Implicit temporal predictability biases slow oscillatory phase in auditory cortex and enhances pitch discrimination sensitivity

Sophie K. Herbst[1,2] and Jonas Obleser[1]

[1]Department of Psychology, University of Lübeck Ratzeburger Allee 160, 23552 Lübeck, Germany
[2]NeuroSpin, CEA, DRF/Joliot, INSERM Cognitive Neuroimaging Unit, Bât 145, Gif s/ Yvette 91190, France

Abstract

Can human listeners use strictly implicit temporal contingencies in auditory input to form temporal predictions, and if so, how are these predictions represented endogenously? To assess this question, we implicitly manipulated foreperiods in an auditory pitch discrimination task. Unbeknownst to participants, the pitch of the standard tone could either be deterministically predictive of the onset of the target tone, or convey no predictive information. Both conditions were presented interleaved in one stream, and separated by variable inter-stimulus intervals such that there was no dominant stimulus rhythm throughout. Even though participants were unaware of the implicit temporal contingencies, pitch discrimination sensitivity (i.e. the slope of the psychometric function) increased when the onset of the target tone was predictable in time (N=49). Concurrently recorded EEG data (N=24) revealed that standard tones which initiated temporal predictions evoked a more negative N1 component than non-predictive standards, and were followed by an increase in delta power during the foreperiod. Furthermore, the phase angle of delta oscillations (1–3Hz) evoked by the standard tone predicted pitch discrimination sensitivity at the target tone (1.75 s later on average), which suggests that temporal predictions can be initiated by an optimized delta phase reset. In sum, we show that auditory perception benefits from implicit temporal contingencies, and provide evidence for a role of slow neural oscillations in the endogenous representation of temporal predictions, in absence of exogenously driven entrainment to rhythmic input.

Significance Statement: Temporal contingencies are ubiquitous in sensory environments, especially in the auditory domain, and have been shown to facilitate perception and action. Yet, how these contingencies in exogenous inputs are transformed into an endogenous representation of temporal predictions is not known. Here, we implicit induced temporal predictability in the absence of a rhythmic input structure, that is without exogenously driven entrainment of neural oscillations. Our results show that even implicit and non-rhythmic temporal predictions are extracted and used by human observers, underlining the role of timing in sensory processing. Furthermore, our EEG results point towards an instrumental role of delta oscillations in initiating temporal predictions by an optimized phase reset in response to a temporally predictive cue.

Acknowledgments: This research was supported by a DFG grant (HE 7520/1-1) to SKH. The authors would like to thank Anne Herrmann for overseeing the data acquisition, Michael Ploechl for technical support, and Virginie van Wassenhove and the Cognition & Brain Dynamics Group at NeuroSpin for helpful discussions.

Conflict of interest: None.

Number of Pages: 24
Number of Figures: 4
Number of Tables: 0
Number of Words: Abstract: 232, Significance Statement: 116, Introduction: 556, Discussion: 1618

Introduction

The human brain is constantly engaged in forming predictions about its environment, concerning the *where*, the *what*, and crucially also the *when* of future events. These temporal aspects of predictions have only recently become the focus of dedicated investigation (Coull and Nobre, 1998; Nobre et al., 2007; Nobre and van Ede, 2018; Rimmele et al., 2018), contrary to spatial or content-based ones. That discrepancy could be explained by the fact that temporal aspects are inherent to any kind of input and are thus difficult to isolate experimentally.

Implicit temporal statistics of visual and auditory input are extracted by the human cognitive system, and benefit perception and action (Cravo et al., 2011; Herbst and Obleser, 2017). In the auditory domain, rhythmic input structure has been shown to improve detection performance and speed (Henry and Obleser, 2012; Lawrance et al., 2014; Rimmele et al., 2011; Stefanics et al., 2010; Wright and Fitzgerald, 2004). Fewer studies have shown that rhythmic temporal regularities can also improve perceptual sensitivity (i.e. discrimination performance) in the auditory (Jones et al., 2002; Schmidt-Kassow et al., 2009; but see Bauer et al., 2015), as well as the visual domain (Cravo et al., 2013; Rohenkohl et al., 2012).

In order to study implicit timing in the absence of a direct rhythmic input structure, we here induced temporal predictability by repeating single predictive intervals in a so-called foreperiod paradigm (Niemi and Näätänen, 1981; Woodrow, 1914). This type of manipulation has been shown to increase visual perceptual sensitivity (Correa et al., 2004, 2005; Cravo et al., 2011; Rolke and Hofmann, 2007). In audition, predictable foreperiods have been found to speed up stimulus processing (Bausenhart et al., 2007) and improve short-term memory performance (Wilsch et al., 2018, 2014). To our knowledge, no study has shown an effect of implicit non-rhythmic temporal predictability on perceptual sensitivity in the auditory domain.

How temporal contingencies present in exogenous inputs are transformed into an endogenous representation of temporal predictability is not well known (van Wassenhove, 2016). Slow neural oscillations (in the delta/1–3 Hz and theta/4–7 Hz frequency bands) have been identified as a prime candidate mechanism for the implementation of temporal prediction at the neural level (Arnal et al., 2014; Besle et al., 2011; Gomez-Ramirez et al., 2011; Herrmann et al., 2016; Kösem et al., 2014; Lakatos et al., 2008; Morillon and Schroeder, 2015; Rimmele et al., 2018; Wilsch et al., 2015) via endogenous phase-resetting and -shifting mechanisms. This mechanism is expected to lead to enhanced phase coherence of slow oscillations in anticipation of temporally predictive input (Breska and Deouell, 2017; Cravo et al., 2013; Henry et al., 2016; Herrmann et al., 2016; Wilsch et al., 2015). However, it is not trivial to disentangle an internalized and endogenously activated representation of temporal prediction from entrainment in the sense of mechanistic input-driven alignment of neural activity to rhythmic input (Breska and Deouell, 2017; Cravo et al., 2011; Haegens and Golumbic, 2017; Obleser et al., 2017).

Here, we set out to investigate the role of neural oscillatory dynamics for an endogenous representation of temporal predictions from auditory inputs, in the absence of direct local temporal structures. To this end, we used a deliberately non-rhythmic foreperiod paradigm in which the rele-

vant intervals have to be internalized by the auditory system (i.e., extracted, storeed, and activated) to actively generate a prediction when a new trial is presented (Haegens and Golumbic, 2017). We implicitly associated temporal predictability to a sensory feature of the standard tone in an auditory pitch discrimination task: the standard's pitch could be deterministically predictive of the time of the *onset* (but not the pitch) of the target tone, or convey no predictive information. Both conditions were presented interleaved in one stream, and separated by variable inter-stimulus intervals such that there was no dominant stimulus rhythm throughout.

We show that, in behavioural terms, temporal predictability increases pitch discrimination sensitivity, assessed via the slope of the psychometric function. Concurrently recorded EEG data provide indices of temporally predictive processing in auditory cortex evoked by both the standard and target tone. Furthermore, we show enhanced delta power in the predictive compared to the non-predictive condition, and (by applying an auditory spatial filter) a predictive relationship between delta phase angle in auditory areas evoked by the temporal cue and pitch discrimination performance. Together, these results suggest an instrumental role of delta oscillations in forming temporal predictions

Methods

Participants

In total, 51 participants were tested (23.6 years on average (SD = 3.5), 28 female, 6 left handed), 26 of which also underwent electroencephalography (EEG). All participants signed informed consent and received either course credit or payment for their participation (8 € per hour). The study was approved by the local ethics committee at the University of Lübeck. From the behavioural sample, we excluded two participants (who were not part of the EEG sample) because of ceiling effects (their slopes for the psychometric function in one of the two conditions exceeded the mean of the slope distributions of all participants by more than 2.5 standard deviations). From the EEG sample we excluded two participants who had blinked in synchrony with the auditory stimulation and for whom we were not able to separate blinks from the auditory evoked potentials during EEG preprocessing. The behavioural data of these two participants were kept in the analyses.

Stimuli and Procedure

The experiment was conducted in an electrically shielded sound-attenuated EEG booth. Stimulus presentation and collection of behavioural responses was achieved using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) under Windows 7. Responses were collected on a standard keyboard. All participants were instructed to use the index and middle fingers of the right hand.

Participants performed a pitch discrimination task, comparing tone pairs embedded in noise, as illustrated in Figure 1A. A black fixation cross was displayed on gray background throughout the whole block. Auditory stimuli were delivered via headphones (Sennheiser HD 25-SP II) at 50 dB

above the individual's sensation level, predetermined for the noise using the method of limits. Low-pass (5kHz) filtered white noise was presented constantly throughout each block. Pure tones of varying frequencies (duration 50 ms with a 10 ms on- and offset ramp), were presented with a tone-to-noise ratio fixed at −18 dB. The first tone, to which we will refer as the *standard* in the following was always at one of two frequencies: 550 or 950 Hz. The second tone, the *target*, was varied in individually predetermined steps around its respective standard. The same step size was used for both standards, but logarithmically transformed and multiplied with the standard frequency, to obtain a log-spaced frequency scale around each standard. To predetermine the step size, each participant was first presented with one experimental block to familiarize themselves with the task. Then, a second block was performed, and if performance was below 65%, the tone-steps were increased, which was repeated up to three times. All participants reached the minimum performance level after minimally two and maximally four rounds of training. As a result of this procedure, the average lowest target tone presented with the 550 Hz standard was 508.3 Hz (range 490.0–519.1 Hz), and the highest target tone 595.3 Hz (range 582.7–617.4 Hz); the lowest target tone presented with the 950 Hz standard was 878.0 Hz (range 846.4–896.7 Hz), and the highest target tone 1028.3 Hz (range 1006.5–1066.3 Hz). The high and low tones never overlapped. In the behavioural experiment, eleven tone frequencies were used from the lowest to highest tone, including the standard; in the EEG experiment we used 7 discrete frequencies. Participants were instructed to indicate after each tone pair whether the second tone was lower or higher than the first. After the target tone, participants had 2 s to respond. The stimulation continued automatically, even if no response was given.

Critically, and unbeknownst to participants, we manipulated the interval between standard and target tones, the *foreperiod*, by either pseudo-randomly drawing foreperiods from a discretized uniform foreperiod duration (ranging from 0.5–3 s, blue distribution in Figure 1 A), or used the same foreperiod duration (1.75 s, green distribution in Figure 1 A). This resulted in one condition in which the target onset was perfectly predictable in time, the *predictive condition*, and one condition in which the target onset was maximally jittered, the *non-predictive condition*. To allow participants to implicitly dissociate the conditions, the foreperiod distributions were associated with one of the standard pitches, for example for one participant the 550 Hz standard was always followed by a predictive foreperiod and the 950 Hz standard was always followed by a non-predictive foreperiod. The assignment was counterbalanced over participants. The two conditions were presented interleaved, such that participants had to encode the standard pitch on each trial. Importantly, the manipulation of foreperiod intervals was strictly implicit, and participants were not informed about it.

To avoid build-up of a rhythm over trials, the inter-stimulus interval between a target tone and the standard tone of the next trial was drawn from a truncated exponential distribution (mean 1.5 s, truncated at 3 s) added to a minimum interval of 3 s (resulting in values between 3–6 s). One block consisted of 22 trials in the behavioural (one repetition per tone step and condition), and 56 trials in the EEG experiment (4 repetitions per tone step and condition). In the behavioural experiment participants performed 20 blocks (440 trials), and in the EEG experiment minimally 12 and maximally 15 blocks (672–840 trials). Between blocks, participants could take breaks of self-

determined length. Feedback was given per trial during the training, and at the end of each block (as proportion of correctly answered trials) during the main experiment.

After the experiment, participants responded to a questionnaire to assess whether they had detected the foreperiod manipulation, including a debriefing about the manipulation. Finally, they filled in a musicality survey (Schaal et al., 2014). The full experimental session lasted about 2.5 h.

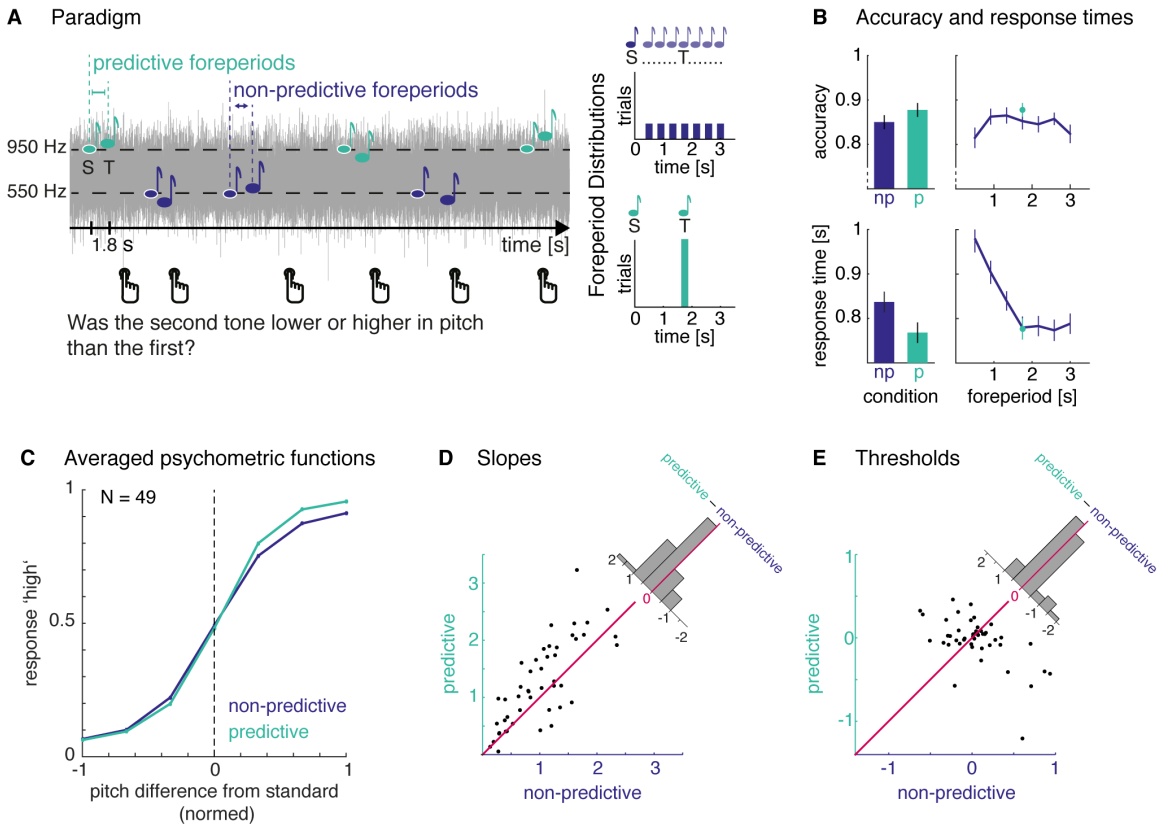

Figure 1: **Paradigm and Behavioural Results. A. Paradigm:** Tone-pairs were presented embedded in low-pass filtered white noise. Participants' task was to judge whether the target tone (T) was lower or higher in pitch than the preceding standard (S). Unbeknownst to participants, the pitch of the standard tone was associated with predictive (green) or non-predictive foreperiod intervals (blue). For the non-predictive condition, foreperiods were drawn from a uniform distribution (upper right panel), while for the predictive condition, foreperiods were fixed at 1.75 s (lower right panel). **B. Accuracy and response times:** Top: Accuracy improved significantly in the predictive condition (left panel), which was nominally also true at the intermediate foreperiod only. Bottom: Response times were faster in the predictive condition (left panel). The difference was driven by slower response times at short foreperiods on the non-predictive condition (right panel) **C. Averaged psychometric functions:** The slope of the psychometric function was steeper in the predictive compared to the non-predictive condition. There were no differences in threshold, guess rate or lapse rate. **D. Slopes for single participants:** for the non-predictive (x-axis) versus predictive (y-axis) conditions. **E. Thresholds for single participants:** for the non-predictive (x-axis) versus predictive (y-axis) conditions.

EEG recording and preprocessing

EEG was recorded with 64 electrodes Acticap (Easy Cap) connected to an ActiChamp (Brain Products) amplifier. EEG signals were recorded with the software Brain Recorder (Brain Products) at a sampling rate of 1 kHz, using no online high-pass filter and a 200 Hz low-pass filter. Impedances

were kept below 10 kΩ. Electrode TP9 (left mastoid) served as reference during recording. Electrode positions were digitized.

EEG data were analysed using the Fieldtrip software package for Matlab (MATLAB 2016a, MATLAB 2017a), and the lme4 package in R (Bates et al., 2015; R Core Team, 2016). First, we re-referenced the data to linked mastoids. Then we applied a low-pass filter to the continuous data (firws filter from the firfilt plugin, Widmann et al., 2015, cut-off 45 Hz, two-pass, transition bandwidth 3 Hz). No high-pass filter was applied. For the time-frequency analysis, we produced a parallel version of the data, using a one-pass high-pass filter (min-phase, 45 Hz). This was done to be able to look at phase coherence in the foreperiod interval just before the target in the predictive condition without backsmearing the target-evoked activity.

Next, we epoched the data around the standard tone onset (−3–6 s), and down-sampled to 100 Hz. All data were visually inspected to mark bad channels that were interpolated (1.2 channels on average). Then ICA were computed using the 'runica' algorithm, with the number of output components adjusted by subtracting the number of bad channels. Blinks, muscular artefacts, and unspecific noise occurring temporarily in a channel or trial were excluded, using the semi-automatic inspection of ICA components provided by the SASICA toolbox for fieldtrip (Chaumon et al., 2015) and removal of these (on average 33.7 components per participant).

Analyses

Analyses of the behavioural data

We analysed accuracy as percentage correct (after removing trials in which the standard and target were equal in pitch) and response times (for the EEG sample only, to inform the interpretation of the target-evoked event related potentials), defined as the interval between the onset of the target tone and the registered button press. Response times shorter than 0.2 s were considered outliers and removed. We compared accuracy and response times between conditions and over foreperiods for the non-predictive condition.

To obtain a measure of pitch discrimination sensitivity, we fitted psychometric functions to model participants' responses in the pitch discrimination task, using bayesian inference, implemented in the *Psignifit toolbox* for Matlab (Version 4, Schütt et al., 2016). The psychometric function describes the relationship between the stimulus level (on the abscissa; here: the difference in pitch between the target and the respective standard tone) and the participant's answer (on the ordinate; here: proportion of trials on which the target pitch was judged as higher). To accommodate the different standard tones per condition, and the individual pitch steps obtained during the training, we normed the discrete pitch differences per participant and condition to range between -1 and 1, with 0 being the pitch of the standard tone.

To select the options for the psychometric function (logistic versus cumulative normal function, number of free parameters), we assessed deviance pooled for both conditions. Deviance reflects a monotonic transformation of the log-likelihood-ratio between the fitted model and the saturated model (a model with no residual error), allowing for an absolute interpretation, or a comparison

between different models (Wichmann and Hill, 2001). The best fits (i.e. lowest deviance, 3.80 for the best model) were obtained by fitting a cumulative normal function with four free parameters: threshold, slope, guess rate, lapse rate.

For a yes-no-task as the one used here, threshold indicates the stimulus level at which a participant is as likely to judge the stimulus as 'low' or 'high'. Divergence from the actual midpoint of all stimulus levels (here: 0) can be interpreted as a response bias. Slope reflects the amount of stimulus change needed to increase the proportion of responding 'high', and can be interpreted as the sensitivity of the listener. Guess and lapse rate indicate the proportion of answering 'high' /'low' for the lowest/highest pitches in the tested range, respectively, that is the errors made by the observer. *Psignifit* uses default priors for the threshold, slope, guess and lapse-rates based on the given stimulus range (Schütt et al., 2016, p.109). Psignifit's version 4 fits a beta-binomial model (instead of a binomial model), which assumes that the probability for a given proportion of answers is itself a random variable, drawn from a beta distribution. This has been shown to provide better fits for overdispersed data, that is data in which answer probabilities over blocks and trials are not independent as assumed by the conventional model.

We fitted psychometric function to each individual's data and separately per condition and compared the resulting parameters between conditions (threshold, slope, guess- and lapse rates) using two-sided t-tests. Additionally, we calculated Bayes Factors for all statistical tests, using the *Bayes Factors* package for Matlab (Rouder et al., 2009).

We computed a logistic regression on the single-trial responses of the pitch-discrimination task, to parallel the analysis of delta phase angles performed for the EEG (see below). Pitch difference and condition were used as interacting fixed effects (with random intercepts and random slopes for both predictors and their interaction), using the lme4 package in R (function *glmer*, Bates et al., 2015) with a binomial link function.

Event related potentials

We examined the time-domain data with respect to responses evoked by standard and target tones, contrasting the predictive and non-predictive condition. For the standard-evoked response, we detrended the data based on the whole epoch and applied baseline correction from −0.1–0 s pre-standard. We only examined the time-window between standard onset and 0.5 s after, because this was the maximal interval in which no target events occurred (earliest target onset was 0.5 s in the non-predictive condition). For the target-evoked response, we first applied detrending and the same pre-standard baseline as described above, and then re-epoched to the target event. We examined the time interval from −0.5–0.5 s around the target event. We averaged over trials within participants and condition, and then over participants, to obtain the average event related potential (ERP, depicted in Figure 2).

To test for statistically significant differences in the time-domain data, we applied cluster permutation tests on two levels. First, we contrasted trials from the non-predictive and predictive condition within each participant using independent samples regression coefficient t-tests implemented in FieldTrip (ft_timelockstatistics). This resulted in weights for each time-electrode data

point for the ERPs. Next, the group-level analysis was performed with a dependent samples t-test to contrast the weights from the subject-level analysis against zero. A permutation test (5000 Monte Carlo random iterations) was performed with cluster-based control of type I error at a level of α=0.05 as implemented in FieldTrip. The condition assignment (i.e. whether the predictive condition was presented at the low or high pitch tones) was added as a control variable. This analysis resulted in time-electrode clusters exhibiting significant condition differences in the ERPs.

Time-frequency representations

Time-frequency representations were computed for epochs time-locked to the standard tones, separately for the predictive and non-predictive condition. We performed this analysis on trials with foreperiods equal or longer then 1.75 s only to avoid evoked activity from target onsets occurring between 0.5–3 s in the non-predictive condition. We matched the smaller number of trials available from the non-predictive condition, by randomly sampling the same number of trials from the predictive condition. To obtain stable results, we repeated the random sampling 50 times and averaged over the resulting time-frequency representations. Additionally, we ruled out potential back-smearing of evoked activity related to target-onset by using data that was filtered with a one-pass low-pass filter, and by replacing all data points after 1.75 s by the value at this time point for the respective trial and channel before performing the time-frequency transformation.

Data were transformed to time-frequency representations for frequencies ranging from 0.5 to 34.5 Hz (linear steps, 1 Hz) and time points between −0.5–2.5 s, using convolution with a single adaptive Hanning taper with frequency-dependent time windows (increasing linearly from 2 to 4 cycles per frequency). To provide sufficiently long data epochs for the lowest frequencies, we appended the epochs (−3–6 s, time locked to the standard tone) with their inverted and right-left flipped version to the left and right before applying the time-frequency transform.

Power estimates were extracted as the squared modulus of the complex-valued Fourier spectra and baseline corrected to relative change (first subtracting, then dividing by the trial-average baseline value per frequency) using the condition average in the interval from −0.5 s to standard onset. Inter-trial phase coherence (ITC) was extracted as the magnitude of the amplitude-normalized complex values, averaged across trials for each time-frequency bin and channel. Statistics were performed in the time-window between 0–1.7 s post standard onset and for all frequencies jointly. For power, we used a two-level procedure as described for the ERPs (but using ft_freqstatistics, 1000 permutations). For the ITC, we only computed the second-level statistics since it represents a measure that already combines single trials.

Delta phase angle analyses

A timing mechanism that predicts the onset of the target tone would have to start timing at the standard tone which serves as a temporal cue. Therefore, we examined the data for any signatures of such a mechanism in the phase of the delta band (see Figure 4B for a schematic depiction). To not distort phase angles by filtering, we used the two-pass filtered version of the data (low-pass only, as for the ERPs). To reduce the dimensionality of the data, and to focus our analysis on auditory activity,

we computed a weighted average of single electrodes at each time point. The weights reflected each participant's N1-peak topography, computed as the average absolute value per channel in the time interval from 0.14–0.18 s following the standard (see topography shown in Figure 4B). We then multiplied the time-domain data at all latencies and channels with these weights and averaged over channels, resulting in one virtual channel.

Target-onset ERPs were muted (as described above) from the time point of target onset on each trial. We then applied a band-pass filter to the data (3rd order Butterworth, two-pass), with cut-off frequencies of 1 and 3 Hz for the delta band. After filtering, we applied the Hilbert transform and extracted phase angles as the imaginary value of the complex fourier spectrum averaged over latencies from 0.14–0.18 s.

We subjected the phase angles to a logistic regression to test for an effect of phase angle on the behavioural response, using the lme4 package in R (function glmer with a binomial link function, Bates et al., 2015). Per trial, we predicted the participant's response in the pitch discrimination task (second tone lower or higher) with two numerical predictors, (1) the normalized pitch difference between standard and target tone (Δpitch in eq. 1, range −1–1, a.u.), and (2) the standard-evoked phase angle extracted as described above (φ), plus their interaction.

The predictors of the logistic regression can be interpreted following the logic of the psychometric function (DeCarlo, 1998), which models a behavioural measure (on the ordinate) based on variations of a stimulus feature (on the abscissa), and is described by two main parameters: threshold and slope. A threshold effect, that is a horizontal shift of the psychometric function, would be reflected by a main effect of the predictor φ. A slope effect, that is a shift in the steepness of the psychometric function, would be reflected by an interaction between the predictors Δpitch and φ. Here, we were particularly interested in a slope effect, that is an interaction between the predictors pitch and phase angle. Due to computational constraints, we only specified a random intercept, but no random slopes for the predictors.

To account for the circularity of the phase angles, we followed an approach previously described by Wyart et al. (2012; see also Cravo et al. 2013) of using the sine and cosine of the phase angles jointly as linear predictors in a regression. For both, the $sin(\varphi)$ and $cos(\varphi)$, we specified an interaction with Δpitch:

$$y = \beta_0 + \beta_1 \cdot (\Delta pitch \cdot sin(\varphi)) + \beta_2 \cdot (\Delta pitch \cdot cos(\varphi)) \tag{1}$$

Then, we recombined the regression weights obtained for the interactions of $sin(\varphi)$ and $cos(\varphi)$ with Δpitch:

$$\beta_{combined} = \sqrt{\beta_1^2 + \beta_2^2} \tag{2}$$

The resulting $\beta_{combined}$ is always positive and can thus not be tested against zero. We computed a reference distribution of $\beta_{combined}$ based on 1000 permutations, by permuting, per participant, the assignment between the response and φ values over trials, recomputed the model and retained the $\beta_{combined}$. To assess significance of the interaction between pitch and phase angle, we assessed 99%

one-sided confidence intervals, and computed p-values from the permutation distribution (following Phipson and Smyth, 2010):

$$p_{perm} = \frac{N(\beta perm_{combined} > \beta_{combined}) + 1}{N(perm) + 1} \tag{3}$$

We additionally computed the phase analysis on data filtered for the low delta (0.5–2 Hz), theta (4–7 Hz), alpha (8–12 Hz), and beta (15–30 Hz) frequency bands and tested the resulting $\beta_{combined}$ for significance using the permutation approach. P-values were Bonferroni-corrected (accounting for five tests with a p-value threshold of 0.05, one for each frequency band), resulting in an adjusted alpha level of 0.01.

To visualize the modulation of pitch discrimination sensitivity over phase angles, we predicted responses from the logistic regression model (using the R package *emmeans*, Lenth, 2018), for a range of $\Delta pitch$, $sin(\varphi)$, and $cos(\varphi)$ values, and plotted the resulting values for the recombined and binned φ (shown in Figure 4C).

Finally, to assess the temporal specificity of the effect in the delta frequency band, we computed the logistic regression over latencies and for each of the two conditions separately. We also computed a permutation distribution over time (using 200 permutations only, due to the time-consuming procedure).

Results

Temporal predictability improves pitch discrimination

On average, participants' responses were correct in 86% percent of trials. Using the full sample of 49 participants, we found that accuracy was significantly higher in the predictive compared to the non-predictive condition (T(48)=3.61, p<0.001, BF=54.86; Figure 1B). We found a nominal, albeit not significant increase in accuracy at the intermediate foreperiod for the predictive compared to the non-predictive condition (Figure 1B), probably due to the small numbers of trials in the non-predictive condition), suggesting that the performance improvement occurred not only at unexpectedly early or late foreperiods, but reflects a difference between conditions. We furthermore analysed response times for the EEG sample, to inform the interpretation of the target-evoked event related potentials. Response times were faster in the predictive (average 0.77 s), compared to the non-predictive condition (0.84 s), by about 70 ms (T(25) = 7.9, p < 0.001). As shown in Figure 1B, the difference is mainly driven by slower responses at early foreperiods in the non-predictive condition.

For the psychometric functions (depicted in Figure 1C), we observed a significant increase of the slope in the predictive compared to the non-predictive condition (T(48)=3.67, p<0.001, Bayes Factor (BF)=64.16; Figure 1D), but no threshold effect (T(48)=0.92, p=0.36, BF=0.31; Figure 1E), nor effects on the guess and lapse rate (p>0.28, BF: 0.22, 0.34).

All of the above results held, albeit with smaller effect sizes, when analysing only the participants for whom we had recorded EEG: predictability resulted in marginally higher accuracy,

(T(25)=1.82, p=0.08, BF=1.07), significantly larger PMF slopes (T(25)=2.60, p=0.02, BF=4.04), and no effects for the threshold, guess, and lapse rate (all $p > 0.18$, BF: 0.43, 0.61, 0.29).

To parallel the analysis of delta phase angles reported below, we also computed logistic regression or the behavioural data with the predictors pitch difference (Δpitch), condition, and their interaction (plus random effects for all three), only for the participants from the EEG sample. The analysis confirms the results described above, namely a significant main effect for Δpitch ($p<0.001$), no main effect for condition (p=0.9), but an interaction between Δpitch and condition ($p<0.01$), that is a slope effect (see Figure 4A).

Finally, we ascertained that the predictability manipulation had not been noted by participants and thus had remained strictly implicit, as intended. During debriefing, the majority of participants reported to not have noticed that either the high or low tones were temporally predictive (3 participants in the EEG sample said they had noticed *after* the experimenter explained the manipulation). When asked to guess whether the low or high standard tones were temporally predictive, performance was at 50% for the EEG-participants.

Temporal predictability affects both, standard- and target-evoked event related potentials:

Standard-evoked activity: Event related potentials were examined time-locked to the standard-tone (Figure 2A). Both conditions showed a negative deflection between 0.1–0.2 s after the standard onset, with a peak at 0.16 s and a fronto-central topography. We refer to this component as the standard-evoked N1. This component was significantly more negative for standards that were temporally predictive of the onset of the target (predictive condition; 0.21–0.26 s, p=0.04). This difference is important in that it shows that standard tones were processed differently if they served as a temporal cue for the target onset versus did not serve as a temporal cue. The latency and topography of the standard-evoked N1 (not the time-range in which the difference was found which was slightly later) was used for the analysis of phase angles described below. When directly comparing the ERPs evoked by the 550 versus 950 Hz standards (randomly assigned to the predictive and non-predictive condition over participants), there was no statistically significant difference in the early time window following the standard tone.

Target-evoked activity: Event related potentials time-locked to the target-tone (Figure 2B) also showed a negative deflection between 0.1–0.2 s after the target onset, with a fronto-central topography. We refer to this component as the target-evoked N1. For targets in the predictive condition, the N1 was larger (0.09–0.14 s, p=0.03). Importantly, the difference is not solely due to the onset time of the target (see inset in Figure 3B), which would be reflected by a difference only for long or short foreperiods in the non-predictive condition.

To test for an apparent latency shift in the N1 between the non-predictive and predictive conditions, we computed the half-area measurement (Luck, 2005), which indexes the time-point at which half the area of a deflection has been reached. Compared to peak-latencies, this measure better accounts for asymmetric deflections. We found a significantly earlier N1-latency for the predictive,

compared to the non-predictive condition (0.13 s versus 0.15 s; T(23)=3.03, $p < 0.01$). Furthermore, there was a difference at the later positive prolonged component, which was positive at posterior and negative at frontal electrodes (0.20–0.38 s, $p<0.01$; 0.28–0.37 s $p = 0.02$ s).

Delta (1–3 Hz) oscillatory changes during predictive foreperiods

We assessed power in a frequency range between 0.5–34.5 Hz for the predictive and non-predictive conditions (see Figure 3A), time-locked to standard onset. Both conditions showed an increase in power in the delta-range (1–3 Hz, Figure 3B) after standard onset, and a prolonged increase in the alpha-range (8–12 Hz) relative to baseline. Subtracting power in the non-predictive from the predictive condition, showed an increase in power for the predictive condition in frequencies ranging

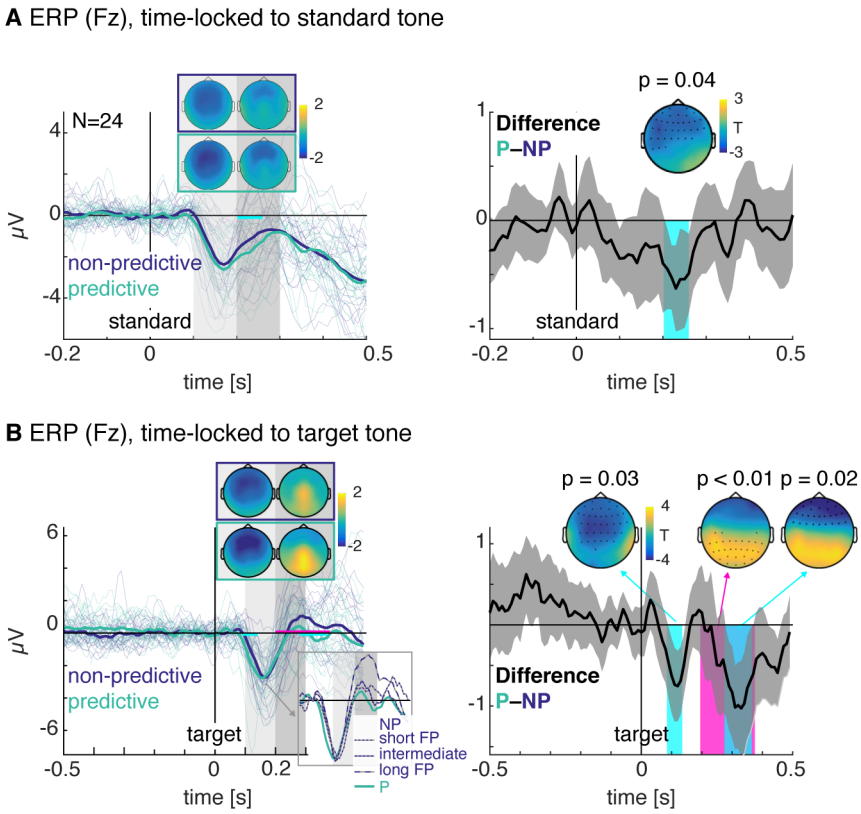

Figure 2: **Event related potentials. A. ERPs time-locked to the standard tone:** Left: The predictive condition (green line) evoked a more negative N1 than the non-predictive condition (blue line). The fine blue and green lines depict single participants' ERPs. The inset shows the topographies in the time windows of 0.1–0.2 s and 0.2–0.3 s for both conditions separately. Right: condition difference. The grey shades indicates the two-sided 95% confidence interval, estimated from the t-distribution. The cyan shade marks the time points at which a significant condition difference occurred, and the topography shows the scalp distribution of the activity during these time windows. Channels at which the difference was significant are marked in black. **B. ERPs time-locked to the target tone:** Left: The predictive condition (green line) evoked an earlier N1 than the non-predictive condition (blue line). The upper inset shows the topographies in the time windows of 0.1–0.2 s and 0.2–0.3 s for both conditions separately. The lower inset shows the target-evoked ERP split into foreperiod (FP) bins for the non-predictive condition, which suggests that the observed latency-shift is not solely due to the foreperiod, but driven by predictability. Right: condition difference. The cyan and pink shades mark the time points at which a significant condition difference occurred, and the topographies show the scalp distributions of the activity during these time windows.

between 0.5–12.5 Hz (0.05–1.7s s, p<0.01). To assess the frequency-specificity of the effect, we lowered the threshold for inclusion of single time-frequency points in a cluster from 0.05 to 0.025, which revealed that the stronger power difference occurred in the delta-range after standard onset (0.5–5.5 Hz, 0.4–1.7s s, p<0.01), mainly over right temporal and occipital electrodes (see Figure 3C). Despite testing for effects at all frequencies jointly, we found no other significant clusters.

When comparing inter-trial phase coherence (ITC) for all frequencies between conditions, no significant differences were observed. However, ITC across the 1–10 Hz range did show the expected increase following the standard tone, ranging from 1–10 Hz, and a prolonged increase in the delta band in both conditions (Figure 3C).

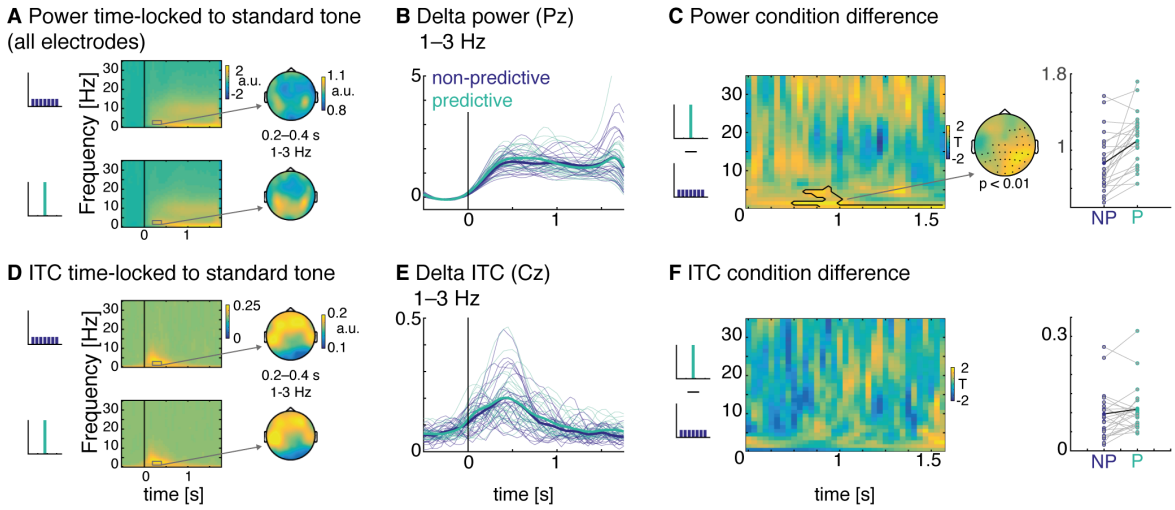

Figure 3: **Time-frequency representations. A. Power, time-locked to standard-onset.** Power estimates were baseline-corrected to the pre-standard interval and display relative change. Top panel: non-predictive condition, bottom panel: predictive condition. The topographies show the power scalp distributions in the interval from 0.2–0.4 s for frequencies from 1–3 Hz. **B. Delta power (1–3 Hz) over time** for the non-predictive (blue) and predictive conditions (green). Fine lines depict single participants' power values. **C. Power-difference between conditions (T-values).** The black shape marks the statistically significant cluster, and the topography shows the scalp distribution of the cluster. The scatterplot to the right shows average power values extracted from the cluster peak for each participant. **D. Inter-trial phase coherence (ITC), time-locked to standard-onset.** Top panel: non-predictive condition, bottom panel: predictive condition. The topographies show the ITC scalp distributions in the interval from 0.2–0.4 s for frequencies from 1–3 Hz. **E. Delta ITC (1–3 Hz) over time** for the non-predictive (blue) and predictive conditions (green). Fine lines depict single participants' ITC values. **F. ITC-difference between conditions (T-values).** No significant condition differences were found. The scatterplot to the right shows average ITC values extracted from the peak of the power-difference cluster (above) for each participant.

Standard-evoked delta phase angle predicts pitch discrimination sensitivity

To test whether delta oscillations play a role in temporally predictive processing in this study, we tested for a relation between delta phase angles evoked by the standard-tone and pitch discrimination performance using a logistic regression approach (see Figure 4B for a schematic depiction). A timing mechanism that predicts the onset of the target tone would have to start timing at the standard tone, which is why we were particularly interested in this time window.

Phase angles in the post-standard time window (0.14–0.18 s) were extracted by applying the Hilbert transform to band-pass filtered (1–3 Hz) single trial data with one virtual channel (see Methods for details) representing the sum of all channels weighted by the N1-topography. We subjected the phase angles (as their sine and cosine) with two numerical predictors, the normalized pitch difference between standard and target tone, and the standard-evoked phase angle, plus their interaction. To assess significance of the interaction effect, we used a permutation approach. We found a significant interaction between pitch and phase angle, which indicates that the slope of the psychometric function varied depending on the delta phase angle evoked by the standard tone (Figure 4 C). The interaction effect was significant only for the delta band (1–3 Hz), but not for other frequency bands tested (0.5–2 Hz; 4–7 Hz; 8–12 Hz; 15–30 Hz; Figure 4 F). Note that this procedure was performed on all trials, without separation into conditions, and thus is generally valid, both for trials on which the standard served as a temporal cue and trials for which it did not.

Next, we tested whether the interaction between delta phase angle and pitch discrimination sensitivity was specifically driven by our manipulation of temporal predictability. We assessed the relationship between phase angle (binned into 6 bins for this purpose) and condition (indexed as -1 for the non-predictive and 1 for the predictive condition; Figure 4 D). If the trials would be equally distributed over conditions per phase angle bin, this should result in an average condition of 0 at all phase angles, which was not the case. Instead we found more trials of the predictive condition to occur at the phase angles at which we had found the higher slopes (Figure 4 C), which suggests that phase angles varied between the two conditions. As a post-hoc test for a quadratic effect of phase bin on condition, we computed a generalized linear model predicting condition from phase bins (coded as 1,0,-1,-1,0,1) and obtained a significant weight for this contrast ($p = 0.03$).

Additional analyses

Influence of delta phase at different time points. We examined the regression weight for the interaction at different time points over the trial, and independently for the predictive and non-predictive conditions. This analysis (Figure 4E) showed that the interaction effect between delta phase angle and the slope of the psychometric function was largest at time points after the standard tone (0–0.5 s), and that the effect was significant (i.e. exceeded the 99% confidence interval of the permutation distribution) only for the predictive condition, suggesting that it was driven by temporal predictability. We therefore conclude that the interaction effect was mainly driven by the predictive condition.

Mediation analysis. We also considered mathematically the possibility that delta phase angle would mediate the effect of temporal predictability on pitch discrimination sensitivity, by comparing the regression weight of the interaction between pitch *times* temporal predictability estimated from a model with no other predictors (as depicted in 4A), and from a model that additionally contained an interaction term for pitch *times* phase angle (Baron and Kenny, 1986; Muller et al., 2005). The negligible change in weight between both models (0.307 to 0.303) indicate that there is no evidence for a mediation effect.

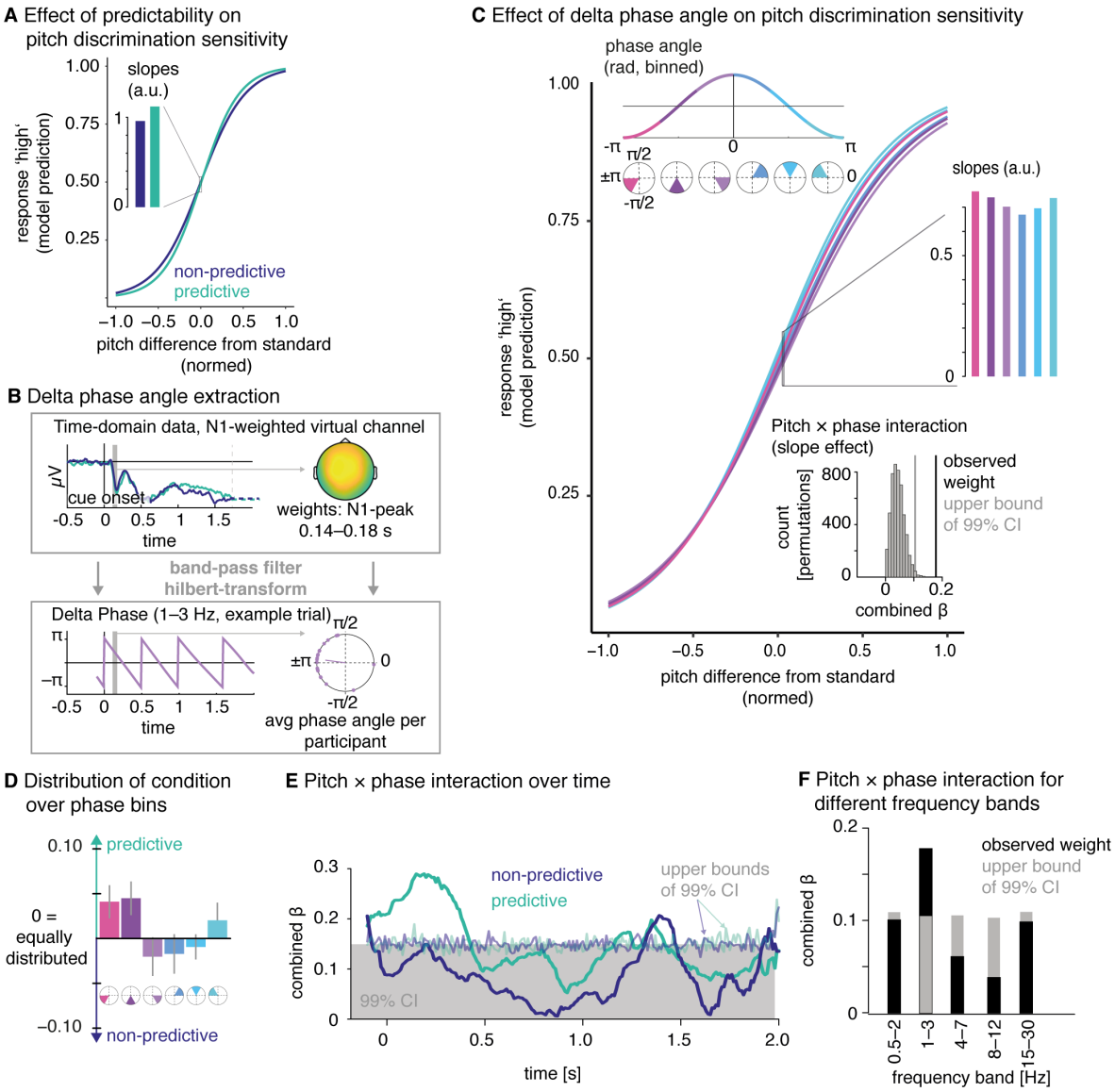

Figure 4: **Delta phase angle predicts pitch discrimination sensitivity. A. Replication of the behavioural effect** (s. Figure 1) with a logistic regression approach. Model predictions from the logistic regression with the predictors pitch (x-axis) and condition (colors). As illustrated by the bar-plot, there was a slope difference between conditions (i.e. an interaction between pitch and condition), with steeper slopes for the predictive condition. **B. Schematic depiction of the delta phase angle analysis.** We extracted the time domain data from single trials, from one virtual channel that reflects the weighted sum of the standard-evoked N1 topography (computed in the interval from 0.14–0.18 s), band-pass filtered (1–3 Hz) and applied the Hilbert transform, to extract the instantaneous phase angles in the time-window of 0.14–0.18 s (the N1-peak). **C. Effect of delta phase angle on pitch discrimination sensitivity:** Model predictions from the logistic regression model with the predictors pitch (x-axis) and phase angle (colors, binned only for visual display). There was a significant interaction between pitch and phase, that is the slopes of the psychometric functions differed depending on delta phase angle (depicted in the bar plot). Note that this analysis was performed on all trials, without separation into conditions. The inset on the bottom right side shows the observed interaction weight (in black) compared to a permutation distribution and its 99% confidence interval (in grey). **D. Distribution of conditions over phase angles.** Conditions were coded as −1 for the non-predictive and 1 for the predictive condition, therefore an equal distribution of conditions over phase angle bins should result in an average condition (colored bars) of 0, which was not the case. Instead, more trials from the predictive condition occurred at the phase angles that were related to a steeper slope of the psychometric function (panel C).

Figure 4: (Continued from previous page.) **E. Pitch × phase interaction over time, separated by condition.** The thick lines indicate the regression weights for the interaction over time for the predictive (green) and non-predictive condition (blue), the thin lines and grey shade indicate the 99% confidence interval computed with the permutation approach. **F. Pitch × phase interaction and confidence intervals for different frequency bands.** The grey bar shows the 99% confidence interval, the black bar the observed weight. Only for the delta band (1–3 Hz) the observed weight significantly exceeds the permuted weights.

Delta phase vs. ERP effect. To distinguish between the ERP effect (found on the N1) and the delta phase effect, we tested whether the N1 amplitude could explain the findings. Computing the same logistic regression model with the N1 amplitude instead of the phase angles as above revealed no significant interaction effect (p=0.15), i.e. the N1 amplitude does not predict pitch discrimination performance on single trials and can thus not simply replace the delta phase angle. However, the N1 amplitude correlated significantly with the standard-evoked phase-angle at all frequency bands, as assessed by a circular-linear correlation (from the Directional package in R Tsagris et al., 2018); R^2: 0.5–2Hz: 0.21, 1–3Hz: 0.27, 4–7Hz: 0.06, 8–12Hz: 0.056, 15–30Hz: 0.004 (all p-values <0.001).

Discussion

In this study, we asked whether human listeners use strictly implicit temporal contingencies in auditory input to form temporal predictions. If so, how are these predictions represented endogenously? We implicitly manipulated temporal predictability by varying the foreperiod (i.e., the interval between standard and target tone) in a pitch discrimination task. Unbeknownst to participants, one of two possible pitches used as the standard tone was indicative to one of two foreperiod distributions, respectively: a uniform distribution, under which the onset of the target tone is unpredictable, and a single-valued distribution under which the onset of the target tone is fully predictable. The data show several indices that participants formed temporal predictions: an increase in pitch discrimination sensitivity in the predictive condition, condition differences in the event-related activity to standard- and target tones, and increased delta power during the foreperiod in the predictive condition. Furthermore, we show that only the phase of delta oscillations in response to the standard tone, which serves as a temporal cue, is indicative of pitch discrimination performance, suggesting that delta oscillations are instrumental for endogenous temporal predictions.

Implicit temporal predictability improves pitch discrimination sensitivity

The observed increase in pitch discrimination sensitivity in the predictive condition suggests that listeners can learn implicitly to associate interval-based temporal predictions with sensory stimulus features like pitch, underlining the relevance of timing to human cognitive processing.

The behavioural data yielded an increase in pitch discrimination sensitivity for the temporally predictive condition, reflected in a steeper slope of psychometric functions modelling the participant's response as a function of the pitch difference between the standard and target tone (Figure 1).

We also show the effect when analysing the data via logistic regression, to match the delta phase angle analyses (Figure 4A). Importantly, participants were not made aware of the predictability manipulation, and the large majority of participants did not notice it (three might have). When asked to guess whether in their case the low- or high-pitched tones were indicating predictability, average performance was 50%, showing that the manipulation was implicit.

To our knowledge, this is the first study to show that pitch discrimination sensitivity is improved by implicit but non-rhythmic temporal predictions. In the auditory domain, detection speed and performance are facilitated by rhythmic temporal predictability (Henry et al., 2014, 2016; Henry and Obleser, 2012; Lawrance et al., 2014; Wright and Fitzgerald, 2004), but the use of detection tasks might strengthen motor aspects of the task. One study by Bausenhart and colleagues (2007) showed that shorter presentation times (difference of about 6 ms) are needed for to achieve correct pitch discrimination performance, when the target tone occurs with a block of constantly short foreperiods. Our study complements these previous findings showing implicit temporal predictability can directly improve auditory pitch discrimination in the absence of an embedding rhythm, or any explicit incentive to engage in timing.

The role of slow neural oscillations in temporal prediction

A central aim of this study was to assess the role of slow neural oscillations in the endogenous representation of temporal predictions. Previous studies convincingly established a mechanism of sensory enhancement via phase alignment of delta oscillations for stimuli that occur during the preferred phase, i.e. in synchrony with the preceding rhythm (Cravo et al., 2013; Henry et al., 2014; Kösem et al., 2017; Lakatos et al., 2008; Schroeder and Lakatos, 2009). An open question is, however, whether the alignment of slow neural oscillations towards predicted stimulus onsets is contingent on rhythmic entrainment of neural oscillations to the exogenous stimulation, or whether slow oscillations also implement endogenous temporal predictions, for example via single-trial phase-reset.

First, increased delta power in the foreperiod interval in the predictive compared to the non-predictive condition (see Figure 3) suggests a relevance of delta oscillations for temporal predictions. Crucially, here the predictive and non-predictive conditions were alternated trial-by-trial, indicating that delta power can be regulated in flexible manner. This suggests that a variant of the rhythmic mode of attention, reflected by increased delta power (Schroeder and Lakatos, 2009) can also apply to interval-based temporal predictions. Enhanced delta power found in the predictive condition possibly reflects an up-regulation, or large-scale synchronization of delta oscillations (Herrmann et al., 2016), necessary to represent temporal predictions.

The absence of condition differences in phase coherence during the foreperiod (Figure 3F) replicates our previous results (Herbst and Obleser, 2017) and suggests that enhanced phase alignment (Breska and Deouell, 2017; Cravo et al., 2011) depends on dedicated or residual periodicity in the stimulation (Obleser et al., 2017). As a side note, it is important to emphasize the methodological challenges one faces when analysing low frequency oscillations in the pre-target window. The probabilistic manipulation of foreperiods as applied here results in differential time-locking of target activity between conditions, and our conservative approach of removing these might have weakened

existing pre-target differences. Thus, our findings suggest that the representation of temporal predictions by enhanced phase coherence – or at least our ability to measure this in human EEG – is contingent on rhythmic stimulation.

Crucially, we found that the absolute phase angle of the delta oscillation in auditory areas shortly after the temporal cue predicted behavioural sensitivity in response to the later-occurring target tone (see Figure 4C). This effect was specific for the delta band (1–3 Hz) with the highest sensitivity occurring at phase angles closest to the trough of the delta oscillation ($\pm\pi$) at the cue and about 1.7 s post-cue (average period of 0.5 s). Albeit interpreting the absolute phase angle from EEG data demands caution, this corroborates the idea that the trough of the delta oscillation is a particularly beneficial state for auditory perception (Henry et al., 2016; Lakatos et al., 2013). The effect was observed for data spatially filtered with a topography relevant for auditory stimulus processing (from the N1), suggesting auditory cortex as the most likely generator. We further showed that the N1 amplitude itself does not show the critical relationship with behavioural sensitivity, arguing for a more specific role of delta oscillations in temporal prediction.

While the relationship between delta phase and behavioural sensitivity was true for all trials, regardless of their experimental condition, follow-up analyses showed that (i) the phase angle evoked by the standard tone differed between the predictive and non-predictive conditions (Figure 4D), and (ii) that the interaction effect was stronger in the predictive condition (Figure 4E), suggesting that the effect was driven by temporal predictability. To our knowledge this represents the first evidence that delta oscillations can be reset on single trials to initialize interval-based temporal predictions.

Clearly, we have not proven a *causal* link of temporal predictability increasing auditory sensitivity *via* an optimized phase angle of delta oscillations. Our analysis failed to establish a mediation effect. We cannot rule out that the different steps necessary to accommodate the complexity of our data in the model (dealing with the circular measure of phase angle and assessing an interaction effect as a measure of behavioural sensitivity), and the small proportion of variance explained by the experimental manipulation (a common problem in cognitive neuroscience) might have prevented us from observing a mediation effect (but see Benwell et al., 2017, for a successful example). Nevertheless, there are likely additional signatures in the neural data that contribute to the endogenous representation of temporal predictions which were not captured by our analysis.

Still, the finding that the phase of delta oscillations in response to a temporal cue predicts behavioural sensitivity, and more strongly so in the predictive condition, combined with enhanced delta power during predictive foreperiods clearly underlines the relevance and specificity of delta oscillations to the endogenous representation of temporal predictions. The adjustment of phase angles at the cue can be seen as the initiation of a timing process, which prepares the system to be in a beneficial state at an anticipated time point.

Temporal predictions affect sensory processing of predictive and predicted tones

Predictive tones (standards). An important indicator of the use of temporal predictability is the condition difference found between the event related potentials evoked by predictive and non-

predictive standard tones (Figure 2A), namely a more negative N1 for predictive compared to non-predictive tones. This shows that participants learned to associate the pitch of the standard tone to temporal predictability, and flexibly used the standard as a temporal cue on a trial-by-trial basis.

Few studies have investigated effects of predictability on the early processing of the cue stimulus itself. In spatial cueing, there is evidence for an effect of predictions on early positive and negative cue-evoked components (100–200 ms post cue Jongen et al., 2007; Nobre et al., 2000; Yamaguchi et al., 1994). In the temporal domain, there is – to our knowledge – only one study that showed an N1-effect directly at the cue (in 8–12 years old children, Mento and Vallesi, 2016). Our results are in line with this finding and reveal that the standard-evoked N1 in adults is affected even by implicit temporal predictability.

Predicted tones (targets). In response to target tones, we found a larger and faster N1 in the predictive compared to the non-predictive condition, suggesting a facilitation for temporally predicted targets (Figure 2B). This result corroborates a large base of studies reporting mainly amplitude effects of temporal predictability on early sensory evoked potentials (Correa et al., 2006; Hsu et al., 2014; Hughes et al., 2013; Kok et al., 2011; Lampar and Lange, 2011; Lange, 2009; Miniussi et al., 1999; Sanders and Astheimer, 2008; Schwartze et al., 2013). The direction of amplitude effects varies with the paradigm used (for an extensive discussion see Lange, 2013) – for probabilistic foreperiod variations as used here both, reduced (Paris et al., 2016; Sherwell et al., 2017) and enhanced N1 amplitudes (Griffin et al., 2002) have been reported.

The observed latency-shift of the N1 by temporal predictions is in line with a previous study on rhythmic temporal predictability (Rimmele et al., 2011). Further evidence comes from experiments reporting a faster N1 for auditory speech and non-speech events combined with visual events (Paris et al., 2017; Stekelenburg and Vroomen, 2007; Vroomen and Stekelenburg, 2010; Wassenhove et al., 2005). However, there is mixed evidence as to whether the temporal information conveyed by the visual event alone leads to the faster N1 (Vroomen and Stekelenburg, 2010), or whether it does so only in combination with sensory predictions (Paris et al., 2017). Note that in our study, the predictive information conveyed by the cue was purely temporal, since the pitch of the target tones varied around their respective standard. Thus, the facilitation of the N1 found here provides novel evidence that temporal predictions alone can enhance early auditory processing.

Conclusions

Here we show that human listeners use strictly implicit temporal contingencies to perform a sensory task for which timing is not an explicit requirement. The neural signatures observed suggest that the standard tone is used as a temporal cue and provokes an optimized phase reset of delta oscillations on single trials and an increase in delta power, associated with enhanced behavioural sensitivity in the predictive condition.

References

Arnal, L. H., Doelling, K. B., and Poeppel, D. (2014). Delta–beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*, 25(9):3077–3085.

Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bauer, A.-K. R., Jaeger, M., Thorne, J. D., Bendixen, A., and Debener, S. (2015). The auditory dynamic attending theory revisited: A closer look at the pitch comparison task. *Brain Research*, 1626:198–210.

Bausenhart, K. M., Rolke, B., and Ulrich, R. (2007). Knowing when to hear aids what to hear. *The Quarterly Journal of Experimental Psychology*, 60(12):1610–1615.

Benwell, C. S., Keitel, C., Harvey, M., Gross, J., and Thut, G. (2017). Trial-by-trial co-variation of prestimulus eeg alpha power and visuospatial bias reflects a mixture of stochastic and deterministic effects. *European Journal of Neuroscience*.

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., Emerson, R. G., and Schroeder, C. E. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *Journal of Neuroscience*, 31(9):3176–3185.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–436.

Breska, A. and Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS biology*, 15(2):e2001665.

Chaumon, M., Bishop, D. V., and Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of neuroscience methods*, 250:47–63.

Correa, Á., Lupiáñez, J., Madrid, E., and Tudela, P. (2006). Temporal attention enhances early visual processing: A review and new evidence from event-related potentials. *Brain research*, 1076(1):116–128.

Correa, Á., Lupiáñez, J., Milliken, B., and Tudela, P. (2004). Endogenous temporal orienting of attention in detection and discrimination tasks. *Perception & Psychophysics*, 66(2):264–278.

Correa, Á., Lupiáñez, J., and Tudela, P. (2005). Attentional preparation based on temporal expectancy modulates processing at the perceptual level. *Psychonomic bulletin & review*, 12(2):328–334.

Coull, J. T. and Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 18(18):7426–7435.

Cravo, A. M., Rohenkohl, G., Wyart, V., and Nobre, A. C. (2011). Endogenous modulation of low frequency oscillations by temporal expectations. *Journal of Neurophysiology*, 106(6):2964–2972.

Cravo, A. M., Rohenkohl, G., Wyart, V., and Nobre, A. C. (2013). Temporal Expectation Enhances Contrast Sensitivity by Phase Entrainment of Low-Frequency Oscillations in Visual Cortex. *The Journal of Neuroscience*, 33(9):4002–4010.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological methods*, 3(2):186.

Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., Schwartz, T. H., and Foxe, J. J. (2011). Oscillatory sensory selection mechanisms during intersensory attention to rhythmic auditory and visual inputs: a human electrocorticographic investigation. *Journal of Neuroscience*, 31(50):18556–18567.

Griffin, I. C., Miniussi, C., and Nobre, A. C. (2002). Multiple mechanisms of selective attention: differential modulation of stimulus processing by attention to space or time. *Neuropsychologia*, 40(13):2325–2340.

Haegens, S. and Golumbic, E. Z. (2017). Rhythmic facilitation of sensory processing: a critical review. *Neuroscience & Biobehavioral Reviews*.

Henry, M. J., Herrmann, B., and Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences*, 111(41):14935–14940.

Henry, M. J., Herrmann, B., and Obleser, J. (2016). Neural Microstates Govern Perception of Auditory Input without Rhythmic Structure. *Journal of Neuroscience*, 36(3):860–871.

Henry, M. J. and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences*, 109(49):20095–20100.

Herbst, S. K. and Obleser, J. (2017). Implicit variations of temporal predictability: Shaping the neural oscillatory and behavioural response. *Neuropsychologia*, 101:141–152.

Herrmann, B., Henry, M. J., Haegens, S., and Obleser, J. (2016). Temporal expectations and neural amplitude fluctuations in auditory cortex interactively influence perception. *NeuroImage*, 124, Part A:487–497.

Hsu, Y.-F., Hamalainen, J., and Waszak, F. (2014). Both attention and prediction are necessary for adaptive neuronal tuning in sensory processing. *Frontiers in human neuroscience*, 8:152.

Hughes, G., Desantis, A., and Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological bulletin*, 139(1):133.

Jones, M. R., Moynihan, H., MacKenzie, N., and Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological science*, 13(4):313–319.

Jongen, E. M., Smulders, F. T., and Van der Heiden, J. S. (2007). Lateralized erp components related to spatial orienting: Discriminating the direction of attention from processing sensory aspects of the cue. *Psychophysiology*, 44(6):968–986.

Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., and De Lange, F. P. (2011). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral cortex*, 22(9):2197–2206.

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A. S., Jensen, O., and Hagoort, P. (2017). Neural entrainment determines the words we hear. *BioRxiv*.

Kösem, A., Gramfort, A., and van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *NeuroImage*, 92:274–284.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. *Science*, 320(5872):110–113.

Lakatos, P., Schroeder, C. E., Leitman, D. I., and Javitt, D. C. (2013). Predictive suppression of cortical excitability and its deficit in schizophrenia. *Journal of Neuroscience*, 33(28):11692–11702.

Lampar, A. and Lange, K. (2011). Effects of temporal trial-by-trial cuing on early and late stages of auditory processing: evidence from event-related potentials. *Attention, Perception, & Psychophysics*, 73(6):1916–1933.

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and cognition*, 69(1):127–137.

Lange, K. (2013). The ups and downs of temporal orienting: a review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Frontiers in Human Neuroscience*, 7.

Lawrance, E. L. A., Harper, N. S., Cooke, J. E., and Schnupp, J. W. H. (2014). Temporal predictability enhances auditory detection. *The Journal of the Acoustical Society of America*, 135(6):EL357–EL363.

Lenth, R. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.1.

Luck, S. J. (2005). An introduction to the event-related potential technique mit press. *Cambridge, Ma*, pages 45–64.

Mento, G. and Vallesi, A. (2016). Spatiotemporally dissociable neural signatures for generating and updating expectation over time in children: A high density-erp study. *Developmental cognitive neuroscience*, 19:98–106.

Miniussi, C., Wilding, E. L., Coull, J. T., and Nobre, A. C. (1999). Orienting attention in time. Modulation of brain potentials. *Brain: A Journal of Neurology*, 122 (Pt 8):1507–1518.

Morillon, B. and Schroeder, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Annals of the New York Academy of Sciences*, 1337(1):26–31.

Muller, D., Judd, C. M., and Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of personality and social psychology*, 89(6):852.

Niemi, P. and Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, 89(1):133–162.

Nobre, A. C., Correa, A., and Coull, J. T. (2007). The hazards of time. *Current Opinion in Neurobiology*, 17(4):465–470.

Nobre, A. C., Sebestyen, G. N., and Miniussi, C. (2000). The dynamics of shifting visuospatial attention revealed by event-related potentials. *Neuropsychologia*, 38(7):964–974.

Nobre, A. C. and van Ede, F. (2018). Anticipated moments: temporal structure in attention. *Nature Reviews Neuroscience*, 19(1):34.

Obleser, J., Henry, M. J., and Lakatos, P. (2017). What do we talk about when we talk about rhythm? *PLoS biology*, 15(9):e2002794.

Paris, T., Kim, J., and Davis, C. (2016). The processing of attended and predicted sounds in time. *Journal of cognitive neuroscience*, 28(1):158–165.

Paris, T., Kim, J., and Davis, C. (2017). Visual form predictions facilitate auditory processing at the n1. *Neuroscience*, 343:157–164.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10(4):437–442.

Phipson, B. and Smyth, G. K. (2010). Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Statistical applications in genetics and molecular biology*, 9(1).

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rimmele, J., Jolsvai, H., and Sussman, E. (2011). Auditory target detection is affected by implicit temporal and spatial expectations. *Journal of cognitive neuroscience*, 23(5):1136–1147.

Rimmele, J. M., Morillon, B., Poeppel, D., and Poeppel, D. (2018). The proactive and flexible sense of timing. *PsyArXiv*, (svf6c).

Rohenkohl, G., Cravo, A. M., Wyart, V., and Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *The Journal of neuroscience*, 32(24):8424–8428.

Rolke, B. and Hofmann, P. (2007). Temporal uncertainty degrades perceptual processing. *Psychonomic Bulletin & Review*, 14(3):522–526.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2):225–237.

Sanders, L. D. and Astheimer, L. B. (2008). Temporally selective attention modulates early perceptual processing: Event-related potential evidence. *Perception & psychophysics*, 70(4):732–742.

Schaal, N. K., Bauer, A.-K. R., and Müllensiefen, D. (2014). Der gold-msi: replikation und validierung eines fragebogeninstrumentes zur messung musikalischer erfahrenheit anhand einer deutschen stichprobe. *Musicae Scientiae*, 18(4):423–447.

Schmidt-Kassow, M., Schubotz, R. I., and Kotz, S. A. (2009). Attention and entrainment: P3b varies as a function of temporal predictability. *Neuroreport*, 20(1):31–36.

Schroeder, C. E. and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1):9–18.

Schütt, H. H., Harmeling, S., Macke, J. H., and Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, 122:105–123.

Schwartze, M., Farrugia, N., and Kotz, S. A. (2013). Dissociation of formal and temporal predictability in early auditory evoked potentials. *Neuropsychologia*, 51(2):320–325.

Sherwell, C., Garrido, M. I., and Cunnington, R. (2017). Timing in predictive coding: The roles of task relevance and global probability. *Journal of cognitive neuroscience*, 29(5):780–792.

Stefanics, G., Hangya, B., HernÃądi, I., Winkler, I., Lakatos, P., and Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *The Journal of neuroscience*, 30(41):13578–13585.

Stekelenburg, J. J. and Vroomen, J. (2007). Neural Correlates of Multisensory Integration of Ecologically Valid Audiovisual Events. *Journal of Cognitive Neuroscience*, 19(12):1964–1973.

Tsagris, M., Athineou, G., Sajib, A., Amson, E., and Waldstein, M. J. (2018). *Directional: Directional Statistics*. R package version 3.3.

van Wassenhove, V. (2016). Temporal cognition and neural oscillations. *Current Opinion in Behavioral Sciences*, 8:124–130.

Vroomen, J. and Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience*, 22(7):1583–1596.

Wassenhove, V. v., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, 102(4):1181–1186.

Wichmann, F. A. and Hill, N. J. (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Attention, Perception, & Psychophysics*, 63(8):1314–1329.

Widmann, A., Schröger, E., and Maess, B. (2015). Digital filter design for electrophysiological data–a practical approach. *Journal of neuroscience methods*, 250:34–46.

Wilsch, A., Henry, M. J., Herrmann, B., Herrmann, C. S., and Obleser, J. (2018). Temporal expectation modulates the cortical dynamics of short-term memory. *Journal of Neuroscience*, pages 2928–17.

Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., and Obleser, J. (2014). Alpha oscillatory dynamics index temporal expectation benefits in working memory. *Cerebral Cortex*, 25(7):1938–1946.

Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., and Obleser, J. (2015). Slow-delta phase concentration marks improved temporal expectations based on the passage of time. *Psychophysiology*, 52(7):910–918.

Woodrow, H. (1914). *The Measurement of Attention (1914)*, volume The Psychological Monographs 17(5).

Wright, B. A. and Fitzgerald, M. B. (2004). The time course of attention in a simple auditory detection task. *Perception & Psychophysics*, 66(3):508–516.

Wyart, V., de Gardelle, V., Scholl, J., and Summerfield, C. (2012). Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron*, 76(4):847–858.

Yamaguchi, S., Tsuchiya, H., and Kobayashi, S. (1994). Electrooencephalographic activity associated with shifts of visuospatial attention. *Brain*, 117(3):553–562.