Title: Synergistic yet dissociable roles of temporal and spectral predictions in auditory detection

Abbreviated title: Synergistic yet dissociable roles of temporal and spectral predictions in auditory detection

Authors: Johannes Wetekam*[,1], Chloé Dumeige[1,2], Manon Beurtey[1,3], Sophie Herbst*[,1]

1. Cognitive Neuroimaging Unit (UNICOG), NeuroSpin
   (CEA, INSERM, Université Paris-Saclay)
   Centre d'études de Saclay, Bâtiment 145,
   91191 Gif-sur-Yvette, France
2. Institute of Biology of the ENS (IBENS), École Normale Supérieure,
   46 Rue d'Ulm,
   75005 Paris, France
3. Laboratory for Interdisciplinary Evaluation of Public Policies (LIEPP), Sciences Po
   27 Rue Saint-Guillaume
   75007 Paris, France

*Corresponding authors: Johannes Wetekam: johannes.wetekam@gmail.com, Sophie Herbst: ksherbst@gmail.com

Number of pages: 32, Number of figures: 3, Number of words Abstract: 235, Number of words Introduction: 659, Number of words Discussion: 900

Acknowledgements

This work was funded by the French National Research Agency (ANR-23-CE28-0006-01).

Conflict of Interest

The authors declare no competing interests.

Abstract

Predictive processing enables the brain to anticipate both *when* and *what* sensory events will occur. Although temporal and feature-based predictions are both known to facilitate perception, their functional contributions and interaction remain poorly understood, particularly under natural listening conditions where predictions must be inferred from implicit statistical regularities rather than explicit cues or rhythms. Here, we orthogonalised temporal and spectral predictability in a non-rhythmic auditory detection task and used a signal-detection framework to dissociate perceptual and decisional processes. Temporal predictions primarily increased response readiness: predictable target timing led to faster responses and more hits but also more false alarms, consistent with a liberal shift in response criterion. Spectral predictions, in contrast, selectively enhanced perceptual sensitivity, mainly by reducing false alarms, without a comparable speeding of responses. Crucially, when both predictive dimensions were available, they interacted synergistically to maximise perceptual sensitivity, revealing how distinct predictive mechanisms combine to optimise behaviour. Furthermore, temporal and spectral variability were internalised differently: performance adapted strongly to the distribution of foreperiod durations but remained largely stable across the spectral distribution, suggesting that the brain encodes temporal and spectral statistics in fundamentally different ways when forming predictions under uncertainty. Together, these findings demonstrate that temporal and spectral predictions influence distinct computational components of auditory decision-making and are integrated synergistically to improve perception under uncertainty. The results provide a mechanistic framework of how multidimensional predictions jointly shape sensory processing in natural listening environments.

1. Introduction

Detecting behaviourally relevant sounds under acoustic uncertainty is a central challenge for the auditory system. In natural environments, sounds are embedded in noise, occur at irregular times and vary in their spectral content. One way the auditory system can reduce this uncertainty is by exploiting regularities in the sensory input to generate predictions about forthcoming events. In audition, stimulus timing ("when") and stimulus content ("what") are two particularly important sources of predictive information. A substantial body of work has shown that temporal and spectral predictions can each facilitate auditory processing (for reviews see Arnal and Giraud, 2012; Heilbron and Chait, 2018; Nobre and van Ede, 2018). However, these two predictive dimensions have largely been studied in isolation, and only a limited number of studies have examined their combined effects. Available evidence suggests that temporal and spectral predictions engage partially dissociable neural mechanisms. Temporal predictions have repeatedly been associated with increased response readiness, reflected in reduced reaction times (Lange, 2008; Hsu et al., 2013; Morillon et al., 2016), as well as modulatory effects on both sensory and motor-related neural processing (Morillon et al., 2015; Auksztulewicz et al., 2018; Rimmele et al., 2018). These effects of temporal predictions have been interpreted as reflecting stimulus-unspecific changes in neural gain (Auksztulewicz et al., 2018, 2019). Spectral predictions, in contrast, are less strongly linked to motor-related activity and instead primarily modulate sensory processing and feature-specific neural responses (Lange, 2008; Schwartze et al., 2013; Carbajal and Malmierca, 2018; Demarchi et al., 2019). Importantly, temporal and spectral predictions often enhance performance in an interactive manner, such that the benefit of one dimension depends on the presence of the other (Hsu et al., 2013; Morillon et al., 2016; Wollman and Morillon, 2018). This combination of partially separable mechanisms and synergistic behavioural benefits raises a central question: How do temporal and spectral predictions interact at the mechanistic level to shape auditory perception?

A major limitation of existing work is that predictions are typically induced using rhythmic stimulation or explicit cues (Lange, 2008; Schwartze et al., 2011, 2013; Hsu et al., 2013; Lakatos et al., 2013; Morillon et al., 2016; Auksztulewicz et al., 2018; Wollman and Morillon, 2018). While such approaches generate strong expectations, they do not require predictions to be inferred under uncertainty. In contrast, natural listening environments are non-isochronous and lack explicit predictive signals, such that predictive information must be extracted implicitly from the statistical regularities of the auditory input. Although some studies have examined auditory predictions under such statistical uncertainty, they have not clearly separated the functional contributions of temporal and spectral predictions (Herbst and Obleser, 2017, 2019). As a result, the role of these predictive dimensions in auditory perception remains incompletely understood. This gap is particularly evident at the level of sound detection, the most basic form of auditory perception. Detection performance is characterised as the joint outcome of separable perceptual and decision processes. Specifically, it reflects the observer's perceptual sensitivity to the signal and the response criterion used to decide for the presence of the signal (Green and Swets, 1966; Summerfield and de Lange, 2014). Prediction-related improvements in auditory detection are often reported without a clear distinction between these components, making it difficult to determine how different forms of predictive information influence behaviour at a mechanistic level. Disentangling the effects of predictions on these processes is therefore essential for understanding the functional role of temporal and spectral predictions as well as their interaction under sensory uncertainty.

Here, we provide a detailed assessment of predictive benefits using a non-rhythmic auditory detection task, in which temporal and spectral predictions could be derived from an orthogonal variation of the of target onset times (foreperiods) and target frequencies (Fig. 1). Inducing predictions based on statistical distributions enabled us to move beyond binary

contrasts between predictable and unpredictable conditions, and to model detection performance relative to the underlying probability of target occurrence. This study thus offers novel insights into how the auditory system encodes and exploits multidimensional predictive information.

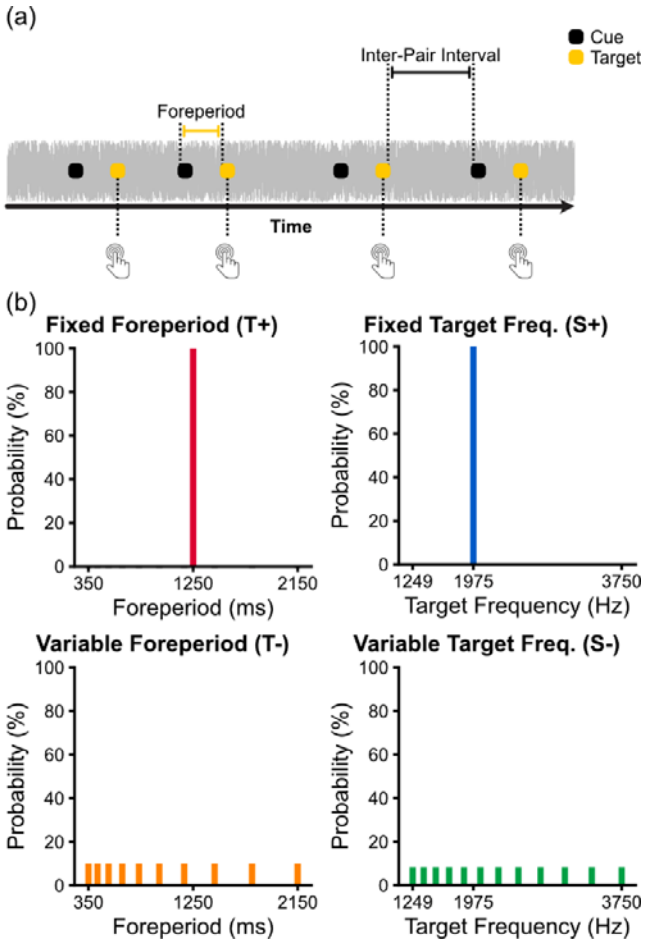


Figure 1 – Foreperiod paradigm and predictability conditions. (a) Schematic illustration of the foreperiod paradigm used in this study. Cue-target tone pairs were presented against continuous background noise. The time interval between the cue and target defines the foreperiod. Inter-pair intervals were jittered to prevent the emergence of isochronous stimulation patterns. In 10 % of trials, the target was omitted (catch trials). (b) Probability distributions of foreperiods and target frequencies. Temporal predictability was manipulated by keeping the foreperiod either fixed (top-left, red; T+) or variable (bottom-left, orange; T−) within a block. Spectral predictability was manipulated by keeping the target frequency either fixed (top-right, blue; S+) or variable (bottom-right, green; S−). Temporal and spectral predictability manipulations were crossed, leading to four different conditions: T+S+, T+S-, T-S+, T-S-.

2. Material and Methods

Participants

For this study, a total of 41 participants aged between 23 and 53 years (27 females, 14 males; mean age = 35.5 years, SD = 7.8 years) were recruited via online announcements distributed through the PRISME platform mailing list (Institut du Cerveau, Paris). Exclusion criteria comprised self-reported auditory impairments and colour blindness. All participants provided written informed consent prior to participation and received financial compensation (20 € in cash per session). The study was approved by the local ethics committee of Université Paris-Saclay (Reference: CER-Paris-Saclay-2022-004/A3) and was conducted in accordance with the World Medical Association (WMA) Declaration of Helsinki.

Stimuli and data acquisition

All experiments were conducted at the Paris Brain Institute PRISME Core Facility (RRID:SCR_026394), Paris, France. A classroom equipped with 12 individual computers enabled parallel data acquisition within a single session. Stimulus presentation, visual display, and behavioural data collection were implemented using the Python library Expyriment. Auditory stimuli were delivered through Logitech G35 headphones at a sampling rate of 44100 Hz.

The experiment was based on a foreperiod paradigm combined with an auditory detection task (Fig. 1a). Sound sequences were presented against a continuous background noise. The background noise consisted of low-pass filtered white noise (cut-off frequency: 5 kHz) and was continuously played throughout each block. Each trial consisted of a pair of tones, with the first tone serving as a cue and the second tone as a target sound. Cue tones had a duration of 250 ms and a fixed frequency of 1975 Hz. Target tones had a duration of 50 ms and their frequency depended on the experimental condition (see below). All tones were pure sine waves and included 10 ms cosine-shaped rise and fall ramps.

Prior to the foreperiod task, the intensity of both the background noise and the tones was individually determined for each participant using staircase-based procedures. First, the background noise level was adjusted relative to each participant's individual sensation level using an ascending and descending method of limits. The final noise level was fixed at 40 dB above the individual sensation level. Next, hearing thresholds for pure tones were measured within this background noise using an adaptive staircase procedure. A 2-down-1-up rule was applied, such that tone intensity was decreased after two consecutive detections and increased after a single miss. This procedure converges on a detection level of approximately 70 % correct responses. Thresholds were estimated separately for three tone frequencies (1250, 1975, and 3750 Hz) and linearly interpolated for the remaining frequencies used in the experiment. During the foreperiod task, cue tones were presented at a level 20 dB above the individual hearing threshold, whereas target tones were presented at hearing threshold level within the background noise. This resulted in clearly audible cues followed by difficult-to-detect target tones.

In 10 % of the tone pairs, the cue was not followed by a target sound (catch trials). Participants were instructed to press the spacebar as quickly as possible whenever they detected a target tone, while avoiding responses on trials in which no target tone was presented. Responses made before target onset were classified as premature responses. Following target onset, participants had a response window of 1 s. After each tone pair, visual feedback indicating the response outcome (too early, hit, miss, correct rejection or false alarm) was presented on the screen. Feedback was displayed immediately following the participant's response or, in the absence of a response, after the 1 s response window had elapsed. A short training sequence of 40 trials preceded the main experimental session.

Temporal and spectral predictability were manipulated using a 2 × 2 blockwise design. Temporal predictability was varied by keeping the foreperiod duration either fixed at 1250 ms

within a block (high temporal predictability, T+) or randomly drawing it from a log-spaced uniform distribution ranging from 350 to 2150 ms (350, 428, 523, 641, 784, 959, 1173, 1436, 1757, and 2150 ms; low temporal predictability, T-). Spectral predictability was manipulated via the target frequency, which was either fixed at 1975 Hz within a block (high spectral predictability, S+) or randomly drawn from a log-spaced uniform distribution ranging from 1249 to 3750 Hz (1249, 1381, 1526, 1686, 1863, 2059, 2275, 2514, 2779, 3071, 3393, and 3750 Hz; low spectral predictability, S-; Fig. 1b). Crossing spectral predictability (S+ / S-) with temporal predictability (T+ / T−) yielded four experimental conditions (T+S+, T+S-, T-S+, and T-S-), one of which was presented per block. Each condition was presented in three blocks of 100 trials, yielding a total of 12 blocks (1200 trials) per participant. Block order was randomised across participants. Trials were separated by a jittered inter-trial interval drawn from a truncated normal distribution (mean = 1375 ms, SD = 62.5 ms, range = 1250–1500 ms), resulting in a non-isochronous stimulation pattern.

Crucially, participants were not explicitly informed about the predictability manipulations. Their only instruction was to report the occurrence of target sounds as quickly and accurately as possible.

Data processing and statistical evaluation

Data analysis was conducted using custom scripts written in Python and R.

Behavioural performance was quantified using four dependent variables derived from participants' key presses. Reaction time was defined as the latency between target onset and the key press and was evaluated on detected signal trials (hits) only. Hit rate was defined as the proportion of signal trials in which a response was recorded within the response window. False alarm rate was defined as the proportion of catch trials in which a response was recorded. In accordance to Hautus (1995), both hit rate and false alarm rate were log-linear

corrected to ensure that they remained strictly between 0 and 1. Sensitivity was computed as an odds-ratio- (OR-) based discriminability measure derived from hit rate (*HR*) and false-alarm rate (*FA*):

$$HR = \frac{hits + 0.5}{signal\ trials + 1}$$

$$FA = \frac{false\ alarms + 0.5}{catch\ trials + 1}$$

$$Sensitivity = \frac{HR/(1 - HR)}{FA/(1 - FA)}$$

In contrast to *d'*, this OR-based sensitivity measure can be obtained directly from the fitted mixed-effects model, allowing main effects, interactions, and simple effects to be estimated and statistically tested within a single modelling framework.

Exclusion criteria were applied at both the participant and trial level. At the participant level, participants were excluded if they did not complete the full experiment (fewer than 12 blocks; $n = 3$). From the remaining sample, participants with extremely low or high overall hit rates were excluded to ensure meaningful task engagement: three participants were excluded for hit rates below 40 %, and one participant was excluded for hit rates above 95 %. These exclusions resulted in a final sample of 34 participants included in the statistical analyses. At the trial level, premature responses (key presses occurring before target onset) were identified and labelled as “too early” and excluded from all analyses. In addition, hit trials with reaction times shorter than 100 ms were removed, as such responses are typically considered biologically implausible. Because temporal predictability could in principle lead to very rapid responses, all analyses were repeated without applying this reaction-time exclusion criterion to ensure that it did not affect the reported results. This robustness check did not change the pattern or statistical significance of the findings.

Statistical analyses were conducted using mixed-effects models implemented in R. Reaction time data were analysed using linear mixed-effects models (LMMs), whereas response-based measures – hit rate, false alarm rate, and sensitivity – were analysed using generalised linear mixed-effects models (GLMMs), as implemented in the *lme4* package (Bates et al., 2015).

To assess the main effects of temporal (T) and spectral predictability (S), as well as their interaction and associated simple effects, RTs were analysed using LMMs fitted to correct responses (hits) on signal trials only. Models included fixed effects of T, S, and their interaction (T×S), along with random intercepts and random slopes for T and S at the participant level. Response-based measures (hit rate, false alarm rate and sensitivity) were analysed using GLMMs. Hit rate and false alarm rate were modelled jointly within a single framework by including signal presence (signal vs. catch trials), temporal predictability (T), spectral predictability (S), and their interactions (T×S) as fixed effects. This approach enabled the simultaneous estimation of hit and false alarm responses and the derivation of sensitivity from the same model. Similar to the LMMs, the GLMMs included random intercepts and random slopes for temporal and spectral predictability at the participant level. Planned contrasts were used to estimate main effects, interaction effects, and simple effects, with p-values adjusted for multiple comparisons using false discovery rate (FDR) correction (Benjamini and Hochberg, 1995). Corrected p-values lower than 0.05 were considered significant.

A second set of analyses examined how performance varied across the distributions of foreperiod durations and target frequencies. In this analysis, performance was quantified using reaction time and hit rate only. Sensitivity and false alarm rate were not analysed in this context, as catch trials do not provide information about foreperiod duration or target frequency. Reaction time was analysed using LMMs, and hit rate was analysed using GLMMs, with foreperiod duration or target frequency entered as continuous predictors

modelled using second-order, mutually orthogonal polynomials. Temporal and spectral predictability were included as fixed effects in the models, and random intercepts were included at the participant level. Random slopes were not included, as models with random slopes did not converge and provided no improvement in model fit. For the model estimates, the predictability factor not directly related to the distribution of interest (i.e. spectral predictability for foreperiod analyses and temporal predictability for target-frequency analyses) was collapsed across its two levels. To quantify the effect size of the curvature of performance across the variable foreperiod and target frequency distributions, a curvature effect size was computed as the difference between the model-predicted minimum and maximum performance values across each distribution. The direction of this effect was defined such that improved performance around the distribution midpoint relative to the edges resulted in a positive effect (or an odds ratio greater than 1), whereas worse performance at the midpoint relative to the edges resulted in a negative effect (or an odds ratio smaller than 1).

Pilot experiment with narrower frequency range and no catch trials

Prior to the main experiment, we conducted an independent pilot experiment with another 38 participants aged between 18 and 45 years (25 females, 13 males; mean age = 28.1 years, SD = 7.2 years) to assess behavioural effects of temporal and spectral predictability. The pilot experiment was identical to the main experiment in terms of recruitment procedures, exclusion criteria, task structure, stimulus timing, and analysis approach. The key differences were that the pilot used a substantially narrower target-frequency range (1724-2224 Hz) in the variable spectral condition and did not include catch trials. The absence of catch trials prevented the calculation of sensitivity measures and resulted in generally higher hit rates, likely leading to ceiling effects. For this reason, the pilot experiment was not suitable for addressing the main questions of the present study and motivated the design of the main

experiment. Here, we use the pilot data only to assess whether performance across variable foreperiod and target-frequency distributions shows a similar qualitative pattern when the spectral range is reduced. Pilot data were analysed using the same mixed-effects modelling approach as in the main experiment, focusing on how hit rate and reaction time varied as a function of foreperiod duration and target frequency under variable predictability. Results of the pilot experiment are reported in Supplementary Figure 1 and are referenced in the Discussion solely to evaluate the robustness of the distribution-level effects.

3. Results

Temporal and spectral predictions benefit behaviour synergistically

The current study examined the effects of temporal and spectral predictions on different aspects of auditory performance: hit rate, false alarm rate, sensitivity and reaction time.

Hit rate: Analysis of hit rate revealed a significant main effect of temporal predictability, as well as a significant interaction between temporal and spectral predictability (T: $p < 0.001$; S: $p = 0.055$; T×S: $p < 0.001$; Fig. 2a-b). This interaction reflected a larger increase in hit rate due to temporal predictions when spectral predictions were present than when they were absent (T at S- and S+: $p < 0.001$). Spectral predictions, in contrast, had no significant effect on hit rate in the absence of temporal predictions (S at T-: $p = 0.811$; S at T+: $p = 0.0018$; Fig. 2c).

False alarm rate: False alarm rate was significantly modulated by temporal predictability, spectral predictability, and their interaction (T: $p < 0.001$; S: $p < 0.001$; T×S: $p = 0.024$; Fig. 2d). Importantly, the two prediction types had opposite effects: temporal predictions were associated with an increase in false alarm rate, whereas spectral predictions were associated with a decrease. In addition, a significant interaction indicated a reduction in false alarm rate when both prediction types were present together compared to when they occurred separately (T at S- and S+: $p < 0.001$; S at T-: $p = 0.012$; S at T+: $p < 0.001$; Fig. 2e-f).

Sensitivity: Sensitivity, which integrates changes in both hit rate and false alarm rate, was significantly affected by temporal predictability, spectral predictability, and their interaction (T, S and T×S: $p < 0.001$; Fig. 2g-h). Simple effects analyses revealed that temporal predictions improved sensitivity only when spectral predictions were present (T at S-: $p = 0.2636$; T at S+: $p < 0.001$). By contrast, spectral predictions also improved sensitivity when present in isolation, although this effect was weaker than when temporal predictions were available as well (S at T- and T+: $p < 0.001$; Fig. 2i).

Reaction time: In contrast to the response-based measures described above, reaction time was significantly reduced by temporal predictability, whereas spectral predictability and the temporal-spectral interaction had no significant effect (T: $p < 0.001$; S: $p = 0.3786$; T×S: $p = 0.1637$; Fig. 2j-l).

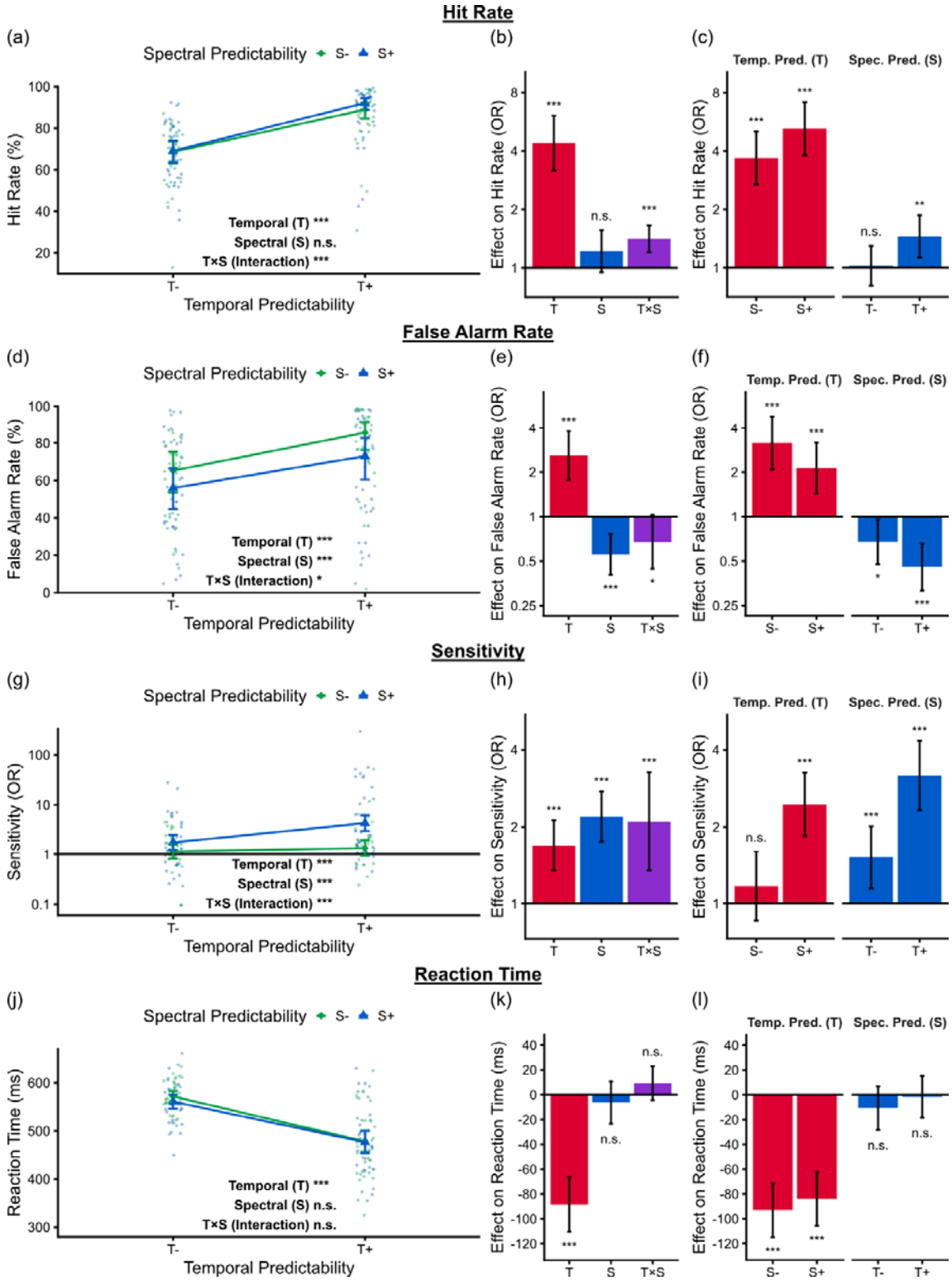


Figure 2 - Temporal and spectral predictions affect auditory performance differentially and complementarily. (a) Hit rate (%) as a function of temporal predictability (x-axis) and spectral predictability (colour-coded). Symbols show model-estimated marginal means for each condition; whiskers indicate 95 % confidence intervals. Small dots represent individual participant means. (b) Main effects of temporal (T) and spectral (S) predictability, as well as their interaction (T×S), on hit rate expressed as odds ratios (OR). An odds ratio greater than 1 indicates a higher likelihood of a hit when predictability is high compared to low, whereas an odds ratio smaller than 1 indicates the opposite pattern. (c) Simple effects of temporal and spectral predictability on hit rate (OR), illustrating how the effect of one predictability factor depends on the level of the

other. (d-f) Same as panels (a–c), but for false alarm rate. (g-i) Same as panels (a–c), but for sensitivity. (j-l) Same as panels (a-c) but for reaction time.

Distinct encoding strategies for temporal and spectral distributions

The dissociable effects of temporal and spectral predictions motivated a closer examination of how observers encode and exploit variability along these two dimensions. Performance under high variability can reveal whether, and in what way, observers encode and adapt to the statistical properties of each distribution. Therefore, we next examined how performance varied across foreperiods and target frequencies in the variable conditions. Specifically, we analysed hit rate and reaction time as a function of foreperiod duration or target frequency, respectively, using GLMMs for hit rate and LMMs for reaction time. In these models, foreperiod duration or target frequency was entered as a continuous predictor, allowing the assessment of non-linear changes in performance across each distribution. Hit rate exhibited a pronounced inverse U-shaped relationship with foreperiod duration in the variable temporal condition (Fig. 3a). Notably, the modulation was slightly asymmetric, with peak performance shifted towards longer intermediate foreperiods rather than being centred exactly on the distribution midpoint. The U-shape pattern was supported by a significant quadratic term in the GLMM and reflected higher performance around the centre of the foreperiod distribution compared to its edges (curvature: $p < 0.001$). In contrast, hit rate showed a significant but markedly smaller U-shaped relationship as a function of target frequency, indicating slightly reduced performance around the centre of the spectral distribution relative to the edges (curvature: $p < 0.001$; Fig. 3b). A direct comparison of curvature effect sizes revealed that, although both effects differed significantly from zero, adaptation to the distribution midpoint was substantially stronger for foreperiod duration than for target frequency (foreperiod and target frequency: $p < 0.001$; Fig. 3c). Reaction time showed a similar pattern. Reaction times were shorter around the centre of the foreperiod distribution than at its edges, reflected in a U-shaped relationship in the variable temporal condition (curvature: $p < 0.001$; Fig. 3d). Similar to hit rate, the modulation was asymmetric, with the fastest responses occurring at longer

intermediate foreperiods rather than exactly at the distribution midpoint. In contrast, no significant curvature was observed for reaction time as a function of target frequency; reaction times remained stable across the spectral distribution (curvature: $p = 0.629$; Fig. 3e). This dissociation is also captured by the curvature effect size, which revealed a pronounced reaction time benefit around the foreperiod distribution centre, but no reliable benefit around the midpoint of the target frequency distribution (foreperiod: $p < 0.001$; target frequency: $p = 0.615$; Fig. 3f).

A direct comparison of model-estimated performance at the fixed foreperiod and target-frequency values revealed no significant difference in hit rate between the fixed and variable temporal conditions at the fixed foreperiod duration (T-: HR = 88.80 ± 1.51 %; T+: HR = 88.17 ± 1.57 %; $p = 0.074$; Fig. 3a). In contrast, hit rate was slightly higher in the fixed spectral condition than in the variable spectral condition when evaluated at the fixed target frequency (S-: HR = 78.88 ± 2.10 %; S+: HR = 81.95 ± 1.86 %; $p < 0.001$; Fig. 3b). Reaction time showed a different pattern. When evaluated at the foreperiod used in the fixed temporal condition, reaction times were slower in the fixed than in the variable temporal condition (T-: RT = 436.58 ± 10.57 ms; T+: RT = 476.66 ± 10.15 ms; $p < 0.001$; Fig. 3d). For target frequency, reaction times were slightly faster in the fixed spectral condition compared to the variable spectral condition at the fixed frequency value (S-: HR = 526.13 ± 8.44 ms; S+: HR = 518.74 ± 8.26 ms; $p = 0.025$; Fig. 3e).

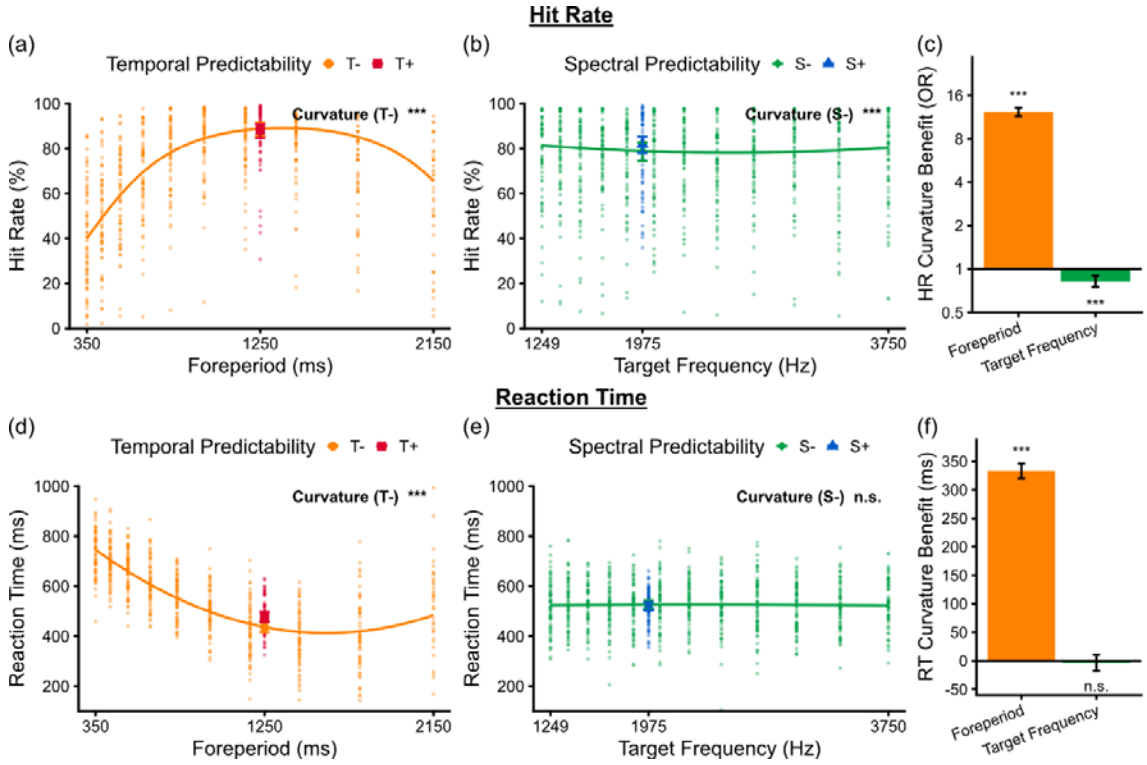


Figure 3 – Distinct encoding strategies for temporal and spectral distributions. (a) Hit rate (%) as a function of foreperiod duration (ms; x-axis) and temporal predictability (colour-coded). The solid line shows model-predicted performance across the variable foreperiod distribution (T-), and dots represent individual participant means. The label "Curvature (T-)" indicates a non-linear change in performance across the variable foreperiod distribution (T-); asterisks denote the statistical significance of this non-linear (quadratic) effect. Symbols show model-estimated marginal means for each condition at the fixed foreperiod duration, with whiskers indicating 95 % confidence intervals. (b) Same as panel (a), but showing hit rate (%) as a function of target frequency (Hz; x-axis) and spectral predictability (colour-coded). (c) Effect size (OR) of the curvature effect on hit rate in the low-predictability condition. An OR greater than 1 indicates better performance around the centre of the distribution compared to its edges, whereas an OR smaller than 1 indicates worse performance around the centre relative to the edges. (d-f) Same as panels (a-c), but for reaction time.

4. Discussion

Synergistic benefits of temporal and spectral predictions on auditory performance

The present results demonstrate that temporal and spectral predictions shape the perception and decision processes underlying auditory detection in distinct ways. Temporal predictability alone increased response readiness: when target onsets were predictable, participants responded faster and produced more hits, but also more false alarms, in line with a liberalisation of response criterion at predicted moments in time. Spectral predictability, in contrast, selectively enhanced perceptual sensitivity, mainly by reducing false alarms, without producing a comparable speeding of responses. Crucially, combining both predictive dimensions yielded a synergistic increase in sensitivity, indicating that temporal and spectral predictions are integrated to optimize performance. Building on previously reported interactions between temporal and spectral predictions (Costa-Faidella et al., 2011; Hsu et al., 2013; Schwartze et al., 2013; Morillon et al., 2016), partitioning performance into sensitivity and criterion unravels the signal-detection mechanisms driving their synergy.

Furthermore, the present findings extend the relevance of temporal predictions beyond rhythmic stimulation and suggest a general account of how the auditory system exploits predictive structure to improve perception and guide decisions under uncertainty: Temporal predictions allow to selectively focus on the most relevant moments in time, often described as temporal orienting of attention (Jones, 1976; Nobre and van Ede, 2018; Denison et al., 2021). Behaviourally, this manifests as faster response times and a more liberal response criterion at predicted moments. Spectral predictions, in contrast, enable the preparation of frequency-selective neural populations, consistent with enhanced perceptual sensitivity through reduced false alarms (Okamoto et al., 2007; De Martino et al., 2015). When both predictive dimensions are available, relevant neural populations can be tuned at the

appropriate moments in time, resulting in maximal perceptual sensitivity (Jaramillo and Zador, 2010; Lakatos et al., 2013).

Importantly, this interpretation suggests that previously reported sensitivity gains attributed to temporal prediction alone may have partly depended on interaction effects with spectral predictability, particularly in paradigms using narrow frequency ranges (Lawrance et al., 2014; Herbst and Obleser, 2019). Indeed, specific temporal preparation is contingent upon feature-specific knowledge: only by anticipating both dimensions can the brain prime the relevant sensory channels. This interpretation is consistent with neurophysiological evidence indicating partly distinct mechanisms: temporal predictions have been associated with widespread gain modulations (Auksztulewicz et al., 2019), whereas content-based predictions are more closely linked to changes in feature-selective sensory representations (Kok et al., 2012, 2014, 2017; Demarchi et al., 2019).

A similar division of labour between two predictive dimensions has been described in the visual domain, where predictions about stimulus probability and stimulus location exert separable effects on response processes: probabilistic cues primarily bias response criterion, whereas location cues selectively enhance perceptual sensitivity (Bashinski and Bacharach, 1980; Reynolds et al., 2000; Carrasco, 2006; Wyart et al., 2012; Summerfield and de Lange, 2014). The present findings suggest that the conceptual principle of complementary predictive effects may generalise across sensory domains.

Differential representation of temporal and spectral uncertainty

Our findings demonstrate a qualitative dissociation in how implicit temporal and spectral variability shape behaviour. When foreperiods varied, performance showed a pronounced inverted u-shaped modulation: Detection was improved around the centre of the temporal distribution relative to its edges. Notably, this profile was tilted towards longer intermediate foreperiods rather than being perfectly symmetric. This asymmetry suggests that temporal

predictions reflect an integration of multiple statistical parameters, consistent with joint sensitivity to the probability density function (Grabenhorst et al., 2019) and the hazard rate (Nobre et al., 2007; Herbst et al., 2018). In addition, integration of temporal statistics across predictive and non-predictive blocks may have contributed to this pattern, as the intermediate foreperiod (1250 ms) was overall the most probable one. Importantly, this centre-weighted modulation was present not only for reaction times but also for hit rates. Thus, under temporal uncertainty, predictions appear to support not only anticipatory response readiness (Grabenhorst et al., 2019, 2026) but also perceptual preparation (Jones, 1976; Nobre and van Ede, 2018; Denison et al., 2021).

In contrast, performance across the spectral distribution showed no comparable centre enhancement and remained largely stable. The same pattern was observed in a pilot experiment using a substantially narrower frequency range (1724–2224 Hz), suggesting that this near-flat spectral profile is largely invariant to the frequency range (Supp. Fig. 1). Several interpretations may account for this dissociation. One possibility is that frequency statistics were represented more veridically compared to temporal statistics: under a uniform distribution, optimal allocation of resources would yield relatively uniform performance across frequencies. Under this hypothesis, participants are able to evenly distribute attention across frequency channels under uncertainty and focus on a particular frequency in the predictive condition to enhance sensitivity. An alternative interpretation is that participants relied on a coarse associative strategy, treating spectral predictability primarily as a binary contingency (fixed versus variable target frequency) rather than exploiting detailed frequency statistics. Future work manipulating the shape of the target-frequency distribution – for example, using Gaussian or bimodal distributions – could help distinguish between genuine statistical learning and coarse contingency learning.

Conclusion

Our results reveal a simple organising principle: temporal and spectral predictions jointly improve performance yet do so via complementary roles and distinct encoding strategies. Temporal predictions alone primarily modulate response readiness, whereas spectral predictions are necessary to selectively enhance perceptual sensitivity at predicted moments. By dissociating these contributions, the present results reconcile behavioural synergy with mechanistic differentiation and provide a general framework for how the auditory system exploits statistical structure to optimise perception under uncertainty.

Author contributions

Study design: Sophie Herbst, Chloé Dumeige, Manon Beurtey, Johannes Wetekam

Experiments: Johannes Wetekam, Chloé Dumeige

Analysis, and original draft of the manuscript: Johannes Wetekam, Sophie Herbst

Review and editing of the original draft: Sophie Herbst, Chloé Dumeige, Manon Beurtey, Johannes Wetekam

Abbreviations

FA:	False alarm rate
FDR:	False discovery rate
GLMM:	Generalised linear mixed-effects model
HR:	Hit rate
LMM:	Linear mixed-effects model
OR:	Odds ratio
RT:	Reaction time
S:	Spectral predictability
SD:	Standard deviation
T:	Temporal predictability

References

Arnal LH, Giraud A-L (2012) Cortical oscillations and sensory predictions. Trends in Cognitive Sciences 16:390–398.

Auksztulewicz R, Myers NE, Schnupp JW, Nobre AC (2019) Rhythmic Temporal Expectation Boosts Neural Activity by Increasing Neural Gain. J Neurosci 39:9806–9817.

Auksztulewicz R, Schwiedrzik CM, Thesen T, Doyle W, Devinsky O, Nobre AC, Schroeder CE, Friston KJ, Melloni L (2018) Not All Predictions Are Equal: "What" and "When" Predictions Modulate Activity in Auditory Cortex through Different Mechanisms. J Neurosci 38:8680–8693.

Bashinski HS, Bacharach VR (1980) Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. Perception & Psychophysics 28:241–248.

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

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57:289–300.

Carbajal GV, Malmierca MS (2018) The Neuronal Basis of Predictive Coding Along the Auditory Pathway: From the Subcortical Roots to Cortical Deviance Detection. Trends in Hearing 22:1–33.

Carrasco M (2006) Covert attention increases contrast sensitivity: psychophysical, neurophysiological and neuroimaging studies. In: Progress in Brain Research, pp 33–70. Elsevier. Available at: https://www.sciencedirect.com/science/chapter/bookseries/abs/pii/S0079612306540038 [Accessed January 22, 2026].

Costa-Faidella J, Baldeweg T, Grimm S, Escera C (2011) Interactions between "what" and "when" in the auditory system: Temporal predictability enhances repetition suppression. Journal of Neuroscience 31:18590–18597.

De Martino F, Moerel M, Ugurbil K, Goebel R, Yacoub E, Formisano E (2015) Frequency preference and attention effects across cortical depths in the human primary auditory cortex. Proceedings of the National Academy of Sciences 112:16036–16041.

Demarchi G, Sanchez G, Weisz N (2019) Automatic and feature-specific prediction-related neural activity in the human auditory system. Nat Commun 10:3440.

Denison RN, Carrasco M, Heeger DJ (2021) A dynamic normalization model of temporal attention. Nat Hum Behav 5:1674–1685.

Grabenhorst M, Michalareas G, Maloney LT, Poeppel D (2019) The anticipation of events in time. Nat Commun 10:5802.

Grabenhorst M, Poeppel D, Michalareas G (2026) The anticipation of imminent events is time-scale invariant. Proceedings of the National Academy of Sciences 123:e2518982123.

Green DM, Swets JA (1966) Signal detection theory and psychophysics. Oxford, England: John Wiley.

Hautus MJ (1995) Corrections for extreme proportions and their biasing effects on estimated values ofd. Behavior Research Methods, Instruments, & Computers 27:46–51.

Heilbron M, Chait M (2018) Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? Neuroscience 389:54–73.

Herbst SK, Fiedler L, Obleser J (2018) Tracking Temporal Hazard in the Human Electroencephalogram Using a Forward Encoding Model. eNeuro 5 Available at: https://www.eneuro.org/content/5/2/ENEURO.0017-18.2018 [Accessed October 23, 2024].

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

Herbst SK, Obleser J (2019) Implicit temporal predictability enhances pitch discrimination sensitivity and biases the phase of delta oscillations in auditory cortex. NeuroImage 203:116198.

Hsu Y-F, Hämäläinen JA, Waszak F (2013) Temporal expectation and spectral expectation operate in distinct fashion on neuronal populations. Neuropsychologia 51:2548–2555.

Jaramillo S, Zador A (2010) Auditory cortex mediates the perceptual effects of acoustic temporal expectation. Nat Prec:1–1.

Jones MR (1976) Time, our lost dimension: Toward a new theory of perception, attention, and memory. Psychological Review 83:323–355.

Kok P, Failing MF, de Lange FP (2014) Prior expectations evoke stimulus templates in the primary visual cortex. J Cogn Neurosci 26:1546–1554.

Kok P, Jehee JFM, de Lange FP (2012) Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75:265–270.

Kok P, Mostert P, de Lange FP (2017) Prior expectations induce prestimulus sensory templates. Proceedings of the National Academy of Sciences 114:10473–10478.

Lakatos P, Musacchia G, O'Connell MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761.

Lange K (2008) Brain correlates of early auditory processing are attenuated by expectations for time and pitch. Brain and Cognition 69:127–137.

Lawrance ELA, Harper NS, Cooke JE, Schnupp JWH (2014) Temporal predictability enhances auditory detection. J Acoust Soc Am 135:EL357–EL363.

Morillon B, Hackett TA, Kajikawa Y, Schroeder CE (2015) Predictive motor control of sensory dynamics in auditory active sensing. Curr Opin Neurobiol 31:230–238.

Morillon B, Schroeder CE, Wyart V, Arnal LH (2016) Temporal Prediction in lieu of Periodic Stimulation. J Neurosci 36:2342–2347.

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

Nobre AC, van Ede F (2018) Anticipated moments: temporal structure in attention. Nat Rev Neurosci 19:34–48.

Okamoto H, Stracke H, Wolters CH, Schmael F, Pantev C (2007) Attention Improves Population-Level Frequency Tuning in Human Auditory Cortex. J Neurosci 27:10383–10390.

Reynolds JH, Pasternak T, Desimone R (2000) Attention Increases Sensitivity of V4 Neurons. Neuron 26:703–714.

Rimmele JM, Morillon B, Poeppel D, Arnal LH (2018) Proactive Sensing of Periodic and Aperiodic Auditory Patterns. Trends in Cognitive Sciences 22:870–882.

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

Schwartze M, Rothermich K, Schmidt-Kassow M, Kotz SA (2011) Temporal regularity effects on pre-attentive and attentive processing of deviance. Biological Psychology 87:146–151.

Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and computational mechanisms. Nat Rev Neurosci 15:745–756.

Wollman I, Morillon B (2018) Organizational principles of multidimensional predictions in human auditory attention. Sci Rep 8:13466.

Wyart V, Nobre AC, Summerfield C (2012) Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. Proc Natl Acad Sci U S A 109:3593–3598.

Figure legends

Figure 1: Foreperiod paradigm and predictability conditions. (a) Schematic illustration of the foreperiod paradigm used in this study. Cue-target pairs were presented against continuous background noise. The time interval between cue and target defines the foreperiod. Inter-pair intervals were jittered to prevent the emergence of a rhythmic stimulation pattern. In 10 % of trials, the target was omitted (catch trials). (b) Schematic representation of the four predictability conditions. Temporal predictability was manipulated by keeping the foreperiod either fixed (top-left, red; T+) or variable (bottom-left, orange; T−) within a block. Spectral predictability was manipulated by keeping the target frequency either fixed (top-right, blue; S+) or variable (bottom-right, green; S−).

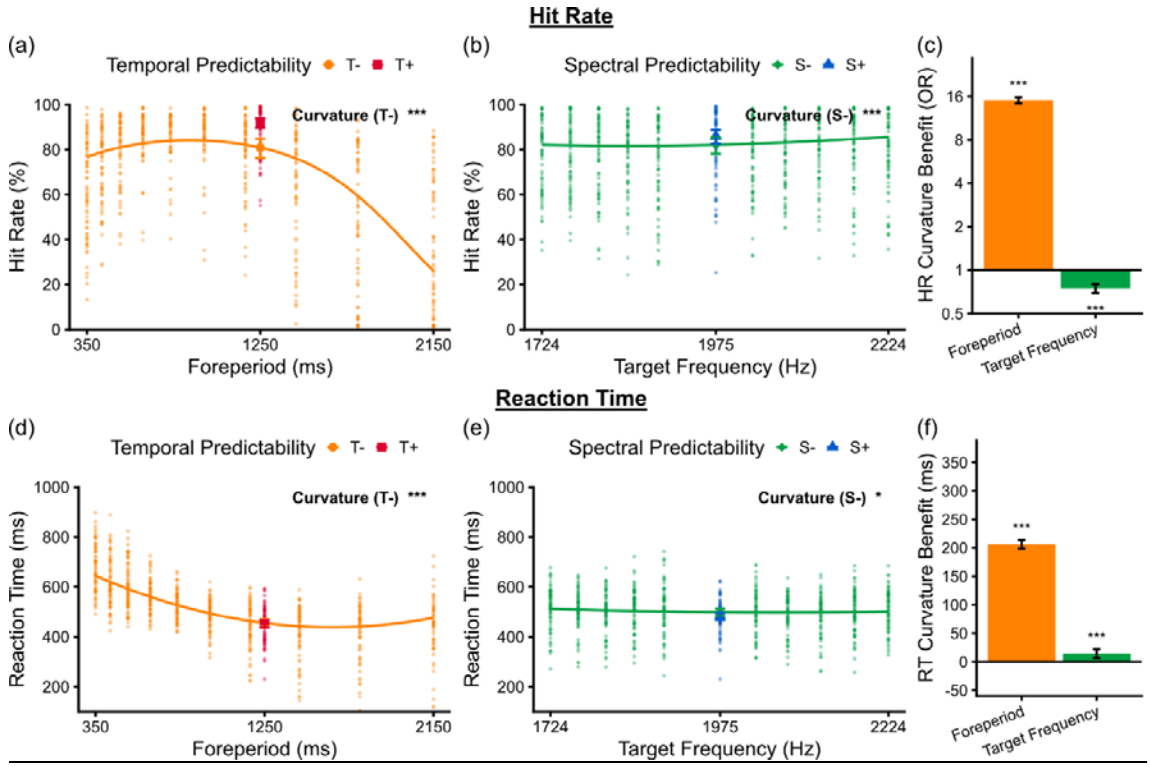
Figure 2: Temporal and spectral predictions affect auditory performance differentially and complementarily. (a) Hit rate (%) as a function of temporal predictability (x-axis) and spectral predictability (colour-coded). Symbols show model-estimated marginal means for each condition; whiskers indicate 95 % confidence intervals. Small dots represent individual participant means. (b) Main effects of temporal (T) and spectral (S) predictability, as well as their interaction (T×S), on hit rate expressed as odds ratios (OR). An odds ratio greater than 1 indicates a higher likelihood of a hit when predictability is high compared to low, whereas an odds ratio smaller than 1 indicates the opposite pattern. (c) Simple effects of temporal and spectral predictability on hit rate (OR), illustrating how the effect of one predictability factor depends on the level of the other. (d-f) Same as panels (a–c), but for false alarm rate. (g-i) Same as panels (a–c), but for sensitivity. (j-l) Same as panels (a-c) but for reaction time.

Figure 3: Distinct encoding strategies for temporal and spectral distributions. (a) Hit rate (%) as a function of foreperiod duration (ms; x-axis) and temporal predictability (colour-coded). The solid line shows model-predicted performance across the variable foreperiod distribution (T-), and dots represent individual participant means. The label “Curvature (T-)” indicates a

non-linear change in performance across the variable foreperiod distribution (T-); asterisks denote the statistical significance of this non-linear (quadratic) effect. Symbols show model-estimated marginal means for each condition at the fixed foreperiod duration, with whiskers indicating 95 % confidence intervals. (b) Same as panel (a), but showing hit rate (%) as a function of target frequency (Hz; x-axis) and spectral predictability (colour-coded). (c) Effect size (OR) of the curvature effect on hit rate in the low-predictability condition. An OR greater than 1 indicates better performance around the centre of the distribution compared to its edges, whereas an OR smaller than 1 indicates worse performance around the centre relative to the edges. (d-f) Same as panels (a-c), but for reaction time.

Supplementary Figure 1: Distinct encoding strategies for temporal and spectral distributions in a pilot experiment with narrower frequency range. (a) Hit rate (%) as a function of foreperiod duration (ms; x-axis) and temporal predictability (colour-coded). The solid line shows model-predicted performance across the variable foreperiod distribution (T-), and dots represent individual participant means. The label "Curvature (T-)" indicates a non-linear change in performance across the variable foreperiod distribution (T-); asterisks denote the statistical significance of this non-linear (quadratic) effect. Symbols show model-estimated marginal means for each condition at the fixed foreperiod duration, with whiskers indicating 95 % confidence intervals. (b) Same as panel (a), but showing hit rate (%) as a function of target frequency (Hz; x-axis) and spectral predictability (colour-coded). (c) Effect size (OR) of the curvature effect on hit rate in the low-predictability condition. An OR greater than 1 indicates better performance around the centre of the distribution compared to its edges, whereas an OR smaller than 1 indicates worse performance around the centre relative to the edges. (d-f) Same as panels (a-c), but for reaction time.

Supplementary material



Supplementary Figure 1 - Distinct encoding strategies for temporal and spectral distributions in a pilot experiment with narrower frequency range. (a) Hit rate (%) as a function of foreperiod duration (ms; x-axis) and temporal predictability (colour-coded). The solid line shows model-predicted performance across the variable foreperiod distribution (T-), and dots represent individual participant means. The label "Curvature (T-)" indicates a non-linear change in performance across the variable foreperiod distribution (T-); asterisks denote the statistical significance of this non-linear (quadratic) effect. Symbols show model-estimated marginal means for each condition at the fixed foreperiod duration, with whiskers indicating 95 % confidence intervals. (b) Same as panel (a), but showing hit rate (%) as a function of target frequency (Hz; x-axis) and spectral predictability (colour-coded). (c) Effect size (OR) of the curvature effect on hit rate in the low-predictability condition. An OR greater than 1 indicates better performance around the centre of the distribution compared to its edges, whereas an OR smaller than 1 indicates worse performance around the centre relative to the edges. (d-f) Same as panels (a-c), but for reaction time.