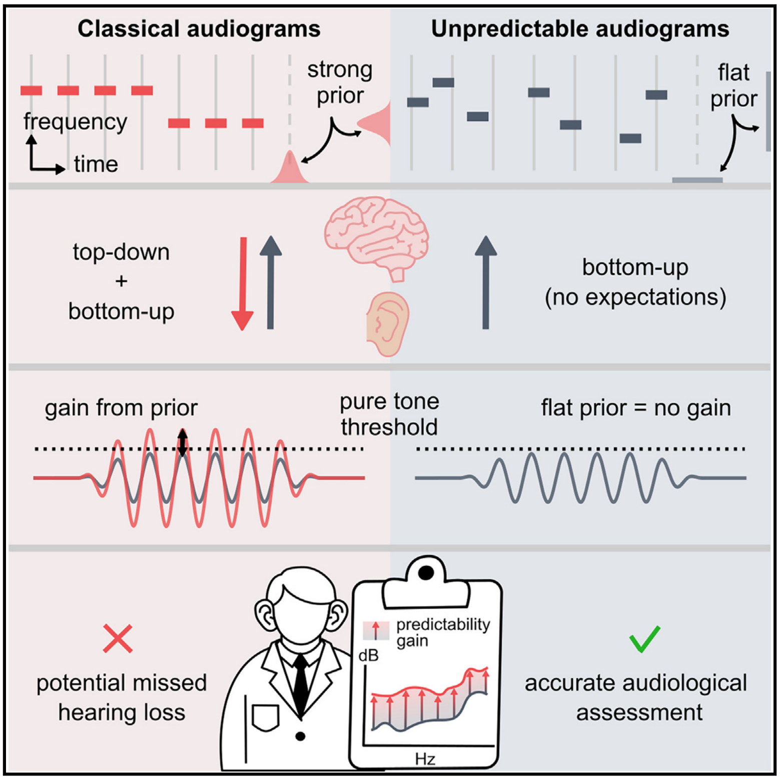
iScience

Article

Predictable sequential structure augments auditory sensitivity at threshold

Graphical abstract



Highlights

- Predictable structure in clinical audiograms augments sensitivity
- Comparing randomized and predictable audiograms showed 6.4 dB increase in thresholds
- Predictive effects are consistent in time and frequency and multiple time scales
- Prediction primarily influences auditory sensitivity rather than decisional bias

Authors

Nadège Marin, Grégory Gérenton, Hadrien Jean, ..., Diane S. Lazard, Keith B. Doelling, Luc H. Arnal

Correspondence

nadege.marin@pasteur.fr (N.M.), keith.doelling@pasteur.fr (K.B.D.)

In brief

Audiology; Sensory neuroscience

Marin et al., 2025, iScience *28*, 112074
March 21, 2025
https://doi.org/10.1016/j.isci.2025.112074

CellPress

iScience

CellPress
OPEN ACCESS

Article

Predictable sequential structure augments auditory sensitivity at threshold

Nadège Marin,[1,*] Grégory Gérenton,[1] Hadrien Jean,[2] Nihaad Paraouty,[2] Nicolas Wallaert,[2] Diane S. Lazard,[1,3]
Keith B. Doelling,[1,4,5,*] and Luc H. Arnal[1,4]
[1]Université Paris Cité, Institut Pasteur, AP-HP, INSERM, CNRS, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, 75012 Paris, France
[2]My Medical Assistant SAS, 51100 Reims, France
[3]Princess Grace Hospital, ENT & Maxillo-facial Surgery Department, Monaco
[4]These authors contributed equally
[5]Lead contact
*Correspondence: nadege.marin@pasteur.fr (N.M.), keith.doelling@pasteur.fr (K.B.D.)
https://doi.org/10.1016/j.isci.2025.112074

SUMMARY

Human hearing is highly sensitive and allows us to detect acoustic events at low levels. However, sensitivity is not only a function of the integrity of cochlear transduction mechanisms but is also constrained by central processes such as attention and expectation. While the effects of distraction and attentional orienting are generally acknowledged, the extent to which probabilistic expectations influence sensitivity at threshold is not clear. Classical audiometric tests, commonly used to assess hearing thresholds, do not distinguish between bottom-up sensitivity and top-down processes. In this study, we aim to decipher the influence of various types of expectations on hearing thresholds and how this information can be used to improve the assessment of hearing sensitivity. Our results raise important questions regarding the conventional assessment of hearing thresholds, both in fundamental research and in audiological clinical assessment.

INTRODUCTION

The sense of hearing allows us to detect and interpret even very weak sounds in our environment. Typically, sensitivity is primarily attributed to the intricate mechanics of the middle and inner ear, and the synaptic transduction of signals to the auditory nerve.[1–3] The efficacy of these mechanics is largely measured using pure-tone detection (PTD) thresholds, assessed via standardized clinical paradigms designed to gauge the softest intensities at which a person can detect pure tones.[4] Traditionally, the audiological field has considered the pure-tone audiogram to primarily reflect bottom-up sensory transduction pathways,[5] where stimulus-driven signals progress from the peripheral sensory organs to higher-order brain regions. However, while these tests emphasize the role of peripheral mechanisms, current audiometry protocols inherently include predictive structures that suggest a more complex interplay of sensory processing mechanisms.

In recent decades, neuroscience has increasingly embraced the idea that sensory processing is more than just a result of bottom-up signal transduction. Instead, perception is now understood as an active process that integrates bottom-up inputs with top-down predictions based on contextual knowledge of prior sensory experiences.[6] These predictions are proposed to be organized in a hierarchical manner across various levels of abstraction, such as semantic, phonemic, or acoustic levels.[7,8] Top-down neural signals – neural processes originating from higher-order brain regions – carry these predictions to shape how sensory input is interpreted by proactively filtering and modulating the gain of expected stimulus inputs.[9] As such, even without conscious awareness, the human brain extracts statistical regularities to exploit predictable structures in its environment.[10] Does the effect of prediction extend to even the most fundamental and low-level auditory function, pure-tone sound detection?

Previous work has shown that hearing sensitivity is contingent on more than the transduction of ascending signals at the cochlear level, but is also affected by the allocation of cognitive constructs such as attentional resources toward upcoming events depending on their behavioral relevance. The probe-signal task,[11] for example, has demonstrated that detection of a tone in noise is enhanced when preceded by a probe tone around the same frequency. A wealth of studies in psychoacoustical experimentation have demonstrated this improvement in performance based on this sort of selective attention *in noise*.[12–14] These studies operate under the assumption that the effects of top-down modulation of signal detection would be the same in quiet at threshold and that the added noise merely externalizes and enhances the neural noise of the system. It remains unclear whether the findings of psychoacoustics in noisy settings apply to signal detection in quiet, or whether expectation only plays a role when separating auditory signal from acoustic noise.

While previous work has studied the role of attention *in quiet*, this has largely been in terms of cross-modal effects, such as auditory processing under visual attention. Rather than manipulating the expected information of attended stimuli, these studies consider attention as a means of removing distraction from one or another sensory domain. Lukas,[15] for example, demonstrated that brainstem responses from auditory nerve to colliculus were filtered by sustained concentrated attention to the visual domain, suggesting top-down control of the entire auditory signal to avoid distraction. Later evidence further suggested that attention has cochlear effects, demonstrated by changes in otoacoustic emissions during attention to one ear vs. the other[16] and by changes in rhythmic modulation of the auditory nerve fiber during attention to the auditory vs. visual domain.[17] These findings provide a mechanism for top-down modulation of broadband cochlear sensitivity via the olivocochlear bundle. Given that a decidedly cortical construct can have direct effect on peripheral function, it is reasonable to explore whether other more *feature-specific* cognitive functions may also influence cochlear processing.

Beyond attention, i.e., the behavioral relevance of the stimulus,[18] auditory perception also relies on the ability to process stimulus statistical properties across multiple dimensions, for instance regarding the content or the timing of a sound. For example, speech contains prosodic cues such as the speeding up or slowing down of syllabic rate which provides further nonverbal information to the listener. As such, predictions about an upcoming sensory event must operate in multiple dimensions. This delineation between content-based (“what”) and time-based (“when”) predictions has been maintained in the field suggesting that the two rely on complementary but distinct top-down mechanisms.[19] In the case of timing, a growing body of evidence supports the coupling of neural oscillations in different timescales to statistical regularities in the sensory input as a mechanism for predictive timing, particularly in the case of auditory stimuli.[20–22] Meanwhile, content predictions are considered as inhibitory top-down signals,[23–25] reducing the processing of predicted signals so that only what is unexpected is given greater neural resources.

These components of prediction, while developed over a large body of research across several decades in the fields of psychology and neuroscience, have not been heavily considered in the audiological domain. PTA thresholds measurements frequently overlook the predictive nature of perception, often presenting the same tone repeatedly and regularly, thereby allowing for the participant to predict when and what a tone is before it arrives. Recent protocols in audiological paradigms have begun to take this into account, instructing clinicians to randomize the timing of tones in pure tone detection tasks to avoid guessing.[26] Still, content predictions (i.e., pitch) have remained ignored: the role such predictions play and how they interact with temporal predictions in threshold evaluation is unknown. Furthermore, most audiograms are conducted manually by a clinician, which limits how unpredictable the resulting sequence can be. Our goal is to assess how much sequential structures contribute to thresholds evaluation under current clinical techniques and to what extent the predictions in either time or content can affect sensory outcomes.

In this study, we leverage recent advances in pure tone detection paradigms[27–29] by employing automated detection threshold techniques using Bayesian machine learning to infer audiograms based on a more flexible paradigm. This approach enables a fully randomized structure in both timing and frequency. We use this advance to compare audiograms within subjects, conducting two experiments in which participants undergo audiometric testing. In the first experiment, we compare directly a gold standard clinical and psychophysical paradigm with the fully randomized paradigm we developed. In a second experiment, we then compare the effect of predictability more directly by contrasting predictability in time and frequency in a highly controlled format at slow (1–3 s) and fast temporal scales (400–800 ms).

In contrast with current clinical expectations, we expect that predictive structure will facilitate sensory detection at threshold, in line with decades of findings in auditory psychophysics using masking in noise. This experiment will provide key information regarding the role of prediction in detection at auditory thresholds in the normal hearing listener and will therefore provide the foundation for future work to show how this relationship may be altered in hearing-impaired populations.

RESULTS

Experiment 1

Predictability augments audiometric sensitivity by 7 dB

First, we compared audiograms from a fully Randomized paradigm and a predictive, three Alternative Forced Choice (3-AFC) paradigm (two-down, one-up adaptive staircase), which represent our most extreme cases of random and predictable respectively, while also being the most standard in terms of previous testing. Details about these paradigms can be found in Figure 1 and the STAR Methods section. Figure 2A shows the average audiogram of all participants for the two paradigms across the tested frequency range from 125 to 8000 Hz. The result shows a significant difference in intensity across frequencies in the two paradigms (two-factor repeated-measures ANOVA, paradigm effect: $F(1, 27) = 271.50$, $p < 0.001$; frequency effect: $F(10, 270) = 11.14$, $p < 0.001$; paradigm × frequency interaction: $F(10, 270) = 2.44$, $p = 0.009$). To estimate the size of this effect, we then extracted, from the Randomized threshold, the discrete threshold values at frequencies that were also tested in the 3AFC and compared the mean values. Figure 2B shows that this comparison yielded a mean improvement of 6.4 dB HL ($t(27) = -13.24$, $p < 0.001$) for the 3AFC paradigm. We repeated this analysis using only the thresholds measured for the frequencies 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz to compute the Pure Tone Average (PTA), a commonly used metric in audiology. For these selected frequencies, the mean improvement in thresholds from Randomized method to the 3-AFC was 7.19 dB HL ($t(27) = -14.41$, $p < 0.001$).

Experiment 2

Predictive structures in time and frequency induce increased sensitivity across timescales

We then tested for what role predictive sequential structure plays in this difference using the fast- and slow-paced paradigms described in the STAR Methods section and Figures 3A and 3B. To control for methodological differences across paradigms,

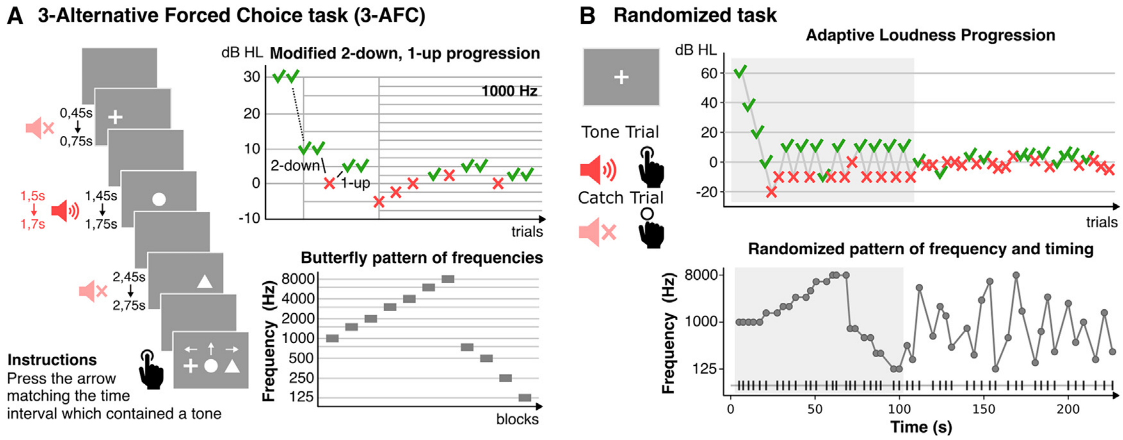


Figure 1. Experiment 1: predictable and unpredictable designs

(A) 3-AFC task. Participants were asked to identify which of 3 possible time intervals contained a tone. A 2-down, 1-up adaptive staircase starting at 30 dB HL determined the level of the presentation of the tones. There is structure in the order of presentation of frequencies, as the next frequency is chosen following a butterfly pattern. The design of this task carries a lot of predictability, both in time and in content.

(B) Randomized task. A fixation cross was displayed on screen while tones chosen by an automated pure-tone audiometry procedure were presented with randomized ISI (1–3 s). Instructions were to press a key as soon as a tone was detected. In this task, the timing and frequency of successive tones is unpredictable.

we calculated a global threshold based on all random conditions, representing an audiogram for unpredictable stimuli. For each condition (FT, F, T, and R), we compared thresholds relative to this global baseline using logistic regression. Figure 3C illustrates this process, showing how pEq values reveal shifts in detection thresholds. For a detailed explanation of the methodology, please refer to the STAR Methods section.

In Figure 4A, we first compared the most random version of each paradigm to assess how changes in protocol alone influence performance. A one-way ANOVA shows a significant difference across paradigms ($F(3, 81) = 191.91$, $p < 0.001$). Pairwise comparisons reveal significant differences between all paradigms, with results summarized in Table 1. The effect of sweeping paradigms compared to the Randomized paradigm reveals that the change in protocols alone - even when cues are reduced as much as possible - provides added information to aid the listener in detection of the tone.

Next, we compared the predictable conditions within the fast- and slow-paced paradigms (Figure 4A, right) to see what role prediction plays in auditory sensitivity outside of the protocol differences analyzed above. To this end, we conducted a separate one-way ANOVA for each paradigm, with predictability (Random (R), Timing (T), Frequency (F), or Frequency and Timing (FT)) as the independent variables. For both paradigms, the ANOVA revealed a significant main effect of predictability (Slow-paced: $F(3, 81) = 6.92$, $p < 0.001$; fast-paced: $F(3, 81) = 14.62$, $p < 0.001$), confirming that even in a tightly controlled paradigm, predictive structure enhances PTD sensitivity. Post-hoc comparisons, summarized in Table 2, further showed that all predictable conditions improved performance compared to the most random condition, with further improvements observed in the fully predictable condition (FT) compared to timing alone (T). In the Slow-paced paradigm, there was a trend toward significance when comparing the fully predictable (FT) condition to the condition with predictable frequency only (F), although this did not reach significance after correcting for multiple comparisons (Table 2). We then compared the gain from having both types of predictability together [FT – R] to the combined gains from each type of predictability independently [(F - R) + (T - R)]. This comparison allowed us to assess whether the combined predictability provides a greater benefit than the sum of its parts, or whether the effects of frequency and time are independent of one another. In the fast-paced condition, a paired t-test revealed that the combined gain from frequency and time independently was significantly greater than the gain from having both jointly predictable, ($t(27) = -2.47$, $p = 0.020$). Conversely, in the slow-paced condition, no significant difference between these gains was found, ($t(27) = 0.32$, $p = 0.749$). These results suggest that in the slow-paced task, the gain from predictability in each dimension is likely independent. However, in the fast-paced task, the gain from the sum of both predictabilities independently is larger than the gain from having both frequency and time jointly predictable, hinting at a different integration mechanism between the two tasks.

Distinct predictive sources across timescales

We next investigated how performance correlated between task pairs. Our aim was to test whether performance was grouped by the category of predictability (frequency vs. timing) or by the paradigmatic setup (slow-vs. fast-paced). The correlation matrix, shown in Figure 4B, reveals that the structure of the task most drives similar performance rather than predictability in a particular dimension like time or frequency. This suggests that the two different time scales lead to different mechanisms through which predictive structure can support performance.

To test this further, we included participant performance into an unsupervised clustering algorithm (4B, right) to assess whether the patterns of participant performance could allow

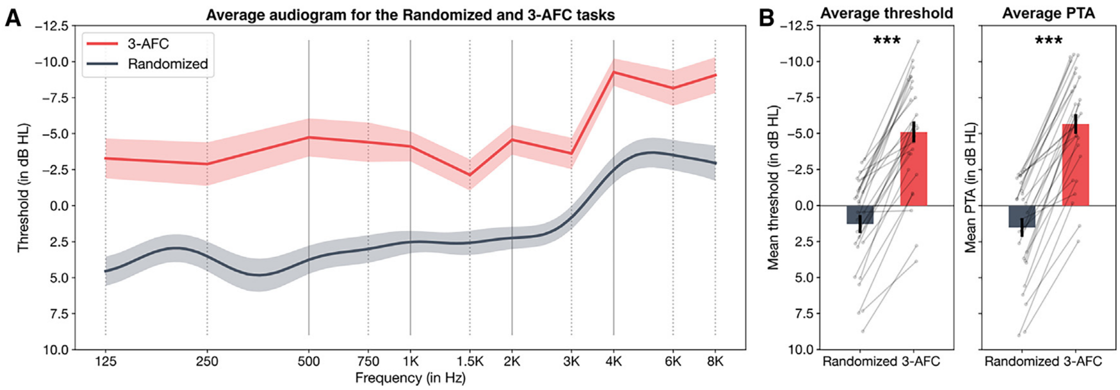


Figure 2. Experiment 1: Thresholds are lower in the highly structured 3-AFC task than in the Randomized task

(A) Average audiograms measured in the Randomized task (in dark gray) and in the 3-AFC task (in red). Shaded areas represent the standard error to the mean (SEM).
(B) Left panel: Average threshold for both tasks, calculated as the mean for the 11 frequencies tested in the 3-AFC paradigm (all vertical lines in panel A). Individual threshold estimates are plotted as dark circles. right panel: Pure Tone Average (PTA) for both tasks, calculated as the mean threshold at 500, 1000, 2000 and 4000 Hz (solid vertical lines in panel A). Individual PTAs are plotted as dark circles. Data are represented as mean ± SEM. Statistical significance was determined using paired t-tests. *** indicates $p < 0.001$.

inferring similar or different mechanisms across paradigms. Interestingly, the clustering algorithm clearly grouped performance by paradigm rather than by the type of predictability in the condition. The slow-paced paradigm grouped with the randomized paradigm, which it most closely resembles in terms of task structure and lack of inherent predictability. In contrast, the fast-paced paradigm clustered with the highly predictive 3-AFC paradigm, as both paradigms involve significant inherent predictability. Specifically, in the fast-paced task, participants focus on the predictability of the last tone within a sequence of five, while in the 3-AFC paradigm, participants focus on the predictability of the tone's occurrence within one of three time intervals.

Predictive gain mostly driven by sensitivity, not decisional bias

Another important question is what role decisional bias may play in this improved performance. As the Experiment 2 paradigms involved a detection task, it could be that participants claimed detection even when they heard nothing. To test for this, we included catch trials in the experiment during which no tone was presented. Figure 4C reveals the results of the catch trial analysis. We conducted a repeated-measures ANOVA with the false alarm rate as the dependent variable, treating each condition as a separate group to allow for comparisons among all conditions. The analysis revealed a significant effect of group ($F(8, 216) = 5.78$, $p < 0.001$). Post-hoc paired t-tests controlling for FDR, summarized in Table 3, indicated that false alarm rates in the Randomized paradigm were significantly lower than in all conditions of Experiment 2. Within the slow-paced paradigm, false alarm rates were higher in both conditions of the task with predictable timing (FT, T) compared to the random condition (R), and they were lower when only the frequency was predictable (F) than when only the time was predictable (T). While temporal predictability in the slow paradigm does appear to be modulated by decisional bias, this effect alone is unlikely to account for the observed improvements in detection thresholds. Predictability in frequency does not appear to affect false alarms (F vs. R) while still enhancing sensitivity. At the same time the fast paradigm shows a nonsignificant reduction in false alarms as predictability increases ($F(3, 69) = 0.668$, $p = 0.575$), excluding decisional bias as a possible explanation of the result. To further test this hypothesis, we correlated false alarm rates of individual participants with their predictive threshold gains (Figure 4D). The slow-paced paradigm shows a significant correlation between false alarm rate and threshold detection when time and frequency are predictable, indicating some contribution of decisional bias to threshold detection (FT: $R^2 = 0.456$, $p < 0.001$). However, the other conditions showed no significant correlations (R: $R^2 = 0.01$, $p = 0.604$; T: $R^2 = 0.109$, $p = 0.086$; F: $R^2 = 0.116$, $p = 0.077$). There is also no such significant correlation in the fast-paced paradigm (R: $R^2 = 0.116$, $p = 0.076$; T: $R^2 = 0.051$, $p = 0.247$; F: $R^2 = 0.026$, $p = 0.408$; FT: $R^2 = 0.052$, $p = 0.244$, Table 4).

Effects of individual differences on experiments 1 and 2

We then sought to explain the variability in participants' ability to extract predictive information in terms of other characteristic features that we had tracked through surveys: musical expertise and age (Figure 5). Neither feature had a particularly strong effect on participant performance. In Experiment 1, we calculated the gain as the average difference in threshold between the two audiological paradigms (Randomized and 3-AFC). Using this measure, we found that neither age ($R^2 = 0.009$, $p = 0.637$) nor musicianship (as indexed by the General Sophistication score of the GMSI, $R^2 = 0.053$, $p = 0.241$) significantly influenced the gain (see Table 4). In experiment 2, in the slow-paced paradigm, Age had no effect on threshold distance (R: $R^2 = 0.001$, $p = 0.878$; T: $R^2 = 0.001$, $p = 0.872$; F: $R^2 = 0.008$, $p = 0.658$; FT: $R^2 = 0.000$, $p = 0.914$). In the fast-paced paradigm, Age had a worsening effect on Random and Time conditions but no effect on Frequency and Both conditions (R: $R^2 = 0.154$, $p = 0.039$; T: $R^2 = 0.226$,

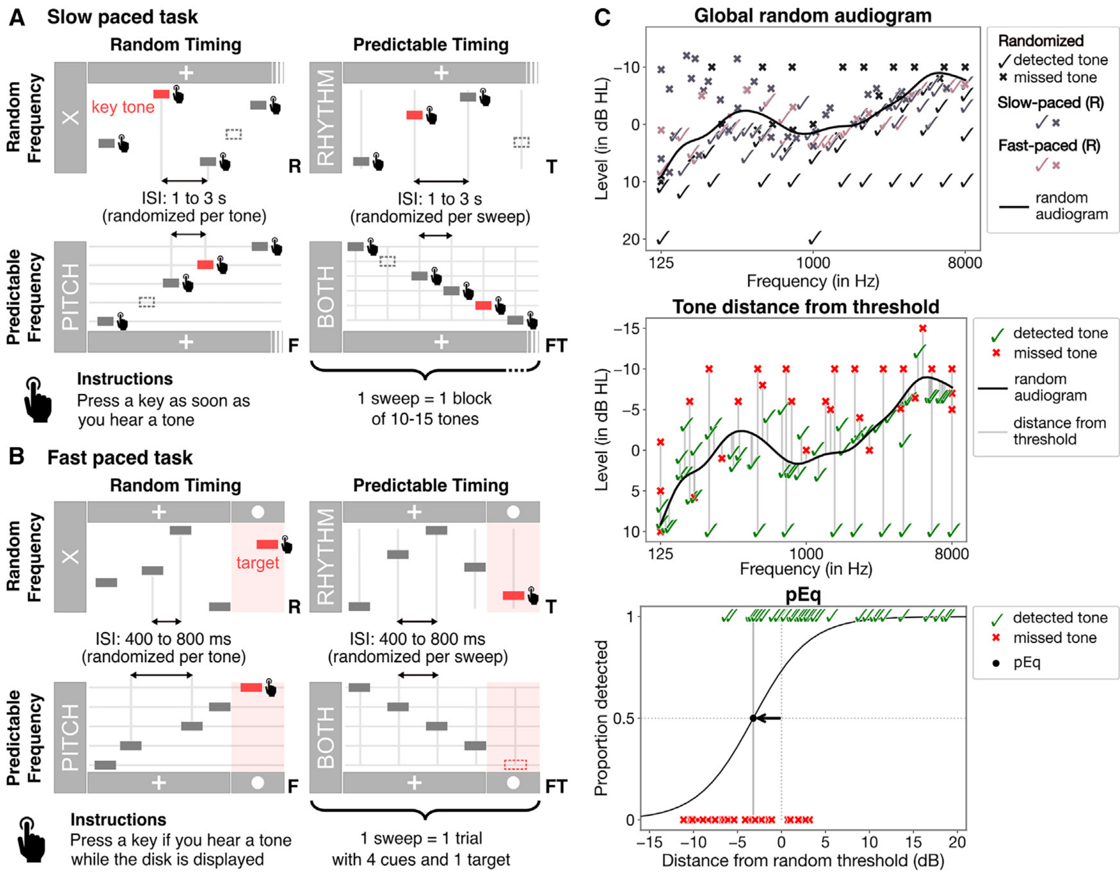


Figure 3. Experiment 2: Slow- and fast-paced designs

Four conditions of predictability were established for both paradigms: predictable timing (T), predictable frequency (F), predictable frequency and timing (FT), or random frequency and timing (R). Predictability in each modality was achieved by designing sweeps with constant intervals, i.e., by regularly spacing tones in time and/or frequency. In conditions of unpredictability, we randomized the intervals (ISIs and/or frequency intervals) for each tone in a sweep.

(A) Slow-paced task. Sweeps consisting of 8–15 tones were presented while a fixation cross was displayed on screen. The instructions were to press a key as soon as a tone was detected. Each sweep contained one catch trial.

(B) Fast-paced task. A cluster of 4 cue tones served to indicate the timing and frequency of a 5th target tone, or not. A fixation cross was displayed on screen until the end of the last cue. A disk then replaced the cross to visually indicate the interval during which the target tone could be presented. Participants were to press a key if they detected a tone while a disk was displayed on screen. 20% of trials contained no target tone, to serve as catch trials.

(C) Method for estimating pEq values (example data from one participant). Top panel: A 'global random' audiogram (solid black line) is computed for each participant from aggregating the answers of the Randomized task and condition R of the Slow- and Fast-paced tasks. Middle panel: for each tested tone in a sweeping condition (R, F, T, or FT; checkmarks: positive detections, crosses: missed detections), the distance from the 'global random' threshold at the tested tone's frequency within each condition is estimated in dB (vertical gray lines). Bottom panel: The tones tested in each condition are ordered by their distance from the 'global' threshold and their proportion of detection is estimated by fitting a sigmoid to the participant's answers. We defined pEq as the distance from the 'global random' threshold corresponding to the 50% mark of detection.

$p = 0.011$; F: $R^2 = 0.007$, $p = 0.68$; FT: $R^2 = 0.011$, $p = 0.601$). Surprisingly, musicianship had no effect on threshold distance in any condition (Fast – R: $R^2 = 0.017$, $p = 0.508$; T: $R^2 = 0.001$, $p = 0.861$; F: $R^2 = 0.061$, $p = 0.204$; FT: $R^2 = 0.008$, $p = 0.653$; Slow – R: $R^2 = 0.026$, $p = 0.416$; T: $R^2 = 0.002$, $p = 0.827$; F: $R^2 = 0.023$, $p = 0.438$; FT: $R^2 = 0.003$, $p = 0.776$).

DISCUSSION

Our findings suggest that the human brain uses predictive information to enhance pure tone auditory sensitivity even at detection threshold, previously thought in clinical domains to be a largely low-level peripheral process. Experiment 1 tested our least and most predictable tasks (the Randomized to the clinical 3AFC paradigm): we showed a 7-dB average difference within the same subjects. Importantly, we found a significant range in differences across subjects (from 2 to 10 dB) suggesting highly variable predictive capacity across participants. This effect is sizable when compared against the 5-dB step size used for clinical measurement.[30] We then used more high-resolution methods to identify how predictive structure in time and frequency contribute to this difference. We showed that predictability in both frequency and time contribute substantially to this predictive effect at two timescales, a slow pace consistent with audiological protocols (1–3 s per tone) and a faster pace consistent with time scales of sequences of acoustic events

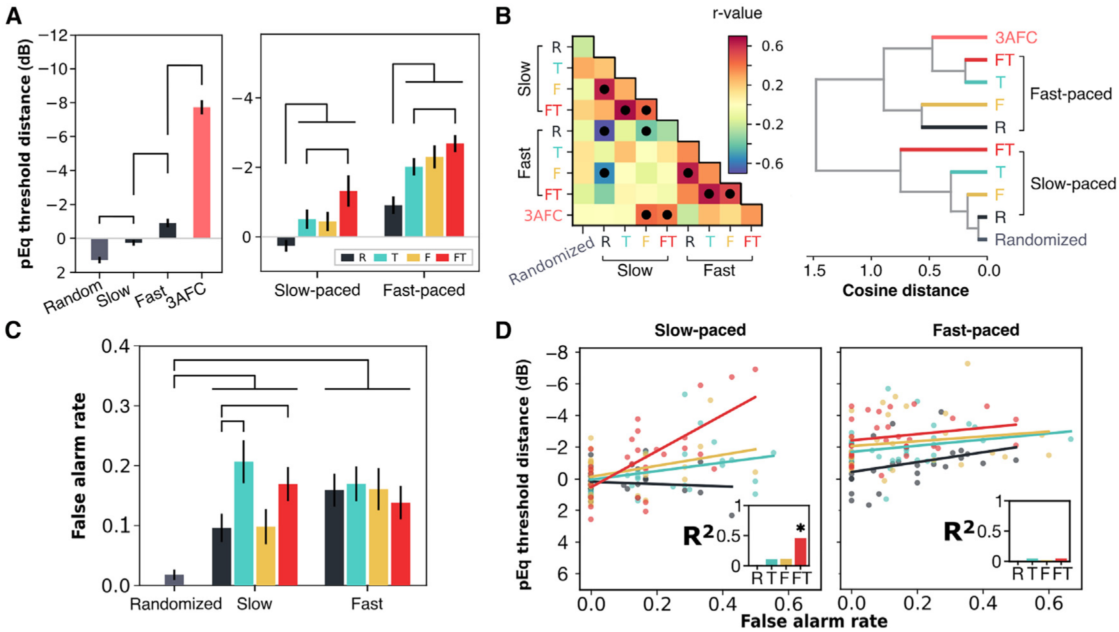


Figure 4. Effect of prediction on detection thresholds

(A) Threshold distance effects due to protocol (left) and to predictability (right). Data are represented as mean ± SEM. Protocol distance is assessed by comparing the random conditions within each protocol. Significant post-hoc comparisons are indicated with black lines. Statistical significance was determined using paired t-tests, with $p < 0.05$. See also Tables 1 and 2.

(B) Protocol clustering based on performance correlations across subjects. Left, Correlation matrix of subject performance by protocol conditions, computed using Pearson's correlation coefficient (r). Filled circles indicate significant correlations ($p < 0.05$). Right, Hierarchical clustering of protocols by cosine distance of subject performance in each cluster.

(C) False alarm rates in each condition. Data are represented as mean ± SEM. Significant post-hoc comparisons (paired t-tests, $p < 0.05$) are indicated with black lines. See also Table 3.

(D) Correlation between threshold distance and false alarm rates for continuous (left) and cluster (right) paradigms. Scatterplots indicate single subject plots, lines indicate linear fit between the two parameters, insets reveal R^2 of the correlation between false alarms and threshold distance. * indicate significance at $p < 0.05$.

like speech and music (400–800 ms). Our findings reveal a contextual influence on peripheral *sensitivity*, likely instantiated via top-down modulation by feedback connections, an effect that remains insufficiently explored and assessed in audiological fields.

Our second experiment investigated the role of prediction under two scenarios: a slow-paced condition in which participants continuously report detection on every tone they hear, spaced out about 1–3 s, and a faster-paced condition, whereby participants first hear a cluster of cue tones (400–800 ms between them) which cue predictively (or not) the timing or frequency (or both) of a target detection tone. Our strongest results occur in the fast-paced case, where tones are detected as part of a sequence occurring at a pace typical of music (~1.2–2.5 Hz). This finding is in keeping with previous work showing that predictions are strong within these sorts of sequences and peak particularly at around this timescale.[20,31] While the slow-paced task also consists of a sequence of tones, the amount of time spaced between tones is quite large making it more likely to be processed as isolated events. Previous work has shown that particularly in non-musicians, sequences are processed differently when item rate is less than 1 Hz.[32]

In our results, we observed that in the slow-paced paradigm, the gain from having both frequency and time predictabilities independently was not significantly different from the gain observed when the two were combined, suggesting that these two forms of prediction linearly combine. This implies that at a slower pace, participants are likely predicting these features independently, perhaps due to the longer intervals allowing for more isolated processing of each predictability cue. Conversely, in the fast-paced paradigm, the gain from the sum of both predictabilities independently was significantly greater than the gain from having both frequency and time jointly predictable. These results suggest that at faster paces, the predictabilities interact in a redundant manner when they are combined, hinting at potentially shared underlying mechanisms. The faster-paced task may, therefore, engage a more complex integrative mechanism, possibly due to the temporal dynamics aligning more closely with naturalistic auditory processing such as in music perception.

Our clustering analysis further supports the conclusion that the two protocols reveal distinct mechanisms of prediction integration. We initially expected that, regardless of pacing, participants would be grouped by the type of prediction (frequency or time). In contrast, we found an opposite effect, where participant performance was best clustered by the timescale of the paradigm. Furthermore, each predictable condition within each protocol clustered together, away from the random condition. These

Table 1. Post-hoc pairwise comparisons between paradigms

Comparison	*t*-value	*p*-value (corr)	Cohen's *d*
3-AFC vs. Randomized	−19.55	<0.001 *	−5.24
3-AFC vs. Fast-paced (R)	−12.79	<0.001 *	−3.72
3-AFC vs. Slow-paced (R)	−17.65	<0.001 *	−4.72
Randomized vs. Fast-paced (R)	6.65	<0.001 *	1.85
Randomized vs. Slow-paced (R)	3.68	0.001 *	1.08
Fast-paced (R) vs. Slow-paced (R)	−3.02	0.006 *	−1.03

Post-hoc paired t-tests comparing conditions of the Randomized, 3-AFC, Fast-paced (random condition, R), and Slow-paced (random condition, R) paradigms. Results include t-values, adjusted *p*-values, and Cohen's d as a measure of effect size. Negative t-values and Cohen's d indicate lower pEq values (i.e., improved sensitivity) in the first condition relative to the second, while positive values indicate higher pEq values (i.e., reduced sensitivity). *p*-values are adjusted using the Benjamini-Hochberg procedure to control the false discovery rate (FDR). Significant comparisons ($p < 0.05$, adjusted) are marked with an asterisk. See also Figure 4A, left panel.

findings suggest that similarity across participants is driven more so by prediction ability at distinct timescales rather than over distinct features. This clustering is not due to the overall mean shift in performance across the protocols. Our clustering method operates over the cosine similarity across participants, which ignores this overall shift. Instead, the result must be due to the overall pattern of individual differences across participants.

Our results exclude the possibility of two alternative hypotheses to explain the improvement in thresholds: (1) a change in decisional biases and (2) a local change in cochlear state. A change in decisional bias would plausibly lead to improved thresholds if, for example, participants indiscriminately indicated detecting the sound regardless of the stimulus. We address this concern by including catch trials in which no sound was presented. By analyzing these trials, we find that decisional bias cannot explain our results. While in the slower paced experiment, false alarms increased with temporal prediction, predictability in frequency shows no increase in false alarms. In the faster-paced experiment, false alarms go down with predictability (though not significantly). From this we can negate the first alternative hypothesis: a significant portion of our results must be due to an increase in *sensitivity* not to a change in decisional bias.

Another hypothesis suggests that the increase in sensitivity is due not to central processing but instead to a local change in the cochlear processor. A myriad of studies have shown that local peripheral responses can change as a result of preceding and concurrent stimuli including masking effects,[33] and noise protection through the modulatory effect of the medial olivocochlear (MOC) reflex.[34] Our experimental design also negates this possibility as the gain in performance is demonstrated in comparison against a random control which maintains the stimulus context in every way except predictability. For example, comparing the effect of predictability in timing vs. random, in this case, the frequency distributions of tones, the temporal intervals between tones, and the visual cue for the target window are exactly the same. The only change between these two conditions is that the temporal intervals are held constant per trial in the predictable case and shuffled for the random case. To take advantage of this, the cochlea would need to execute high-level computations that are typically expected in cerebellum, basal ganglia and motor cortex for a recent review, see.[35] From our perspective, a temporal prediction redundancy in the cochlea is a far less plausible hypothesis.

Instead, we propose that our findings result from an interaction of top-down predictive information with cochlear responses, enhancing sensitivity for *specific frequencies* and time points based on statistical expectation. This interaction may occur at either peripheral or central levels of the auditory system. At the peripheral level, descending auditory pathways —particularly the MOC system— are known to modulate cochlear responses.[36] These efferent fibers can alter cochlear micromechanics, facilitating sensory gain control and selective attention, effects that can be measured by otoacoustic emissions.[37,38] Additionally, the cortico-olivocochlear pathway has been implicated in top-down control mechanisms that prioritize relevant stimuli and suppress irrelevant noise, providing a pathway for cognitive processes to influence early sensory responses.[36,39]

At the central level, predictive coding frameworks propose that hierarchical feedback signals from higher cortical areas, such as the frontal cortex, regulate sensory processing in subcortical and primary sensory regions. Computational[40] and

Table 2. Post-hoc pairwise comparisons within both sweeping paradigms of experiment 2

	Slow-paced			Fast-paced		
Comparison	*t*-value	*p*-value (corr)	Cohen's *d*	*t*-value	*p*-value (corr)	Cohen's *d*
FT vs. R	−3.60	0.008 *	−0.87	−6.29	<0.001 *	−1.35
F vs. R	−3.22	0.010 *	−0.58	−5.28	<0.001 *	−0.89
T vs. R	−2.63	0.028 *	−0.63	−3.85	0.001 *	−0.83
F vs. T	0.20	0.841	0.05	−0.83	0.416	−0.18
FT vs. T	−2.39	0.036 *	−0.41	−3.16	0.006 *	−0.51
FT vs. F	−2.12	0.052	0.44	−1.33	0.234	0.25

Post-hoc paired t-tests comparing conditions within the Slow-paced and Fast-paced paradigms. Results include t-values, adjusted *p*-values, and Cohen's d as a measure of effect size. *p*-values are adjusted using the Benjamini-Hochberg procedure to control the FDR. Significant comparisons ($p < 0.05$, adjusted) are marked with an asterisk. See also Figure 4A, right panel.

Table 3. Post-hoc comparisons of false alarm rates

Comparison	t-value	*p*-value (corr)	Cohen's d
Randomized vs. Fast F	−3.84	0.004 *	−1.06
Randomized vs. Fast FT	−4.21	0.002 *	−1.1
Randomized vs. Fast R	−5.0	<0.001 *	−1.31
Randomized vs. Fast T	−5.02	<0.001 *	−1.33
Randomized vs. Slow F	−2.58	0.047 *	−0.7
Randomized vs. Slow FT	−5.13	<0.001 *	−1.38
Randomized vs. Slow R	−3.15	0.018 *	−0.82
Randomized vs. Slow T	−5.56	<0.001 *	−1.47
Fast F vs. Fast FT	0.56	0.717	0.14
Fast F vs. Fast R	0.05	0.967	0.01
Fast F vs. Fast T	−0.24	0.941	−0.05
Fast F vs. Slow F	1.64	0.214	0.42
Fast F vs. Slow FT	−0.04	0.967	−0.01
Fast F vs. Slow R	1.76	0.192	0.42
Fast F vs. Slow T	−1.02	0.445	−0.26
Fast FT vs. Fast R	−0.71	0.622	−0.14
Fast FT vs. Fast T	−1.1	0.424	−0.21
Fast FT vs. Slow F	1.4	0.296	0.32
Fast FT vs. Slow FT	−0.84	0.542	−0.17
Fast FT vs. Slow R	1.36	0.302	0.32
Fast FT vs. Slow T	−2.3	0.070	−0.43
Fast R vs. Fast T	−0.32	0.904	−0.07
Fast R vs. Slow F	1.68	0.207	0.47
Fast R vs. Slow FT	−0.1	0.967	−0.02
Fast R vs. Slow R	2.0	0.125	0.48
Fast R vs. Slow T	−1.28	0.331	−0.31
Fast T vs. Slow F	2.71	0.038 *	0.52
Fast T vs. Slow FT	0.21	0.941	0.05
Fast T vs. Slow R	2.42	0.058	0.54
Fast T vs. Slow T	−1.01	0.445	−0.24
Slow F vs. Slow FT	−2.49	0.054	−0.5
Slow F vs. Slow R	−0.15	0.965	−0.02
Slow F vs. Slow T	−3.07	0.019 *	−0.73
Slow FT vs. Slow R	2.97	0.022 *	0.51
Slow FT vs. Slow T	−1.54	0.242	−0.29
Slow R vs. Slow T	−3.22	0.017 *	−0.75

Post-hoc paired t-tests comparing false alarm rates in all conditions of experiments 1 and 2. Results include t-values, adjusted *p*-values and Cohen's d as a measure of effect size. *p*-values are adjusted using the Benjamini-Hochberg procedure to control the FDR. Significant comparisons ($p < 0.05$, adjusted) are marked with an asterisk. See also Figure 4C.

neurophysiological studies[39,41] suggest that inhibitory feedback projections from frontal areas shape sensory responses based on expectations formed by prior knowledge. This aligns with predictive coding principles, where unexpected sensory inputs generate prediction error signals that propagate up the cortical hierarchy to update internal models.[19,23,25] Descending projections from the auditory cortex to subcortical structures, such as the inferior colliculus, further highlight the central-peripheral interaction. For instance, optogenetic studies in mice have shown that cortico-collicular feedback regulates predictive coding metrics, such as prediction error, at the level of the inferior colliculus.[42] These findings suggest that central predictive mechanisms can fine-tune subcortical auditory processing to optimize responses based on contextual expectations.

Taken together, our findings support the notion that top-down predictive processes augment auditory sensitivity. Such an effect may occur through direct modulation of central sensory responses or indirect modulation via descending pathways, such as the MOC system. Both mechanisms may involve dynamic adjustments in neural excitability and gain control along the auditory pathway, enabling the integration of frequency and temporal predictions across varying timescales. Whether this interaction involves a change in cochlear processing directly or instead of the central interpretation of cochlear responses will require future work.

Either hypothesis has major implications for the fields of audiology and neuroscience. For example, if the hypothesis of altering cochlear responses via central input is correct, it would provide a mechanism for phase locked loops whereby the central perceptual system can endogenously and voluntarily alter its bottom-up input, similar to eye movements or whisking but in the spectral domain.[43–45] Alternatively, if the results are due to a change in central interpretation, the findings upend the assumption in audiology that pure tone audiogram measure purely cochlear health.[5]

While audiologists are aware that predictability can alter their results, it is standard practice in several countries to present tones with random timing. For example, this effect is thought to be due to our first alternative hypothesis: decisional bias, patients may report detecting a sound regardless of what they have heard purely because they know where the tone should have been. Our results confirm this concern showing in the slow-paced paradigm that temporal predictability can result in decisional biases indicating detected tones regardless of perception. However, we additionally show for the first time that prediction can not only alter decisional bias but also enhance sensitivity and to a significant degree. Therefore, depending on the chosen clinical setup of the paradigm, the pure tone audiogram may be diagnosing more fully the auditory neural apparatus, mixing effects of both peripheral cochlear health and central, predictive ability. From the perspective of standard practice, this finding may not represent a significant issue for clinicians: the typical sounds that patients experience in their lives have some degree of predictability in them and as such current clinical setups match patients' daily experience. Still, our findings reveal two separable components present in sensory detection which may have separate trajectories of development and degradation over the course of a lifespan. Future work will investigate these trajectories more fully.

To that end, although the current task was not designed to study individual differences in perception, we took advantage of the natural variance in our participant cohort to explore how age (<45 years old) might influence predictive processing. We only observed a significant effect of age in the random and timing predictions of the fast-paced task. Interestingly, this effect disappeared under conditions with frequency predictability

Table 4. Effects of age and musicianship on sensitivity

Paradigm	Condition	Age (R^2, *p*-value)	Musicianship (R^2, *p*-value)
Experiment 1	Gain	0.009, 0.637	0.053, 0.241
Experiment 2 - Slow	R	0.001, 0.878	0.026, 0.416
	T	0.001, 0.872	0.002, 0.827
	F	0.008, 0.658	0.023, 0.438
	FT	0.000, 0.914	0.003, 0.776
Experiment 2 - Fast	R	0.154, 0.039 *	0.017, 0.508
	T	0.226, 0.011 *	0.001, 0.861
	F	0.007, 0.680	0.061, 0.204
	FT	0.011, 0.601	0.008, 0.653

Statistical results from regression analyses examining the influence of age and musicianship (General Musical Sophistication Index, GMSI) on gain (Experiment 1) and threshold distance (Experiment 2) across conditions. Experiment 1 reports the average gain between Randomized and 3-AFC paradigms. Experiment 2 evaluates the effects separately for the slow-paced and fast-paced paradigms. R^2 values represent the proportion of variance explained, and *p*-values assess significance. Significant results ($p < 0.05$) are marked with an asterisk. See also Figure 5.

(F and FT). This may suggest that improved predictable structure in stimulus content can mask even early age-related reductions in sensitivity. Future work will further investigate these effects, particularly in older populations and individuals with clinical sensory hearing loss, to assess how prediction can be used either as a coping mechanism or an indicator of future decline.

Together, our findings reveal that pure tone audiometric measurements conflate the influence of both bottom-up sensory responsivity and top-down predictive modulation. We show that prediction in both frequency and time and at multiple timescales can influence the sensitivity of the auditory system. Signal detection is most improved in the context of sequences with timing similar to those of typical auditory stimuli like speech and music. We also show key differences in behavior between these timescales including that temporal predictions at slower timescales may influence more decisional bias whereas frequency predictions improve sensitivity. Lastly, we look for individual differences to understand the variability in predictive performance. Even in our younger population, we find an effect of age on sensory detection in fast paced sequences which can be recovered by frequency predictability. We expect our findings to inspire audiologists and auditory neuroscientists to track raw predictive ability more carefully across age and lifespan as a potential key factor in understanding hearing abilities.

Limitations of the study

The tones in this experiment were presented binaurally, which limits its immediate applicability to clinical audiometry settings that often use monaural stimuli for more precise ear-specific measurements. Future research could adapt this paradigm to monaural presentations to better align with standard audiometric practices.

Additionally, our participant cohort was narrow in its scope as it consisted exclusively of young, healthy individuals with normal hearing (<20 dB HL hearing loss), leaving open the question of how these findings generalize to elderly populations with and without hearing loss. While we show that audiograms bias sensitivity in the predictability of their stimulus presentation, this bias could be significantly altered with hearing loss and age, either increasing bias – using predictive mechanisms to compensate for degraded auditory input – or decreasing bias as degraded sounds engage the patient less. This raises the possibility that predictability could be used to categorize patients based on their reliance on temporal or frequency cues, potentially serving as a diagnostic tool. Future studies should explore whether leveraging predictability in audiometric assessments could uncover meaningful distinctions in auditory processing among

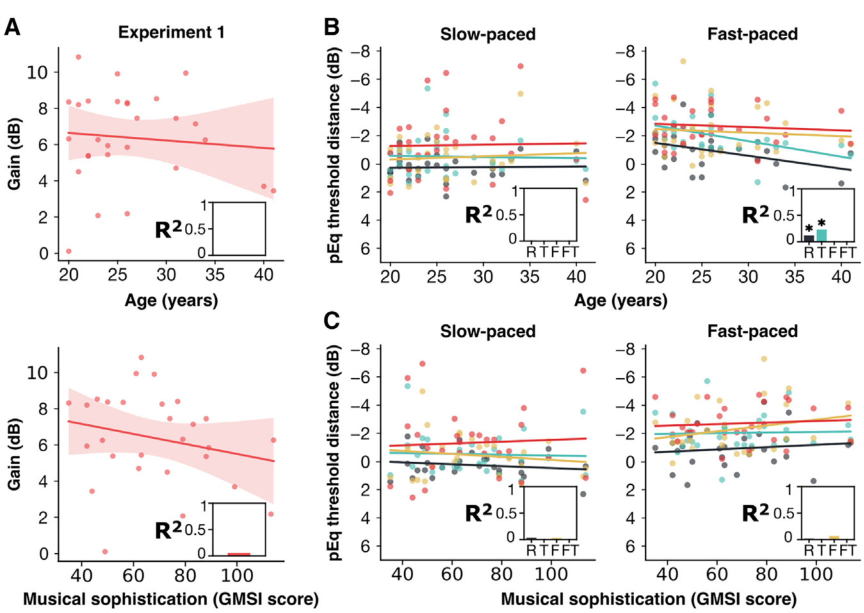


Figure 5. Predictive effects variability largely unexplained by age and musicianship

(A) Correlation between gain and participant Age (top) and gain and GMSI score (musicianship; bottom) in Experiment 1. Scatterplots indicate individual participant values, lines indicate linear regression, and shaded areas represent the 95% confidence interval. R^2 values indicating variance explained are reported in the lower right insets, and asterisks denote significance.

(B) Correlation between pEq threshold distance and participant Age, in the slow-paced (right) and fast-paced (left) paradigms of experiment 2. Scatterplots indicate individual participants, lines indicate linear regression. R^2 of variance explained is reported in lower right insets. Asterisks indicate significance. Confidence intervals were omitted to reduce visual clutter.

(C) Correlation between pEq threshold distance and participant musicianship in experiment 2. Plot organization is the same as in B. See also Table 4.

* indicates p < 0.05

individuals with hearing impairments, particularly those with varying degrees of hearing loss or reliance on predictive strategies.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Keith B. Doelling (keith.doelling@pasteur.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data have been deposited on Gitlab at https://gitlab.pasteur.fr/ida-public/PredAudio and are publicly available as of the date of publication.
- All original code has also been deposited on Gitlab at https://gitlab.pasteur.fr/ida-public/PredAudio and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

The authors would like to thank Benjamin Morillon, Céline Quinsac, Paul Avan, and Lavinia Slabu for their insights and discussions on this project.

The authors also thank the Pole de Coordination de la Recherche Clinique at Institut Pasteur for their efforts in ensuring our protocols meet ethical standards.

This work was supported by the Fondation Pour l'Audition (RD-2020-10, L.H.A.), a French government grant managed by the Agence Nationale de la Recherche under the France 2030 program (ANR-23-IAHU-0003, L.H.A.), a Fondation Fyssen Postdoctoral Fellowship (K.B.D.), BPIFrance i-Nov grant DOS0127610/00 (N.W.), and Fondation d'entreprise Optic 2000 - Lissac - Audio 2000 (Donation to D.S.L. and L.H.A.).

AUTHOR CONTRIBUTIONS

N.M., D.S.L., L.H.A., and K.B.D. designed research. G.G., N.P., N.W., and H.J. provided materials for experimental setup. N.M. and K.B.D. performed research and analyzed data. N.M., L.H.A., and K.B.D. wrote the first draft of the paper. N.M., G.G., H.J., N.P., N.W., D.S.L., L.H.A., and K.B.D. edited the paper.

DECLARATION OF INTERESTS

The authors declare the following competing interests: NW has a patent pending on technology described in the manuscript. NW has equity ownership in My Medical Assistant SAS. HJ, and NP receive salary from My Medical Assistant SAS.

NM, GG, DSL, LHA and KBD declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Stimuli and material
  - Experimental setup
  - Experiment 1: Comparison of audiometric paradigms
  - Experiment 2: Time vs. frequency predictions
  - Common elements across paradigms
  - Further details on slow-paced paradigm
  - Further details on fast-paced paradigm
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis for experiment 1
  - Analysis for experiment 2
  - Statistics for experiments 1 and 2

Received: September 13, 2024
Revised: January 15, 2025
Accepted: February 17, 2025
Published: February 20, 2025

REFERENCES

1. Hudspeth, A.J. (2014). Integrating the active process of hair cells with cochlear function. Nat. Rev. Neurosci. *15*, 600–614. https://doi.org/10.1038/nrn3786.
2. Marin, N., Lobo Cerna, F., and Barral, J. (2022). Signatures of cochlear processing in neuronal coding of auditory information. Mol. Cell. Neurosci. *120*, 103732. https://doi.org/10.1016/j.mcn.2022.103732.
3. Pickles, J. (1998). An Introduction to the Physiology of Hearing. In An Introduction to the Physiology of Hearing, Fourth Edition (Brill).
4. Brant, L.J., and Fozard, J.L. (1990). Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging. J. Acoust. Soc. Am. *88*, 813–820. https://doi.org/10.1121/1.399731.
5. Musiek, F.E., Shinn, J., Chermak, G.D., and Bamiou, D.-E. (2017). Perspectives on the Pure-Tone Audiogram. J. Am. Acad. Audiol. *28*, 655–671. https://doi.org/10.3766/jaaa.16061.
6. Friston, K. (2005). A theory of cortical responses. Phil. Trans. R. Soc. B *360*, 815–836. https://doi.org/10.1098/rstb.2005.1622.
7. Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., and de Lange, F.P. (2022). A hierarchy of linguistic predictions during natural language comprehension. Proc. Natl. Acad. Sci. USA *119*, e2201968119. https://doi.org/10.1073/pnas.2201968119.
8. Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., and Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc. Natl. Acad. Sci. USA *108*, 20754–20759. https://doi.org/10.1073/pnas.1117807108.
9. Hesselmann, G., Sadaghiani, S., Friston, K.J., and Kleinschmidt, A. (2010). Predictive Coding or Evidence Accumulation? False Inference and Neuronal Fluctuations. PLoS One *5*, e9926. https://doi.org/10.1371/journal.pone.0009926.
10. Saffran, J.R., Johnson, E.K., Aslin, R.N., and Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. Cognition *70*, 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4.
11. Greenberg, G.Z., and Larkin, W.D. (1968). Frequency-Response Characteristic of Auditory Observers Detecting Signals of a Single Frequency in Noise: The Probe-Signal Method. J. Acoust. Soc. Am. *44*, 1513–1523. https://doi.org/10.1121/1.1911290.
12. Borra, T., Versnel, H., Kemner, C., Van Opstal, A.J., and Van Ee, R. (2013). Octave effect in auditory attention. Proc. Natl. Acad. Sci. USA *110*, 15225–15230. https://doi.org/10.1073/pnas.1213756110.
13. Scharf, B., Quigley, S., Aoki, C., Peachey, N., and Reeves, A. (1987). Focused auditory attention and frequency selectivity. Percept. Psychophys. *42*, 215–223. https://doi.org/10.3758/BF03203073.
14. Tanner, W., and Norman, R. (1954). The human use of information–II: Signal detection for the case of an unknown signal parameter. Trans. IRE Prof. Group Inf. Theory *4*, 222–227. https://doi.org/10.1109/TIT.1954.1057462.
15. Lukas, J.H. (1980). Human Auditory Attention: The Olivocochlear Bundle May Function as a Peripheral Filter. Psychophysiology *17*, 444–452. https://doi.org/10.1111/j.1469-8986.1980.tb00181.x.
16. Giard, M.-H., Collet, L., Bouchet, P., and Pernier, J. (1994). Auditory selective attention in the human cochlea. Brain Res. *633*, 353–356. https://doi.org/10.1016/0006-8993(94)91561-X.

17. Gehmacher, Q., Reisinger, P., Hartmann, T., Keintzel, T., Rösch, S., Schwarz, K., and Weisz, N. (2022). Direct Cochlear Recordings in Humans Show a Theta Rhythmic Modulation of Auditory Nerve Activity by Selective Attention. J. Neurosci. *42*, 1343–1351. https://doi.org/10.1523/JNEUROSCI.0665-21.2021.
18. Summerfield, C., and Egner, T. (2009). Expectation (and attention) in visual cognition. Trends Cognit. Sci. *13*, 403–409. https://doi.org/10.1016/j.tics.2009.06.003.
19. Arnal, L.H., and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. Trends Cognit. Sci. *16*, 390–398. https://doi.org/10.1016/j.tics.2012.05.003.
20. Arnal, L.H., Doelling, K.B., and Poeppel, D. (2015). Delta-Beta Coupled Oscillations Underlie Temporal Prediction Accuracy. Cerebr. Cortex *25*, 3077–3085. https://doi.org/10.1093/cercor/bhu103.
21. Doelling, K.B., Assaneo, M.F., Bevilacqua, D., Pesaran, B., and Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. Proc. Natl. Acad. Sci. USA *116*, 10113–10121. https://doi.org/10.1073/pnas.1816414116.
22. Nobre, A.C., Correa, A., and Coull, J. (2007). The hazards of time. Curr. Opin. Neurobiol. *17*, 465–470. https://doi.org/10.1016/j.conb.2007.07.006.
23. Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical microcircuits for predictive coding. Neuron *76*, 695–711. https://doi.org/10.1016/j.neuron.2012.10.038.
24. Friston, K. (2010). The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. *11*, 127–138. https://doi.org/10.1038/nrn2787.
25. Heilbron, M., and Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? Neuroscience *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061.
26. British Society of Audiology (2018). Pure tone air and bone conduction threshold audiometry with and without masking (UK: British Society of Audiology).
27. Schlittenlacher, J., Turner, R.E., and Moore, B.C.J. (2018). Audiogram estimation using Bayesian active learning. J. Acoust. Soc. Am. *144*, 421–430. https://doi.org/10.1121/1.5047436.
28. Wallaert, N., Perry, A., Jean, H., Creff, G., Godey, B., and Paraouty, N. (2024). Performance and Reliability Evaluation of an Automated Bone-Conduction Audiometry Using Machine Learning. Trends in Hearing *28*, 23312165241286456. https://doi.org/10.1177/23312165241286456.
29. Wallaert, N., Perry, A., Quarino, S., Jean, H., Creff, G., Godey, B., and Paraouty, N. (2024). Performance and reliability evaluation of an improved machine learning-based pure-tone audiometry with automated masking. World Journal of Otorhinolaryngology - Head and Neck Surgery *n/a*. https://doi.org/10.1002/wjo2.208.
30. Favier, V., Vincent, C., Bizaguet, É., Bouccara, D., Dauman, R., Frachet, B., Le Her, F., Meyer-Bisch, C., Tronche, S., Sterkers-Artières, F., and Venail, F. (2018). French Society of ENT (SFORL) guidelines (short version): Audiometry in adults and children. Eur. Ann. Otorhinolaryngol. Head Neck Dis. *135*, 341–347. https://doi.org/10.1016/j.anorl.2018.05.009.
31. Morillon, B., Arnal, L.H., Schroeder, C.E., and Keitel, A. (2019). Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. Neurosci. Biobehav. Rev. *107*, 136–142. https://doi.org/10.1016/j.neubiorev.2019.09.012.
32. Doelling, K.B., and Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. Proc. Natl. Acad. Sci. USA *112*, E6233–E6242. https://doi.org/10.1073/pnas.1508431112.
33. Harris, D.M., and Dallos, P. (1979). Forward masking of auditory nerve fiber responses. J. Neurophysiol. *42*, 1083–1107.
34. Morand-Villeneuve, N., Garnier, S., Grimault, N., Veuillet, E., Collet, L., and Micheyl, C. (2002). Medial olivocochlear bundle activation and perceived auditory intensity in humans. Physiol. Behav. *77*, 311–320.
35. Cannon, J.J. (2021). How Beat Perception Co-opts Motor Neurophysiology. Trends Cognit. Sci. *14*.
36. Elgueda, D., and Delano, P.H. (2020). Corticofugal modulation of audition. Current Opin. Physiol. *18*, 73–78. https://doi.org/10.1016/j.cophys.2020.08.016.
37. Perrot, X., Ryvlin, P., Isnard, J., Guénot, M., Catenoix, H., Fischer, C., Mauguière, F., and Collet, L. (2006). Evidence for Corticofugal Modulation of Peripheral Auditory Activity in Humans. Cerebr. Cortex *16*, 941–948. https://doi.org/10.1093/cercor/bhj035.
38. Srinivasan, S., Keil, A., Stratis, K., Woodruff Carr, K.L., and Smith, D.W. (2012). Effects of cross-modal selective attention on the sensory periphery: Cochlear sensitivity is altered by selective attention. Neuroscience *223*, 325–332. https://doi.org/10.1016/j.neuroscience.2012.07.062.
39. Wittekindt, A., Kaiser, J., and Abel, C. (2014). Attentional Modulation of the Inner Ear: A Combined Otoacoustic Emission and EEG Study. J. Neurosci. *34*, 9995–10002. https://doi.org/10.1523/JNEUROSCI.4861-13.2014.
40. Spratling, M.W. (2010). Predictive Coding as a Model of Response Properties in Cortical Area V1. J. Neurosci. *30*, 3531–3543. https://doi.org/10.1523/JNEUROSCI.4911-09.2010.
41. Sohoglu, E., Peelle, J.E., Carlyon, R.P., and Davis, M.H. (2012). Predictive Top-Down Integration of Prior Knowledge during Speech Perception. J. Neurosci. *32*, 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012.
42. Lesicko, A.M.H., Angeloni, C.F., Blackwell, J.M., De Biasi, M., and Geffen, M.N. (2022). Corticofugal regulation of predictive coding. Elife *11*, e73289. https://doi.org/10.7554/eLife.73289.
43. Ahissar, E., and Kleinfeld, D. (2003). Closed-loop Neuronal Computations: Focus on Vibrissa Somatosensation in Rat. Cerebr. Cortex *13*, 53–62. https://doi.org/10.1093/cercor/13.1.53.
44. Ahissar, E., and Arieli, A. (2012). Seeing via Miniature Eye Movements: A Dynamic Hypothesis for Vision. Front. Comput. Neurosci. *6*, 89. https://doi.org/10.3389/fncom.2012.00089.
45. Ahissar, E., Nelinger, G., Assa, E., Karp, O., and Saraf-Sinik, I. (2023). Thalamocortical loops as temporal demodulators across senses. Commun. Biol. *6*, 562–611. https://doi.org/10.1038/s42003-023-04881-4.
46. Wallaert, N., Perry, A., Quarino, S., Jean, H., Creff, G., Godey, B., and Paraouty, N. (2024). Performance and reliability evaluation of an improved machine learning-based pure-tone audiometry with automated masking (World Journal of Otorhinolaryngology-Head and Neck Surgery).
47. Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J.K. (2019). PsychoPy2: Experiments in behavior made easy. Behav. Res. Methods *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y.
48. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.
49. Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. SciPy *7*, 92–96. https://doi.org/10.25080/Majora-92bf1922-011.
50. Vallat, R. (2018). Pingouin: statistics in Python. J. Open Source Softw. *3*, 1026.
51. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
52. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B *57*, 289–300.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and analyzed data	This paper	gitlab.pasteur.fr/ida-public/PredAudio
Software and algorithms		
iAudiogram algorithm	Wallaert et al.[46]	My Medical Assistant SAS (Reims, France)
Original code used in this project	This paper	gitlab.pasteur.fr/ida-public/PredAudio
PsychoPy 2022.1.1	Peirce et al.[47]	github.com/psychopy/psychopy

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

31 participants aged 18 to 45 with self-reported normal hearing were recruited using the RISC (Relai d'Information sur les Sciences de la Cognition) volunteers' database (Risc., https://www.risc.cnrs.fr/.). Normal hearing was confirmed (Pure Tone Average <20 dB HL) in the first audiometry test (described below) and no participants were removed on this basis. Three participants were excluded from all the analyses, for either not following the instructions, or due to a data collection error. We present the results from the remaining 28 participants (12 men, 16 women) with an average age of 26.25 years-old (sd: 5.73). All participants are involved in all experiments reported in this paper. Informed consent was obtained prior to testing and participation in this study was compensated. One session lasted about 1.5h in total. We assessed self-reported musical ability by administering the Goldsmiths Musical Sophistication Index (Gold-MSI). This study was approved by the Comité de Protection des Personnes Tours OUEST 1 on 10/09/2020 (project identification number 2020T317 RIPH3 HPS).

METHOD DETAILS

Stimuli and material

All experiments were created using the open-source python library Psychopy 2022.1.1.[47] Stimuli consisted of 200 ms pure tones with frequencies spanning the 125–8000 Hz range. Stimuli were presented binaurally in a sound attenuated booth, through Etymotic Research ER-2 insert earphones fitted with 13 mm Echodia plugs and connected to a Solid State Logic SSL 2 soundcard. The 200 ms pure tones were apodized with 5 ms hanning windows and hearing thresholds were measured for tones calibrated to ISO 389-2, with reference to an IEC-711 occluded-ear simulator. Psychopy's sound module interprets loudness as a number comprised between 0 and 1, therefore the calibration was performed for an arbitrary Psychopy volume of 0.005 before applying the correction to dB HL. The presentation level was linearly interpolated for frequencies not included in the correction table.

Experimental setup

In experiment 1, participants completed two audiometry tasks: a Randomized paradigm and a 3-alternative forced choice (3-AFC) adaptive staircase paradigm. The 3-AFC adaptive staircase was designed to reproduce the current gold-standard in audiometry procedures and allow us to compare its threshold estimation with other assessments made using unpredictable tones. The Randomized task measures thresholds for pure tones as unpredictable in timing and frequency as possible within this study's limitations.

In experiment 2, the same participants completed two additional audiomety tasks: a fast-paced paradigm, and a slow-paced paradigm, which recorded 4 audiograms each, in different conditions of predictability in time and frequency.

As all volunteers participated in both experiment 1 and 2, we used the randomized paradigm in experiment 1 to confirm inclusion in the study (<20 dB HL hearing loss) and to choose the sound presentation levels in both paradigms of Experiment 2 (see below). To estimate thresholds, all tasks except for the 3-AFC adaptive staircase of Experiment 1 rely on a state-of-the-art, automated pure-tone audiometry procedure by My Medical Assistant SAS[28,29] validated against customary adaptive staircase procedures and based on the automated machine learning described previously.[27]

Experiment 1: Comparison of audiometric paradigms

Three alternative forced choice (3-AFC) paradigm

At each trial, participants were asked to identify one of three possible time intervals during which a tone had been played (Figure 1A). A sequence of 3 shapes flashed on screen (a cross, a disk and a triangle respectively on the left, in the middle and on the right) to indicate the three intervals in time. Each shape was displayed during 300 ms and the next one appeared after a delay of 700 ms. Participants had to press the keyboard arrow corresponding to the correct interval, randomly chosen on a trial-by-trial basis. The experiment started with five practice trials to familiarize participants with the task. We tested 11 frequencies ordered in a butterfly pattern spanning the 125–8000 Hz range (listed in order of presentation: 1000 Hz, 1500 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, 8000 Hz, 750 Hz, 500 Hz, 250 Hz, 125 Hz). The first tone of each frequency was presented at 30 dB HL. The level of presentation of the next tone was determined based on a 2 down – 1 up adaptive staircase procedure. The initial step of 10 dB HL was reduced to 5 dB HL after the first decrease in level. We halved the step again after 3 reversals of the staircase and stopped after 6 reversals. Participants had the opportunity to take a short break before starting the next frequency's staircase. Pure tone thresholds were estimated as the average level across the last 4 reversals. The task lasted on average 20.11 min (sd: 3.36 min).

Randomized paradigm

Participants were asked to detect a tone at low decibel levels, indicating their detection by pressing a key when they heard the tone (Figure 1B). They were informed that detections would only be considered correct if they occurred within a 1-s window following the tone. A fixation cross was displayed on the screen during the whole task. Participants were familiarized with the paradigm using three practice trials at 40 dB HL. The task starts with an initialization phase testing tones spanning the frequency range (listed in order of presentation: 1000Hz, 1500Hz, 2000Hz, 3000Hz, 4000Hz, 6000Hz, 8000Hz, 750Hz, 500Hz, 250Hz, 125Hz) to provide an initial coarse estimation of the audiogram. During this phase, the frequency of tones is predictable, as one frequency is tested several times in a row at decreasing hearing levels, until a negative answer is recorded for that frequency. After the initialization phase, the following sound presentation levels and frequencies were chosen to maximally reduce the uncertainty of the threshold estimation using Bayesian inference. The estimation is based on recent work designed to automate pure tone detection threshold.[27] It uses a Gaussian Process with prior expectation about the relationship between points in the frequency-level space, to assign a probability that a tone of any level and frequency will be heard. The prior is encoded through kernels: one with a covariation matrix using a squared exponential in frequency space, denoting that neighboring frequencies are correlated, and another as a linear function in intensity space, denoting that increased loudness should increase the probability of detection. The next tone is chosen as the point leading to the greatest decrease in the uncertainty of these probabilities, so its frequency cannot be predicted by the participant. The timing in between each tone, the Inter-Stimulus Interval (ISI), was randomly chosen for each tone to be between 1 and 6 s to reduce the participants' ability to predict the timing of upcoming tones. Each tone had a 20% chance of being replaced with silence to be analyzed as catch trials, a fact that participants were not informed of. The true average percentage of catch trials across participants amounted to 18.81% ± 1.57% of trials. The paradigm completes when the Bayesian estimator reaches a confidence interval of 6 dB. With this constraint, the complete paradigm comprised an average of 60.9 trials (min: 51, max: 101, sd: 17.9), including an initialization phase of 25.1 trials (min: 22, max: 28, sd: 1.4), for a total average duration of 6.24 min (sd: 2.20 min).

Experiment 2: Time vs. frequency predictions

For both the slow- and fast-paced paradigms in experiment 2, tones were organized in 'sweeps' whose structures defined four conditions of predictability: predictable timing (T), predictable frequency (F), predictable frequency and timing (FT), or random frequency and timing (R) (Figures 2A and 2B). While both paradigms followed this general structure, the primary difference between these two tasks is the timescale of the Inter-Stimulus Interval (ISI) used to separate successive tones.

In the slow-paced task, participants reported the detection of each tone within a sweep, mimicking clinical procedures that involve explicit responses for every tone. This required using large ISIs (1-3s) to allow time for participants to answer. Each tone thus consisted in a trial and each sweep represented one block. In contrast, the fast-paced paradigm used shorter ISIs (400–800 ms) and clusters of cue tones preceding the target tone, in a setup more aligned with prediction studies. In this task, one sweep represented one trial.

Common elements across paradigms

Sweeps in both paradigms followed a consistent structure to create predictable or unpredictable conditions: For predictable timing, the ISI was kept constant and was randomly selected from a paradigm-dependent range (Slow-paced: 1-3s, Fast-paced: 400-800ms). For predictable frequency, the interval between tones was selected from a set of three musical intervals (whole tone, major third, or perfect fifth) and applied uniformly within each sweep. To create unpredictable sweeps, the intervals between successive tones were randomized (ISIs for time and musical intervals for frequency) from the same distributions chosen for predictable sweeps. We subsequently shuffled the order of the tones for sweeps with unpredictable frequency (T, R).

We measured a separate audiogram in each of the four conditions of predictability and the order of presentation of the conditions was randomized for each sweep. Participants were cued visually about the predictability of each sweep. For predictable frequency conditions, the word "PITCH" flashed on screen for 1 s. For predictable timing, "RHYTHM" flashed similarly. Both words flashed simultaneously before FT sweeps, while fully random sweeps were preceded by a flashing cross.

We used the same model for automated pure-tone audiometry as in the Randomized paradigm of Experiment 1 to estimate audiograms in Experiment 2. The procedure normally starts by collecting initialization data, testing tones spanning the frequency range (1000 Hz, 1500 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, 8000 Hz, 750 Hz, 500 Hz, 250 Hz, 125 Hz) to provide an initial coarse estimation of the audiogram. During this initialization phase, the frequency of the tones is predictable, as one frequency is tested several times in a row at decreasing hearing levels, until a negative answer is recorded for that frequency. To avoid repeating this phase in the two Sweeping paradigms of Experiment 2 where we aim to control the predictability of oncoming tones, we reused the initialization data recorded during the Randomized task to initialize all the audiograms measured in the Sweeping paradigms.

Further details on slow-paced paradigm

Instructions were the same as those for the Randomized task: participants had to press the spacebar less than 1 s after the onset of tones they managed to detect. Sweeps contained 8 to 15 tones, centered around a key tone selected using the Bayesian algorithm from the Randomized paradigm to ensure coverage of tones where the model was most uncertain. After flashing the next sweep's condition for 1 s, a fixation cross was displayed until the end of the sweep. Tones outside the frequency range (125–8000 Hz) were excluded, with sweeps redesigned to maintain a minimum of 8 tones. The presentation levels of non-key tones were randomly set between their detection threshold (established in the prior randomized task from Experiment 1) and the level of the key tone chosen by the model based on data from this paradigm. ISIs were randomly chosen from a range of 1–3 s. Each sweep contained one catch trial which replaced a randomly chosen tone in the sweep from the third position onward (9.20% ± 0.99% of tones across participants). Data from the first 2 tones of each sweep were excluded from analyses in the F, T, and FT conditions since they cannot be predicted using previous tones. The average duration of the slow-paced paradigm was 15.52 min (sd: 3.55 min).

Further details on fast-paced paradigm

Participants were instructed to respond to a target tone by pressing a key while a disk was displayed on the screen. Before the target tone, a cluster of four pure tones was presented alongside a fixation cross, serving as cues for the target tone. Predictable timing conditions used a fixed ISI within each trial (400–800 ms), while random conditions used ISIs randomly drawn from the same range.

After the final cue tone, a circle appeared on the screen to prompt participants to respond, remaining visible for 1.2 s. Participants were instructed to press the key within 1 s of the target tone's onset for their detection to be recorded as valid. Participants were familiarized with the experiment in an initial training phase consisting of five mock trials with feedback.

We again used the algorithm from the randomized task in Experiment 1 to choose the presentation level and frequency of the target tone in every other trial. The target frequency of the remaining 50% of trials was randomly picked from a log-uniform distribution ranging 125 to 8000 Hz, while the hearing level was interpolated from the audiogram estimation made in the Randomized paradigm to ensure that selected tones could not be predicted by any unrealized patterns in Bayesian Inference model. The level of the cues was also interpolated from the Randomized audiogram and raised by 6 dB to increase the probability of their detection. Each target tone had a 20% chance of being replaced with silence to be analyzed as a catch trial (19.53% ± 6.47% of trials across participants). The fast-paced paradigm lasted an average of 25.01 min (sd: 8.91 min).

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis for experiment 1

The 3-AFC and Randomized paradigms use different algorithms to calculate the final threshold: Adaptive Staircase and Bayesian Active Learning, respectively. These methods differ in the continuity of the outputted thresholds (the staircase is necessarily discrete, whereas the Bayesian algorithm is continuous), but otherwise are readily comparable. As such, to account for this difference we sample the threshold values of the Randomized paradigm at the frequencies tested in the 3-AFC. We then average the values across frequency for each participant to get a mean detection threshold comparable with the 3-AFC in two ways: first, averaging across all 11 frequencies of the 3-AFC, and second, averaging over the 4 critical frequencies commonly tested to estimate the pure tone average (PTA) in the clinic, 500, 1000, 2000 and 4000 Hz. The difference in thresholds from each metric are compared using a T-test.

While the 3-AFC is designed to infer a threshold at 70.7% correct, this percentage corresponds to a 56.5% chance of tone detection (accounting for guessing at 33% when tones go undetected). We consider this percentage to be negligibly different from the 50% chance inferred by the Randomized. Furthermore, correcting for the 6.5% difference, if possible, would only enhance the size of the effect between the two conditions as we expect the more predictable task (3-AFC) to result in lower thresholds.

Analysis for experiment 2

After the initial comparison in experiment 1, it is important to control for the difference in methods used to assess thresholds and compare all paradigms under the same footing. As such, when considering all paradigms together (Randomized, 3-AFC, and slow and fast-paced paradigms), we first calculate a global threshold using Bayesian Active Learning as in the Randomized paradigm incorporating tones from all random conditions (Randomized paradigm of Experiment 1 and Random conditions from Experiment 2 protocols - Figure 3C, top panel). This threshold represents an audiogram derived from unpredictable stimuli, independent of specific paradigm protocols. Then for each condition of the slow- and fast-paced tasks (FT, F, T and R), we consider the distance of each tone presented in terms of its intensity from this global threshold (Figure 3C, middle panel). We use this distance as an independent variable in a logistic function to predict whether each tone will be detected or not (Figure 3C, bottom-panel). If the point of equivalence (pEq) of the logistic function is equal to 0, then there is no difference in threshold for this condition relative to global threshold. However, if there is a significant difference in either direction, we can assume the threshold has shifted by this amount. This method allows us to treat each condition in the same manner comparing its outcome as a relative distance from all conditions with unpredictable stimuli.

Statistics for experiments 1 and 2

All statistical analyses were conducted using Python libraries, including SciPy,[48] Statsmodels,[49] Pingouin[50] and Scikit-learn.[51] In experiment 1, audiograms were compared across frequencies and paradigms using two-factor repeated-measures ANOVA, with Greenhouse-Geisser correction applied where the sphericity assumption was violated. Paired t-tests were performed to compare mean thresholds across frequencies and paradigms, using the Benjamini Hochberg method to control the False Discovery Rate (FDR).[52] The false discovery rate (FDR) refers to the expected proportion of false positives among significant results. The Benjamini-Hochberg (BH) procedure controls the FDR by ranking p-values and identifying a threshold at which the proportion of false positives is controlled at a specified FDR level.

To ensure a consistent comparison across frequencies, paradigms and conditions, all subsequent analyses in experiment 2 made use of the calculated pEq values, which serve as a normalized metric representing the relative deviation of the detection threshold from the reference threshold. Separate one-way ANOVAs were implemented for the fast- and slow-paced paradigms to assess differences within paradigms based on predictability levels. Post-hoc pairwise comparisons with Benjamini-Hochberg procedure were performed to determine significant differences between different conditions.

To investigate potential decisional bias, we incorporated catch trials into all paradigms except the 3-AFC task and computed false alarm rates from these trials. We conducted a repeated-measures ANOVA, treating each testing condition – be it the single condition in the Randomized task or any of the four distinct conditions in the fast- and slow-paced paradigms – as a separate group. To identify significant variations in false alarm rates across these conditions, we performed post-hoc comparisons using False Discovery Rate correction.[52] We performed Pearson's correlation analysis to investigate relationships between false alarm rates and pEq threshold differences and between participants' characteristic features (age and musicianship) and their ability to benefit from predictive information. In all tests, the significance level was set at $p < 0.05$. We also used unsupervised hierarchical agglomerative clustering to evaluate the patterns of participant performance across paradigms and predictability conditions. Clustering was performed using SciPy's hierarchical clustering module with linkage using an average method and cosine distance to avoid clustering based on overall performance. All data are presented as means ± standard error of the mean.