. Rev.* **83**, 323–355 (1976).
60. J. J. Cannon, A. D. Patel, How Beat Perception Co-opts Motor Neurophysiology. *Trends Cogn. Sci.* **25**, 137–150 (2021).
61. G. Pfurtscheller, F. H. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* **110**, 1842–1857 (1999).
62. A. Breska, R. B. Ivry, Context-specific control over the neural dynamics of temporal attention by the human cerebellum. *Sci. Adv.* **6**, eabb1141 (2020).
63. A. C. Nobre, F. van Ede, Anticipated moments: temporal structure in attention. *Nat. Rev. Neurosci.* **19**, 34–48 (2018).
64. M. R. Jones, *Time Will Tell: A Theory of Dynamic Attending* (Oxford University Press, 2018).
65. L. Grabot, *et al.*, The Strength of Alpha–Beta Oscillatory Coupling Predicts Motor Timing Precision. *J. Neurosci.* **39**, 3277–3291 (2019).
66. P. Sauseng, W. Klimesch, What does phase information of oscillatory brain activity tell us about cognitive processes? *Neurosci. Biobehav. Rev.* **32**, 1001–1013 (2008).
67. N. Kopell, G. B. Ermentrout, M. A. Whittington, R. D. Traub, Gamma rhythms and beta rhythms have different synchronization properties. *Proc. Natl. Acad. Sci.* **97**, 1867–1872 (2000).
68. E. W. Large, J. S. Snyder, Pulse and Meter as Neural Resonance. *Ann. N. Y. Acad. Sci.* **1169**, 46–57 (2009).
69. R. Polak, *et al.*, Rhythmic Prototypes Across Cultures: A Comparative Study of Tapping Synchronization. *Music Percept.* **36**, 1–23 (2018).
70. M. Johansson, Non-Isochronous Musical Meters: Towards a Multidimensional Model. *Ethnomusicology* **61**, 31–51 (2017).
71. T. Kvifte, Categories and Timing: On the Perception of Meter. *Ethnomusicology* **51**, 64–84 (2007).
72. M. R. Haugen, A. Danielsen, Effect of tempo on relative note durations in a performed samba groove. *J. New Music Res.* **49**, 349–361 (2020).
73. E. F. Clarke, Some Aspects of Rhythm and Expression in Performances of Erik Satie's Gnossienne No. 5. *Music Percept.* **2**, 299–328 (1985).