Cerebral Cortex, 2021;00: 1–18

https://doi.org/10.1093/cercor/bhab200
Original Article

ORIGINAL ARTICLE

Context Sensitivity across Multiple Time scales with a Flexible Frequency Bandwidth

Tamar I. Regev[1,4], Geffen Markusfeld[2], Leon Y. Deouell[1,2,†] and Israel Nelken[1,3,†]

[1]Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel, [2]Department of Psychology, The Hebrew University of Jerusalem, Jerusalem 9190501, Israel, [3]Department of Neurobiology, The Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel and [4]Current address: MIT Department of Brain and Cognitive Sciences, Cambridge, MA 02139, USA

Address correspondence to Tamar I. Regev. Email: tamar.regev@mail.huji.ac.il; tamarr@mit.edu

[†]Leon Y. Deouell and Israel Nelken have equal senior author contribution.

Abstract

Everyday auditory streams are complex, including spectro-temporal content that varies at multiple timescales. Using EEG, we investigated the sensitivity of human auditory cortex to the content of past stimulation in unattended sequences of equiprobable tones. In 3 experiments including 82 participants overall, we found that neural responses measured at different latencies after stimulus onset were sensitive to frequency intervals computed over distinct timescales. Importantly, early responses were sensitive to a longer history of stimulation than later responses. To account for these results, we tested a model consisting of neural populations with frequency-specific but broad tuning that undergo adaptation with exponential recovery. We found that the coexistence of neural populations with distinct recovery rates can explain our results. Furthermore, the adaptation bandwidth of these populations depended on spectral context—it was wider when the stimulation sequence had a wider frequency range. Our results provide electrophysiological evidence as well as a possible mechanistic explanation for dynamic and multiscale context-dependent auditory processing in the human cortex.

Key words: adaptation, computational modeling, EEG, ERP, human auditory cortex

Introduction

To function efficiently, sensory systems should interpret incoming stimuli in the context in which they are embedded. Contextual effects in audition are often studied using sound sequences that have some regularity, which is infrequently violated by "deviant" sounds. These studies show that in humans, auditory event-related potentials (ERPs) depend on the preceding statistics of the sequence (Sussman 2007; Garrido et al. 2013; Herrmann et al. 2015). However, processing of context is important for any stimulus sequence structure and not just for detecting regularities or change thereof. Here, we studied contextual effects beyond deviance processing, using human EEG recordings, and modeling.

Context integration mechanisms are diverse. One way of efficiently representing context is by summary statistics of past stimulation. Indeed, humans can reliably report the mean pitch of several pure tones (Albrecht et al. 2012; Piazza et al. 2013), and sound textures are represented using time-averaged statistics (McDermott et al. 2013). As the environment constantly changes, estimating summary descriptors dynamically may optimize information transmission (Brenner et al. 2000; Fairhall et al. 2001). Across species and modalities, neuronal

input–output functions scale with statistical properties of the stimulus distribution including mean (Dunn and Rieke 2006; Nagel and Doupe 2006; Dean et al. 2008; Dahmen et al. 2010), variance (Blake and Merzenich 2002; Maravall et al. 2007; Rabinowitz et al. 2011; Herrmann et al. 2013, 2014, 2015), or higher order moments (Kvale and Schreiner 2004; Herrmann et al. 2018, 2020).

The timescales over which context influences neural activity vary widely, from tens of milliseconds to minutes (Fairhall et al. 2001; Khouri and Nelken 2015). Presumably, this variation is necessary because the natural auditory environment contains relevant information at all of these timescales. Sensitivity of auditory neural responses to regularities established across multiple timescales was reported in single A1 neurons in animals (Ulanovsky et al. 2004), human MEG (Teng et al. 2017; Maheu et al. 2019; Teng and Poeppel 2020), or the EEG components MMN and P2 (Costa-Faidella et al. 2011). However, all of the above studies concentrated on regularity/deviance detection.

We investigated context-dependent auditory processing that does not involve regularity/deviance-detection mechanisms by measuring EEG responses to tone sequences with varying frequencies in which all stimuli were equiprobable and task-irrelevant. First, we present a new analysis of data from the control conditions of 2 experiments (1 and 2) previously published (Regev et al. 2019). We show here that in these data, the N1 and P2 event-related EEG components (peaking ~100 and ~180 ms following stimulus presentation, respectively) were sensitive to preceding tone frequencies on distinct timescales. We then designed another study (Experiment 3) to replicate and generalize these exploratory observations.

Previous studies have established that the timescale of N1 sensitivity is longer than ~1 s (Zacharias et al. 2012; Okamoto and Kakigi 2014; Herrmann et al. 2016). Further, the context sensitivity of the N1 component, but not of the P2 component, has been well explained by adaptation models (Herrmann et al. 2013). Importantly, Herrmann et al. (2013) had used a predetermined adaptation time constant to model both N1 and P2. Here, we use a frequency-specific adaptation model (Herrmann et al. 2013, 2014, 2015) to explain our previous results as well as the new dataset, but instead of using a predetermined time constant, we develop a rigorous methodology to quantitatively estimate, from the data, the effective timescales governing context sensitivity at multiple post-stimulus latencies. Our methodology further allowed us to statistically compare between the time constants estimated for distinct response components (N1 and P2).

The hypothesis driving our model in this work is that auditory responses at different latencies (i.e., the latencies of the N1 and P2 components) might be explained by frequency-specific adaptation with different adaptation timescales. Therefore, we expected that our previous observation of a longer timescale of context sensitivity at the earlier post-stimulus latency (of the N1 relative to the P2) will replicate in Experiment 3.

Previous studies had also suggested that the adaptation bandwidth of N1, but not of P2, dynamically adapts to the range of frequencies presented in the stimulus stream (Herrmann et al. 2013) and that this could be caused by cross-frequency adaptation (Taaseh et al. 2011, also termed co-adaptation in Herrmann et al. 2013). Cross-frequency adaptation is the attenuation of the responses of frequency-selective neurons to their characteristic frequency by previous presentations of frequencies away from their characteristic frequency. Therefore, in Experiment 3, we manipulated the overall range of frequencies in the sequences. Following Herrmann et al. (2013), we hypothesized that the adaptation bandwidth of N1 would depend on the spectral range of the stimulation sequence. Importantly, we wanted to know whether this result generalizes to the P2 latency.

To foretell, our model, applied to the extant and to new data, allowed us to confirm that N1 and P2 are sensitive to different timescales of past stimulation, to replicate the dynamic spectral-context-dependent adaptation bandwidth of N1, and to extend it to the P2 latency. We conclude that this type of adaptation, with varying bandwidths and timescales, is a general mechanism shaping auditory ERPs.

Methods

Participants

Eighty-nine healthy adults participated in all 3 experiments—25 musicians, 29 musicians, and 35 nonmusicians in Experiments 1, 2, and 3, respectively. The reason musicians participated in Experiments 1 and 2 is not relevant to the current study and is explained by Regev et al. (2019). Participants were recruited from The Hebrew University of Jerusalem, from Bezalel Academy of Arts and Design, and from the Jerusalem Academy of Music and Dance and could either receive 40 NIS (~12 US$) per hour or course credit for participation in the experiment. The data of 7 participants (4, 1, and 2 from Experiments 1, 2, and 3, respectively) were excluded due to technical difficulties with the recording or excessive rates of artifacts. The analysis therefore included the data of 82 participants—21 in Experiment 1 (7 female, mean age = 29.2 years, standard deviation [SD] = 9 years), 28 in Experiment 2 (15 female, mean age = 24.6 years, SD = 3.6 years), and 33 participants in Experiment 3 (19 females, mean age = 24.6, SD = 2.9 years old). Three additional participants (1 from Experiment 2 and 2 from Experiment 3) were later excluded from data analysis due to unclear auditory responses, resulting in 79 participants overall in the analysis, as explained further in the Data Processing section. All participants self-reported normal hearing and no history of neurological disorders. The experiment was approved by the ethical committee of the faculty of social science at The Hebrew University of Jerusalem, and informed consent was obtained after the experimental procedures were explained.

Stimuli and Apparatus

In all experiments, participants were seated in a dimly lit, sound-attenuated, and echo-reduced chamber (C-26, Eckel) in front of a 17-inch CRT monitor (100-Hz refresh rate), at a viewing distance of about 90 cm. The screen was concealed by a black cover, with a rectangular window in the middle (14 by 8.5 cm), through which they viewed the visual display. Auditory stimuli were presented through headphones (Sennheiser HD25, having a relatively flat frequency response function in the range of frequencies used in the experiment) that were placed over the EEG cap. The experiment was run using the Psychophysics toolbox (Brainard 1997) for MATLAB (version 2013b, MathWorks) running on a 32-bit Windows XP system. Auditory stimuli were synthesized using MATLAB. The experiment included only pure tones, each of 100 ms duration with a 30 ms linear rise and fall ramps. The relatively long ramp was used to prevent low-level onset differences between the pure tones with lowest and highest frequencies, which were spaced up to 3 or more octaves in some of

the sequences. Stimuli were presented at a sound pressure level that was comfortable for the participants. At the beginning of the experiment, each participant adjusted the relative amplitudes of each individual tone, such that all tones had approximately the same subjective loudness.

Experiment Design

Participants viewed silent black and white films while tones were presented to them through headphones. The participants could choose either "The Artist" (Michel Hazanavicius, 2011) or "The Kid" (Charlie Chaplin, 1921), both silent movies. The participants were instructed to ignore the sounds. Each tone sequence in all 3 experiments was composed of pure tones of 5 different frequencies. The specific frequencies varied between block types (see Fig. 1A,B, A and B for an illustration of the stimuli, and a detailed description below). To create the sequences, random permutations of the 5 tones were concatenated successively. If the first tone of the next random permutation was the same as the last tone of the previous permutation, the order of tones in next permutation was reversed. As a result, the order of the tones was random with 3 constraints: 1) each tone occurred exactly in 20% of the sound presentations, 2) a repetition of the same frequency never occurred, and 3) two successive presentations of the same tone frequency were separated by no more than 8 other sounds, imposing a substantial uniformity of tone occurrences over time.

Experiment 1

Two block types that served as control conditions in Experiment 1 of Regev et al. (2019) were used for the current study. Condition 1 included 5 pure tones: Db4, B4, G5, Eb6, and A6 (letter and number notations refer to the pitch class and the octave, International Standards Organization (ISO) system for register/octave designations. The frequencies of the tones were 277.2, 493.9, 784, 1244.5, and 1760 Hz, respectively). Hence, the tones spanned two and two thirds octaves (32 semitones) and the inter-neighbor intervals (frequency intervals between adjacent tones on the sequence-specific frequency axis, see Fig. 1B) were 10, 8, 8, and 6 semitones from low to high frequency. The mean frequency (computed on the logarithmic frequency axis) was that of G5—784 Hz. Condition 2 included Eb4, Db5, A5, F6, and B6 (311.1, 554.4, 880, 1397, and 1975.5 Hz, respectively). These were similar to the tones in Condition 1, but all shifted 2 semitones down. Three blocks of each condition were presented, and their order was counterbalanced between participants. Each block included 500 trials, 100 of each specific tone. This resulted in 300 trials for each specific tone in each condition. The tones were presented with an stimulus-onset asynchrony (SOA, i.e., the time interval between the onsets of 2 consecutive stimuli) of either 450 or 550 ms, randomly (average SOA was 500 ms). As a result, each block lasted 250 s, and there were at least 30 s of silence between the blocks (at the participant discretion). See Figure 1B for illustration of stimuli.

Experiment 2

One block type that served as a control condition in Experiment 2 of Regev et al. (2019) was used for the current study. This condition included 5 tones with equal probabilities: Db4, C5, F#5, D6, and B6 (277.9, 523.2, 740, 1174.7, and 1975.6 Hz, respectively). Hence, the tones spanned two and five sixths octaves (34 semitones) and the inter-neighbor intervals were 11, 6, 8, and 9 semitones from low to high frequency. The mean frequency (computed on the logarithmic frequency axis) was that of F#5–740 Hz. Two blocks of this condition were presented. Each block included 550 trials presented with an SOA of 400 ms. This resulted in 220 trials for each specific tone. Each block lasted 220 s, and there were at least 30 s of silence between the blocks (at the participant discretion). See Figure 1B for illustration of stimuli.

Experiment 3

In this experiment, we manipulated the overall range of frequencies in the sequences (Table 1). Five block types were used. Block 1 with a wide range of frequencies (Wide; three and a quarter octaves between the lowest and highest tone, 39 semitones), blocks 2a and 2b with a medium range (Medium; two and a third octaves, 28 semitones), and blocks 3a and 3b with a narrow range (Narrow; one and a sixth octaves, 14 semitones). The mean inter-neighbor intervals were 9.75, 7, and 3.5 semitones in the Wide, Medium, and Narrow range conditions, respectively. For the Medium and Narrow conditions, we designed 2 different block types (a and b), transposed by 7 semitones, in order to generalize the results beyond the specific (absolute) mean or range of frequencies (Fig. 1B). Each block type was presented 3 times, resulting in 15 blocks altogether. The order of the blocks was randomized for each participant separately, with the constraint that successive blocks had to be of different types. In every block 540 tones were presented in total, resulting in 324 trials overall for each tone frequency in each block type.

The frequencies of all tones were taken from the C major scale, in order to prevent a situation in which one of the tones would become harmonically deviant and therefore result in stronger ERP responses (Poulin-Charronnat et al. 2006; Koelsch 2009).

The SOA was randomly set to one of 5 possible values: 450, 475, 500, 525, or 550 ms. As a result, the duration of each block was about 270 s (4.5 min), and there were at least 45 s of silence between blocks (at the participant discretion). In total, the EEG was recorded for approximately an hour and a half.

EEG Recording and Preprocessing

EEG was recorded from 64 pre-amplified Ag/AgCl electrodes using an Active 2 system (BioSemi, the Netherlands), mounted on an elastic cap according to the extended 10–20 system (https://www.biosemi.com/pics/cap_64_layout_medium.jpg), with the addition of 2 mastoid electrodes and a nose electrode. Horizontal electrooculogram (EOG) was recorded from electrodes placed at the outer canthi of the right and left eyes. Vertical EOG was recorded from electrodes placed below the center of both eyes and above the center of the right eye. This resulted in 72 recording electrodes overall. The EEG signals were measured during acquisition relative to a Common Mode Sense active electrode placed over a parietal area. They were sampled at a rate of 512 Hz (24 bits/channel) with an online antialiasing low-pass filter set at one-fifth of the sampling rate and stored for offline analysis.

EEG preprocessing was conducted using BrainVision Analyzer 2.0 (Brain Products) and MATLAB (2016b, MathWorks). The following preprocessing pipeline was applied in all experiments: First, since we paused recording during the break between blocks, de-trending was applied using MATLAB, subtracting long-term linear trends from each block, thus zeroing the signal at beginning and end of blocks and avoiding discontinuities at the border between blocks. Then, further preprocessing

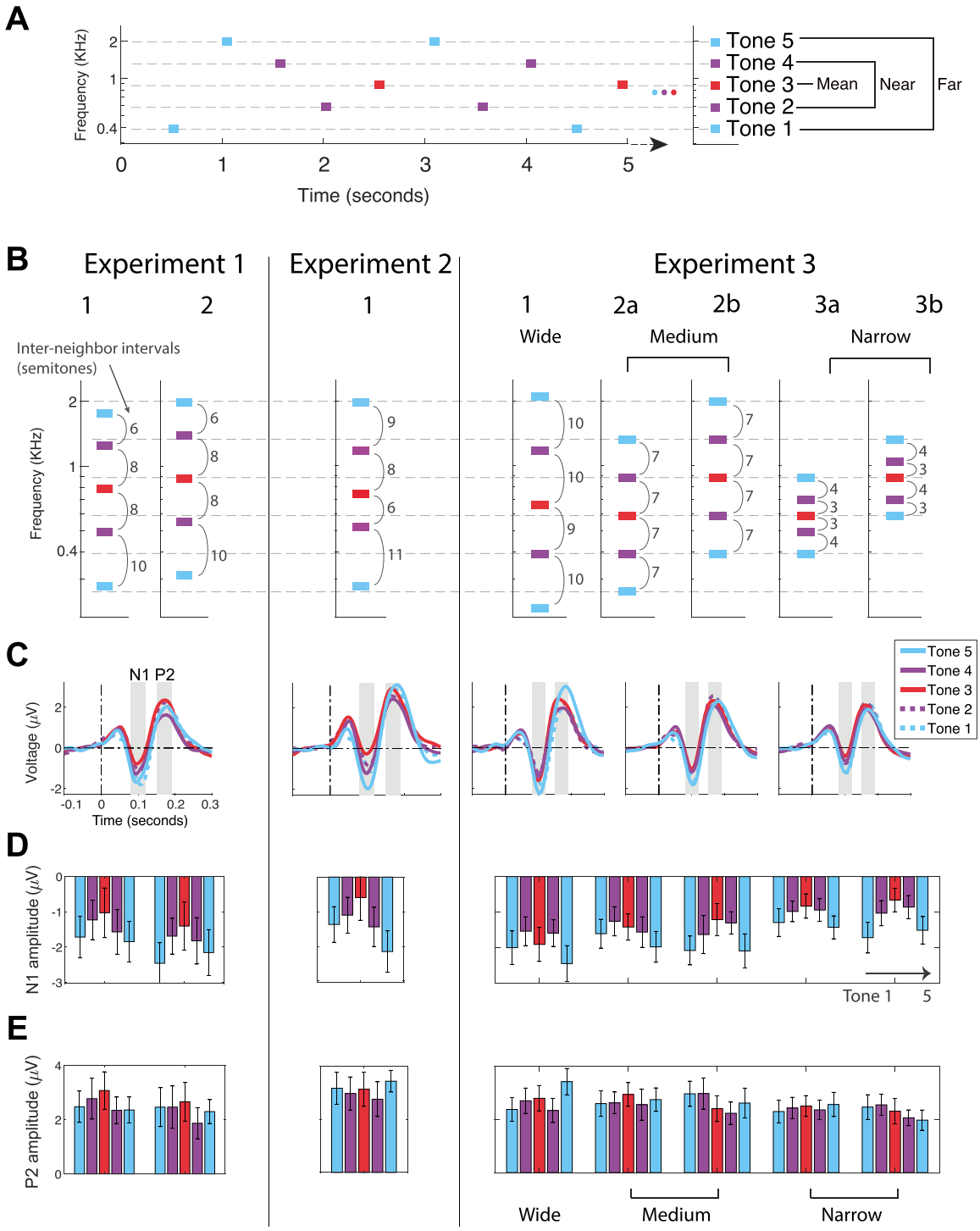

Figure 1. N1 but not P2 is sensitive to long-term context. (*A*) An example segment of a tone sequence in the experiment (from block type 2b of Experiment 3). (*B*) Stimuli used in all experiments and conditions. Intervals between neighboring tones on the frequency axis (inter-neighbor intervals) are displayed in semitones. (*C*) ERPs for tones 1–5 (low to high frequency), calculated for each experiment. For Experiment 3, ERPs are plotted for each frequency range, pooling together block types 2a + 2b and 3a + 3b. (*D*) Mean and 95% CIs (across participants) of N1 peak amplitudes. For each block type, the 5 bars correspond to tones 1, 2, 3, 4, and 5 (lowest to highest) from left to right and the bar colors match the color scheme in *A* and *B*. (*E*) Same as *D* for peak amplitudes of P2.

was done in Analyzer, using the following pipeline: 0.1 Hz high-pass, zero-phase-shift second-order Butterworth filter; referencing to the nose electrode; correction of ocular artifacts using independent component analysis (Jung et al. 2000) based on typical scalp topography and time course; and, finally, discarding epochs that contained other artifacts. The latter part was semi-automatic—first, an algorithm marked artifacts based on predefined criteria: absolute difference between samples

Table 1. Description of stimuli in Experiment 3

Condition	1 (wide)	2a (medium)	2b (medium)	3a (narrow)	3b (narrow)
Notes	A3, G4, E5, D6, C7	C4, G4, D5, A5, E6	G4, D5, A5, E6, B6	G4, B4, D5, F5, A5	D5, F5, A5, C6, E6
Frequencies (Hz)	220, 392, 659.26, 1174.7, 2093	261.6, 392, 587.33, 880, 1318	392, 587.3, 880, 1318, 1975.5	392, 493.88, 587.33, 698.46, 880	493.88, 698.46, 880, 1046.5, 1318
Inter-neighbor intervals[a] (low to high, semitones)	10, 9, 10, 10	7,7,7,7	7,7,7,7	4,3,3,4	3,4,3,4
Mean frequency (note, Hz)	F5, 698.46	D5, 587.33	A5, 880	D5, 587.33	A5, 880
Range (octaves)	$3\frac{1}{4}$	$2\frac{1}{3}$	$2\frac{1}{3}$	$1\frac{1}{6}$	$1\frac{1}{6}$
Range (semitones)	39	28	28	14	14
Mean Inter-neighbor interval (semitones)	9.75	7	7	3.5	3.5

Note: Tone properties are listed from low to high.
[a]Mean interval between nearby tones on the sequence-specific frequency scale, Figure 1B.

>100 μV within segments of 100 ms; gradient >50 μV/ms; absolute amplitude >100 μV; or absolute amplitude <0.5 μV for a duration of more than 100 ms. If an artifact was detected using any of these criteria, an epoch of 200 ms around it was marked. Then, we performed visual inspection of all data to remove or add rare artifacts that were missed or marked by mistake. Artifact rejection was performed on 30 Hz low-passed data, causing the artifact rejection process to be blind to high-frequency noise, which did not interfere with our analysis. Then, the data were exported from Analyzer to Matlab (without the 30 Hz low-pass filter). Finally, using MATLAB, a 1–20 Hz band-pass zero-phase-shift fourth-order Butterworth filter was applied to the continuous data, followed by segmentation and averaging.

Data Processing

We calculated ERPs locked to auditory stimulus presentation. The ERP amplitudes were measured from the midline central Cz electrode. This location was selected because it maximizes the N1 and P2 responses and is typically used to measure these components (e.g., as in Tremblay et al. 2001). The data were parsed into segments beginning 100 ms before the onset of tone presentation and ending 400 ms after the onset of tone presentation. The average amplitude of the 100 ms pre-stimulus time served as a baseline for amplitude measurements. ERPs were obtained for each participant, by separately averaging trials of every block type and every tone, conditioned on every possible previous tone. This resulted in 5 tones × 4 previous tones (because there were no repetitions) × number of block types (2, 1, or 5 in Experiments 1, 2, and 3, respectively) ERPs. We then calculated the peak amplitudes of the N1 and P2 components for each ERP yielding 40, 20, and 100 measurements per participant and potential type in Experiments 1, 2, and 3, respectively.

N1 and P2 peak amplitudes were calculated in 2 stages. First, to minimize misidentification of peaks due to noise, we determined the peak time from the average of all presentations of each tone (i.e., not conditioned on the previous tone frequency) for each block type and participant. The resulting ERPs were thus based on a large number of trials and had satisfactory signal-to-noise ratio. We defined the N1 latency as the time of the absolute minimum (most negative) in the time window between 50 and 150 ms after stimulus onset and the P2 latency as the time of the absolute maximum in the time window between 130 and 250 ms after stimulus onset. If the peak latencies corresponded to the edge of the corresponding time windows, the participant was excluded from data analysis. In total, 0, 1, and 2 participants were excluded for this reason from Experiments 1, 2, and 3, respectively, leaving 20, 27, and 31 participants in the analysis of the experiments. We then computed the average voltage in a 12-ms window around the detected peak time for every possible combination of tone frequency with a previous tone frequency. Thus, peak latencies were determined for each tone frequency and block type regardless of previous tones, whereas the peak amplitudes were measured around these latencies, contingent on the previous tones. Overall this resulted in 4480 data points for each potential type (participants × measurements per participants, see end of previous paragraph; 21 × 40 = 840, 27 × 20 = 540 and 31 × 100 = 3100 for Experiments 1, 2, and 3, respectively), or 8960 (4480 × 2 potential types, N1 or P2) altogether.

Statistical Analysis of ERP Peak Potentials

To test the effect of both long- and short-term context on N1 and P2 peak amplitudes, we used linear mixed-effects (LME) models. LME models were run in Matlab 2016b using the *fitlme* function. To be able to compare between the N1 and P2 amplitudes, the data were standardized using a z-score transform on the N1 and P2 separately, after multiplying N1 data points by −1. N1 and P2 amplitude values were modeled using 8 fixed factor predictors, 4 for each potential type. The 4 predictors were as follows: an intercept, 2 continuous slope variables termed: "Interval-Mean" (long-term context: frequency interval between current tone and mean frequency overall in the sequence, semitones), "Interval-Previous" (short-term context: interval between current and previous tone frequencies, semitones) and another slope variable representing interaction between the 2 latter variables, encoded as their product: "Interval-Mean ∗ Interval-Previous." Random effect factors were added for all of the 8 fixed factors, grouped by participant number. In general, to determine whether a factor should be part of the model, we compared the 2 (nested) models trained with and without this specific factor using a likelihood ratio test (Matlab "compare" routine). If the likelihoods of the 2 models were not significantly different, that

factor was excluded. Thus, in this model, we omitted the random (but not the fixed) slope of the interaction variables "Interval-Mean ∗ Interval-Previous" (both for N1 and P2) since they did not contribute to the overall explained variance (Likelihood ratio test between the model with and without these factors; $\chi^2(2) = 3$, $P = 0.21$). This resulted in 6 random effect terms in the LME model. The final LME model is described with the following Wilkinson formula (Wilkinson and Rogers 1973):

$$\begin{aligned}\text{Voltage} \sim\ & I(N1) + I(P2) + \text{Dist}_{\text{mean}} \cdot I(N1) + \text{Dist}_{\text{mean}} \cdot I(P2) \\ & + \text{Dist}_{\text{prev}} \cdot I(N1) + \text{Dist}_{\text{prev}} \cdot I(P2) + \text{Dist}_{\text{mean}} \cdot \text{Dist}_{\text{prev}} \cdot I(N1) \\ & + \text{Dist}_{\text{mean}} \cdot \text{Dist}_{\text{prev}} \cdot I(P2) + (I(N1)|s) + (I(P2)|s) \\ & + (\text{Dist}_{\text{mean}} \cdot I(N1)|s) + (\text{Dist}_{\text{mean}} \cdot I(P2)|s) + (\text{Dist}_{\text{prev}} \cdot I(N1)|s) \\ & + (\text{Dist}_{\text{prev}} \cdot I(P2)|s) \end{aligned} \quad (1)$$

where Voltage is either N1 or P2 standardized amplitudes of the responses to a specific tone (given all previous tone possibilities, for each experiment, condition and participant). $I(N1)$ and $I(P2)$ are indicator functions for N1 or P2 (each being 1 when the voltage belongs to the corresponding class and 0 otherwise). They therefore represent separate intercepts for N1 and for P2. $\text{Dist}_{\text{mean}}$ stands for Interval-Mean, $\text{Dist}_{\text{prev}}$ stands for Interval-Previous. For example, $\text{Dist}_{\text{prev}} \cdot I(N1)$denotes the $\text{Dist}_{\text{prev}}$ slope variable contributing to the N1 amplitudes. $(X|s)$ denotes the random variable X grouped by participant number. The variable X is always assumed to be normally distributed with a mean of 0, and its variance is estimated from the data. Thus, $(I(N1)|s)$ denotes a subject-specific contribution to the intercept for the N1 measurements; $(\text{Dist}_{\text{prev}} \cdot I(N1)|s)$ is a subject-specific contribution to the corresponding slope. This LME model was estimated from data points from all 3 experiments together, resulting in 8960 data points overall (see end of Data Processing section), collected from 79 participants.

Note that this is almost identical to modeling each potential type (N1 or P2) separately by:

$$\begin{aligned}\text{Voltage} \sim\ & I + \text{Dist}_{\text{mean}} + \text{Dist}_{\text{prev}} + \text{Dist}_{\text{mean}} \cdot \text{Dist}_{\text{prev}} + (I|s) + \\ & (\text{Dist\_prev} * I|s) + (\text{Dist\_mean} \cdot I|s) \end{aligned} \quad (2)$$

However, there are some distinctions. For instance, in the way we estimated the model the residual error is calculated overall from all data together, and therefore, it is more appropriate for statistical comparisons between the estimates of N1 and P2, which was one of our main goals.

For each fixed-effect coefficient, a standardized effect size, Cohen's *d*, was computed by dividing the estimate by its SD. The SD was calculated from the estimate standard error (SE) provided by Matlab function fitlme, multiplying it by the square root of the number of degrees of freedom, DF (DF = 78; number of participants – 1). The significance value of each individual coefficient (ANOVA comparing it to 0) was given by the *fitlme* model output. To statistically compare between the contributions of 2 (or more) coefficients, we ran a post hoc coefficient test (*F*-test) for LME estimates, using the *coefTest* function in Matlab. Cohen's *d* of these effects was calculated as the square root of F/DF (DF = 78, as above).

To ensure the robustness of the results, we also ran a 2-level analysis, commonly used for example to analyze group results in functional MRI studies. A linear regression was run for each participant with regressors similar to the fixed effects above. We then performed a second-level analysis of the estimates using paired *t*-tests on participant-specific estimate values.

To test the interaction of the overall frequency range of tones in the sequence with the effects estimated by the LME analysis described above, we ran another LME model adding interactions with the continuous slope variable range (overall frequency range in the sequence, semitones: Wide: 39, Medium: 28, and Narrow: 14). For each term in the above model, we added another interaction term with range. This resulted in a large number of variables in the model and therefore we omitted the higher order interactions that did not contribute significantly to the overall explained variance, according to a likelihood ratio test as described above. All random terms including the range variable as well as both fixed and random effects of the Interval-Mean ∗ Interval-Previous terms did not contribute significantly to the overall explained variance (Likelihood-ratio test between the models with and without all of the latter terms; $\chi^2(12) = 6.5$, $P = 0.88$) and therefore were omitted from this model. However, since the fixed factors of the interaction terms Interval-Mean ∗ Interval-Previous were included in the previous model, we estimated as well a model including these variables and verified that the results were comparable with and without them. These LME models were run only on data points from Experiment 3, in which the range was manipulated within subject, resulting in 6200 data points (3100 datapoints per potential type × 2 potential types, N1 or P2, see end of first paragraph of Data Processing section) collected from 31 participants. Effect sizes were calculated similar to the above, using DF = 31 – 1 = 30.

Single-Trial EEG Amplitude Extraction

We calculated single-trial EEG amplitudes as the average voltage in a window of 12 ms centered around the latency of the N1 and P2 as determined by the subject's ERP (i.e., average across trials). After excluding trials with artifacts, we were left with 2660, 1100, and 7449 trials per participant on average in Experiments 1, 2, and 3, respectively (resulting in 55 863, 29 701, 230 931, or 316 495 trials overall used to train the model for Experiments 1, 2, 3, or when using all data together, respectively) for either N1 or P2. We also fitted the model separately for each of the frequency range conditions (Wide, Medium, and Narrow) in Experiment 3. The total number of observations used in each of these conditions was 46 136, 92 874, and 91 921, respectively (recall that there were 2 versions of the Medium and Narrow conditions, Fig. 1A, hence the larger number of trials).

Adaptation Model

We used frequency-specific adaptation equations (Herrmann et al. 2013, 2014, 2015) to model single-trial N1 and P2 responses. The model assumes frequency-tuned neural populations with Gaussian response profiles over a log-frequency axis. Each population has resources, which determine the size of the responses that it can generate and are depleted in proportion to these responses. Therefore, resource depletion causes the contextual attenuation of the responses. Between sound presentations, resources recover exponentially with time. The amount of resources that are depleted is termed response adaptation (RA) following Herrmann et al. (2013, 2014, 2015). The RA values just before the presentation of a tone were calculated recursively for

each specific stimulus sequence, trial by trial, using Equation (3).

$$RA_{i,t+1} = \left(RA_{i,t} + (1 - RA_{i,t})\, e^{-\frac{1}{2}\left(\frac{\log(f_i)-\log(f_t)}{\sigma}\right)^2} \right) \cdot e^{\frac{-\Delta t_{S_t \to S_{t+1}}}{\tau}} \quad (3)$$

Here $RA_{i,t}$ is the response adaptation of neural population i centered around frequency f_i, at time step t of the stimulus sequence, in which stimulus S_t with frequency f_t was presented. RA ranges between 0 (no adaptation) and 1 (maximal adaptation) and therefore 1-RA is the amount of available resources and is therefore proportional to the response of neurons (full adaptation corresponds to minimum responsiveness and vice versa). The Gaussian term reflects the frequency tuning of the neural population. It determines the size of the response evoked from population i by the stimulus presented at time t. The response causes further resource depletion that is determined by the frequency interval between the current tone presented in the sequence and f_i, the best frequency of neural population i. The time interval $\Delta t_{S_t \to S_{t+1}}$ is the time between the onset of the stimulus at time step t and the next stimulus at time step $t+1$. During this time, resources recover and thus RA decreases exponentially.

There are 2 parameters in Equation (3): the Gaussian width of the frequency profiles—σ, and the time constant of the exponential recovery—τ. These were termed together $\vec{\theta} = (\sigma, \tau)$. Once $\vec{\theta}$ is given, Equation (3) allows the computation of the adaptation level for each neuronal population and at each time point.

Next, the model assumes a linear relationship between the measured EEG data and RA of the population at the presented tone frequency (Equation 4).

$$\text{Voltage} = a + b \cdot RA \quad (4)$$

where Voltage is the measured EEG amplitude, either N1 or P2, and a and b are the linear factors associating RA and the data. Since RA is defined between 0 and 1, these factors allowed for appropriate scaling and shift to the EEG units of measurement. The inverse relation between RA and the responsiveness of neurons (1-RA) is captured by the values of a and b.

$RA_{i,t}$ was first computed for all the 5 tone frequencies i and all t time steps in a specific sequence. Each participant heard different stimulus sequences and therefore RA was computed for each participant separately. Then, to predict responses at each time step t, the relevant RA was taken as that of the neural population i, corresponding to the tone frequency presented at time step t, as in Herrmann et al. (2013, 2014, 2015). The last step assumes that EEG responses measured at time step t are dominated by neural populations centered around the frequency of the presented stimulus $f_i = f_t$. The procedure we used was the one used by Herrmann et al. (2013, 2014, 2015), and we used it in order to be consistent with them. A more natural procedure would consist of integrating the RA values for all neuronal populations, weighted by their tuning profiles (the exponential term in Equation 3). We verified that the 2 methods are comparable by calculating the full RA predictions, weighted by the tuning profiles, for one specific participant and plotting them against the simplified model. The 2 models were indeed strongly linearly related to each other (Supplementary Fig. S1).

Model Fit and Parameter Estimation

All data analysis related to model fitting was carried out using Matlab 2016b or 2019b (Mathworks, MA, USA), and all the code is available online (http://osf.io/mswhv). In total, the model had 4 free parameters—the 2 mechanistic parameters $\vec{\theta} = (\sigma, \tau)$, and the linear factors a and b. All parameters were estimated from the data. However, $\vec{\theta}$ and the linear factors were treated differently. $\vec{\theta}$ are the parameters of interest for this study, while the linear factors are "nuisance" parameters that have to be fitted but are not interpreted. Therefore, fitting the model was done in 2 steps. First, single-trial model predictions (RA) were calculated for a range of pre-defined possible $\vec{\theta}$ values; 18 values for σ spanning 1–18 semitones and 25 for τ spanning 0.2–5 seconds with 0.2 second steps (resulting in 450 possible $\vec{\theta}$ combinations). Second, for each $\vec{\theta}$, single-trial EEG responses were regressed on the computed RA values. In practice, a LME model was fitted (using Matlab *fitlme*) using RA as a continuous fixed effect with random intercept and slopes grouped by participant using the Wilkinson formula (Wilkinson and Rogers 1973):

$$\text{Voltage} \sim 1 + RA + (1 + RA|\text{participant}) \quad (5)$$

This approach made it possible to estimate participant-specific linear factors while reducing the amount of overfitting that would be generated by estimating the linear factors of each participant separately. Log-likelihood (LL) statistics of the LME models were extracted for each value of $\vec{\theta}$. Then, $\vec{\theta}^{max}$, the $\vec{\theta}$ value resulting in the maximum likelihood, was selected as the best estimate for the parameters of interest.

We fitted the model to each experiment separately and to the data from all experiments together as well. When fitting the model to Experiment 3, we estimated separate fixed-effect slopes and intercepts for the different conditions composing this experiment:

$$\text{Voltage} \sim 1 + RA * \text{condition} + (1 + RA|\text{participant}) \quad (6)$$

where condition is a categorical variable encoding the 3 range conditions. When fitting the model using the data of all experiments together, we used the same formula as for Experiment 3 (Equation 6) encoding 5 different conditions (Experiments 1, 2, and the 3 separate conditions of Experiment 3). Including the RA * condition interaction in the random term resulted in much longer calculation times, which were unfeasible for the nonparametric statistical tests (bootstrap, permutations, see below), but we verified that it yielded identical $\vec{\theta}^{max}$ values for the parameter fits of each experiment separately.

Testing Significance of Model Fit

To test whether the adaptation model generally described the data better than chance, we repeated all the steps of parameter estimation for randomly permuted measurements. For permuted data, the model was not expected to perform better than chance and therefore this allowed calculation of the distribution of the LL statistic under the null hypothesis of no effect of sequence order. We used 100 permutations. In particular, for each permutation of the data, we collected the maximum log-likelihood (LL^{max}) values (over all possible $\vec{\theta}$ values) and plotted the null distributions of maximum LL values for N1 and P2

separately. The maximum LL of real data was compared to the null distributions of maximum LL.

Confidence Regions for the Parameter Estimates

To calculate a confidence region around $\vec{\theta}^{max}$, we asked which $\vec{\theta}$ values are significantly different from $\vec{\theta}^{max}$. Due to Wilk's theorem the quantity

$$D\left(\vec{\theta}\right) = -2 \cdot \ln\left(\frac{\text{likelihood}\left(\vec{\theta}\right)}{\text{likelihood}\left(\vec{\theta}^{max}\right)}\right) \quad (7)$$

should be distributed as χ^2 with 2 DF under the null hypothesis that $\vec{\theta}^{max}$ does not describe the data better than any other parameter value, and for a big enough sample size. We also verified this assumption empirically, simulating the null distribution of D at 2 representative values of $\vec{\theta}$ for N1 and P2 data and showing that it is comparable to χ^2 with 2 DF. For the confidence regions, we calculated the *D* statistic for all possible $\vec{\theta}$ values and asked whether it is significant using a χ^2 test with 2 DF. A large *D* is expected for $\vec{\theta}$ values that are significantly different from $\vec{\theta}^{max}$. The *D* values at $\vec{\theta}^{max}$ always equal to 0 by definition. The threshold of $D < 6$, which is approximately the value corresponding to a $P = 0.05$ for the χ^2distribution with 2 DF, was used to define the 95% confidence region around $\vec{\theta}^{max}$.

Comparing Parameter Estimates for N1 and P2

We statistically contrasted the values of $\vec{\theta}^{max}$ estimated for N1 versus P2 using 3 methods. First, we compared the $\vec{\theta}^{max}$values of N1 and of P2 by checking whether the $\vec{\theta}^{max}$ of, for example, N1, fell outside the confidence region of P2 and vice versa (see previous section for the calculation of confidence regions). Second, we performed bootstraping (random sampling with replacement) on the groups of participants (of either Experiment 1 2 or 3 or when using all participants together), and repeated the estimation procedure 100 times. This resulted in an estimate of the distribution of $\vec{\theta}^{max}$ values (under the assumption that the group of participants represents the population). The distribution of differences of $\vec{\theta}^{max}$values calculated in the same bootstrap repetition for N1 versus P2 was compared to 0. Third, we used data permutations to create the null distribution of the difference between LL at $\vec{\theta}^{max}$ of the 2 potential types (termed log-likelihood differences, LLD). For each of 100 permutation repetitions, we flipped between N1 and P2 of single trials, with probability 0.5 per each trial flip. For each iteration, we repeated the estimation procedure and computed the LLD. We thus estimated the null distribution of LLD under the assumption that the parameters of N1 and P2 were identical and calculated the *P*-value of the LLD of the actual data by comparing it to this null distribution. We emphasize that our main goal was to compare between parameter values estimated for the N1 and P2 latencies in each experiment rather than giving exact absolute values for the parameters.

Comparing Parameter Estimates for Different Frequency Ranges

$\vec{\theta}^{max}$ values were estimated for each of the stimulus frequency range conditions (Wide, Medium, and Small) in Experiment 3 by fitting the model to the data of each of the range conditions using Equation 5 and selecting $\vec{\theta}^{max}$ in each condition separately. Parameter estimation was then repeated 100 times using a bootstrap procedure—simulating new groups of participants by random sampling with replacement from the pool of 31 participants. To statistically test the effect of frequency range on adaptation bandwidth of N1 and P2, we fitted an LME model to the bootstrapped σ^{max} values. The bootstrap repetition number was used as the grouping variable for the random effects. The stimulus frequency range variable was modeled as a continuous fixed effect with random intercept and slope. Thus, this LME model is described with the following Wilkinson formula:

$$\begin{aligned}\sigma^{max} \sim\ & I(N1) + I(P2) + \text{range} \cdot I(N1) + \text{range} \cdot I(P2) + (I(N1)|\text{bootN}) \\ & + (I(P2)|\text{bootN}) + (\text{range} \cdot I(N1)|\text{bootN}) \\ & + (\text{range} \cdot I(P2)|\text{bootN}) \end{aligned} \quad (8)$$

where $\sigma^{max}$are the estimated adaptation bandwidths in each of 100 bootstrap runs, in each of the 3 range conditions and both for N1 and P2 data (600 values in total), and $I(N1)$ and $I(P2)$ are indicator functions for N1 or P2, respectively. Thus, the I represents separate intercepts for N1 and P2. Range is the overall range of frequencies presented in a sequence, in semitones (so range · I(N1) represents the effect of range on the N1 bandwidths) and bootN is the bootstrap run number serving as a grouping variable for the random effects. Effect size (Cohen's *d*) was calculated as above (DF = 100 − 1 = 99, for 100 bootstrap repetitions).

Results

In 3 EEG experiments, 79 participants (21, 27 and 31 in Experiments 1, 2, and 3) were presented with sequences of 5 equiprobable pure tones (Fig. 1A), which they were instructed to ignore while concentrating on a silent film. Using a passive paradigm allowed us to investigate neural effects that are elicited automatically and do not depend directly on attention or on any active task. In Experiments 1 and 2, we analyzed the control conditions from a previously published study (Regev et al. 2019; the published study included also conditions involving rare deviants, which were the focus of that study. We do not analyze these conditions here as we focus on equiprobable sequences). The sequences in these experiments had a relatively wide frequency range (32 and 34 semitones between lowest and highest tones used in the sequence, respectively, Fig. 1B, Methods). Experiment 3 included 3 frequency range conditions: Wide, Medium, and Narrow (39, 28, and 14 semitones, respectively). We examined the amplitudes of the N1 and P2 auditory-evoked responses and their dependence on 2 features of the tones along the sequence: 1) The frequency interval between the current tone and the mean sequence frequency (Interval-Mean) and 2) the frequency interval between the current tone and the previous tone frequency (Interval-Previous). Interval-mean represented a long-term context variable, because in order to show sensitivity to the overall mean in the sequence, integration of several previous tones should occur (at least 5 previous tones, ~2.5 s). Interval-previous represented a short-term context variable, at the scale of 1 SOA (~0.5 s).

Table 2. Linear mixed-effects (LME) results—effect of long- and short-term context on N1 and P2

		Estimate	SE	F(1,8952)	P-value	*d*
Predictors						
Intercept	N1	−0.51	0.078	42.2	8.8E−11	−0.73
	P2	−0.18	0.074	5.80	0.016	−0.27
Interval-Mean	N1	0.04	0.004	70.0	6.7E−17	0.94
	P2	−0.002	0.005	0.15	0.7	−0.04
Interval-Previous	N1	0.01	0.003	26.9	2.2E−7	0.58
	P2	0.02	0.003	31.8	1.7E−8	0.63
Interval-Mean ∗ Interval-Previous	N1	−0.0008	0.0002	10.08	0.0015	−0.36
	P2	−0.0002	0.0002	0.96	0.32	−0.11
Pairwise comparisons between predictors						
N1 vs. P2 Interval-Mean				35.5	2.6E−9	0.67
N1 vs. P2 Interval-Previous				0.23	0.62	−0.05
N1 Interval-Mean versus -previous				25.4	4.6E−7	0.57
P2 Interval-Mean versus -previous				21.41	3.7E−6	−0.52
N1 Interval-Mean versus -previous versus P2 Interval-Mean versus -previous				46.3	1.03E−11	0.77

Note: Top—N1 and P2 amplitudes (standardized using a z-score transform after negating N1 data points) were modeled using the 8 predictors listed in the first (left-most) column. The model consisted of fixed and random factors (grouped by participant) for each of the listed predictors (except for the interaction term Interval-Mean ∗ Interval-Previous, which did not have a random factor since the latter did not contribute to explained variance), Methods. Columns 2–5: Fixed-effect estimates (Estimate), standard errors of the estimates (SE), *F*-statistic used for ANOVA comparing the estimates to 0, with degrees of freedom, significance level (*P*-value) of the latter test, standardized effect size (Cohen's *d*, Methods). The predictors Interval-Mean and Interval-Previous stand for the frequency interval between the current tone and sequence mean or current tone and previous tone frequency, respectively, in semitones. Bottom—Post hoc pairwise coefficient comparison between predictors. The LME model was run on 8960 observations collected from 79 participants overall in the 3 experiments (Methods).

N1 but Not P2 Is Sensitive to Long-Term Context

Absolute N1 amplitudes increased as a function of the frequency interval between the current tone and overall mean frequency in the sequence. This dependence manifested itself as a typical inverted U-shape pattern, so that the most negative N1 amplitudes were elicited in response to the most extreme tones and the least negative N1 was elicited by the middle tone (which was also approximately equal to the mean frequency of the sequence). This phenomenon was robust and replicated in all 3 experiments (Fig. 1C,E). In contrast, P2 amplitudes did not show significant dependence on the mean sequence frequency (Fig. 1D,E). To quantify this effect, we used a linear mixed-effects (LME) model including data from all experiments together (Table 2 and Fig. 2). The difference from the mean significantly affected the N1 amplitude, but not the P2 amplitude (Table 2). The difference between the N1 and P2 Interval-Mean effects (slopes) was significant (Table 2, Fig. 2—contrast #3). A two-stage procedure including linear regression on individual participants followed by second-level analysis at the group level gave similar results (Supplementary Fig. S2 and Supplementary Table S1).

P2 Is More Sensitive than N1 to Short-Term Context

N1 and P2 absolute amplitudes increased as a function of the interval between the current and previous tone frequencies, but this effect was larger for P2 than for N1 (Fig. 3 and Supplementary Fig. S3). To visualize this, we pooled the possible combinations of current and previous tones according to the "degree of neighborhood." Neighbor 1–4 denotes the proximity of tones on the frequency axis in a specific sequence (Fig. 3A,B). Figure 3C,D compares the N1 and P2 peak amplitudes for when the previous stimulus was "Neighbor 1" versus "Neighbor 2." We concentrated just on the "Neighbor 1" and "Neighbor 2" groups since they were composed of more combinations of current and previous tones and included all current tones, whereas by design only extreme tones in every sequence could have neighbor "3" and "4." The difference between "Neighbor 1" and "Neighbor 2" amplitudes was larger for P2 than for N1 in 7 out of 8 block types overall in the 3 experiments (Fig. 3D).

The LME model including data from all experiments together (Table 2, Fig. 2) indicated that the frequency interval between the current and previous tone (Interval-Previous, semitones) significantly affected both N1 and P2 amplitudes (Table 2). The effect size of Interval-Previous was nominally larger for P2 than for N1 but they were not significantly different. Notably, excluding the interaction term between Interval-Mean and Interval-Previous from the LME model reduced the short-term (Interval-Previous) context effect for N1 but not for P2 (Supplementary Table S2), resulting in a significant difference between N1 and P2 Interval-Previous effect. Thus, whereas the dependence of P2 on short-term context was robust, the dependence of N1 on short-term context interacted with its dependence on long-term context. Regressions on individual participants and second-level analysis of regression estimates gave similar results (Supplementary Fig. S1 and Supplementary Table S1).

Sequence Frequency Range Affects N1 but Not P2 Amplitudes

N1 amplitudes were reduced when the frequency range in the sequence was smaller, while P2 amplitudes were not affected much by the frequency range manipulation (Fig. 1C,D, Experiment 3), consistent with the fact that N1 amplitudes were more affected by long-term adaptation throughout the sequence than P2. To test this statistically, we ran another LME model including "range" as a predictor (Table 3). This analysis confirmed a significant effect of range on the N1 amplitude but not on the P2 amplitude and a significant difference between the effect of range on N1 and P2 amplitudes (Table 3).

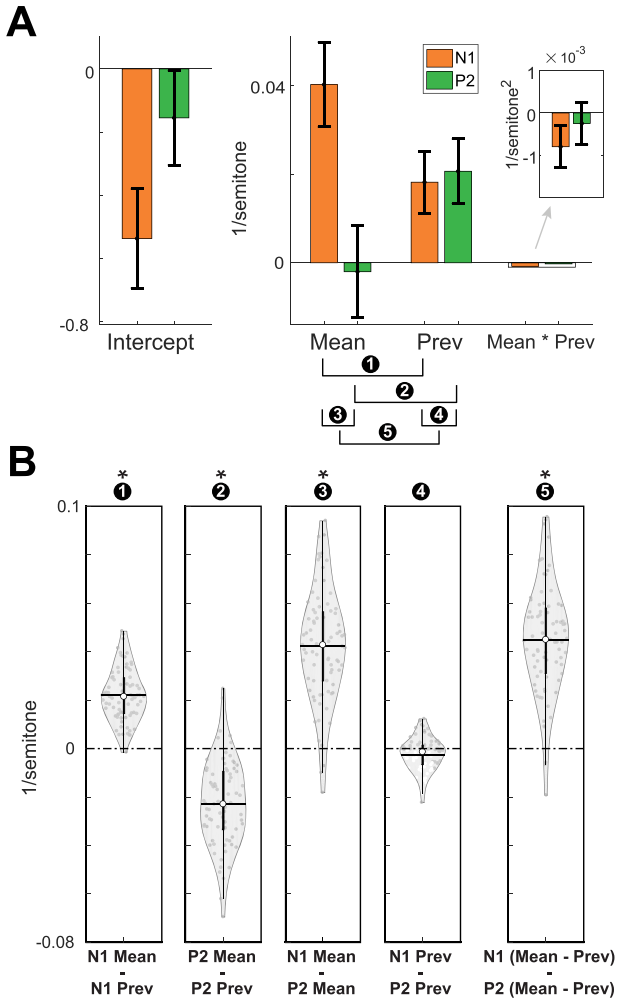

Figure 2. Long- and short-term context effects on N1 and P2 amplitudes. (*A*) Bar-graphs illustrate fixed-effect estimates values from a linear mixed-effects (LME) model (Table 2 and Methods for further specification). "Mean" and "Prev" stand for Interval-Mean and Interval-Previous, denoting the frequency intervals between the current tone and the sequence mean or previous tone, respectively, in semitones. The predicted N1 and P2 voltages were *z*-scored (after reversing the sign of the N1 data points). Error bars represent 95% CIs around the estimate (calculated by multiplying the SE of the estimate by the 95% inverse *t*-distribution value (DF = 78)). (*B*) Violin plots illustrate comparisons between LME estimates. Each dot represents one participant. White numbers in black circles above the violin plots indicate to which comparison they correspond (displayed under *A*). The significant contrasts are marked with an asterisk. Participant-specific estimates were calculated by adding the common fixed-effect estimates to participant-specific random effects. Horizontal lines represent the mean, white circles the medians, and thick and thin black vertical lines represent the 25% and 75% percentiles, respectively.

Sequence Frequency Range Attenuates Long- and Short-Term Context Effects

The latter LME model also indicated that both the short- and long-term context effects were attenuated for sequences with larger frequency ranges, consistent with adaptation with a limited bandwidth. There was a significant interaction between the range and the short-term context effect (Interval-Previous), such that for both N1 and P2, the effect of short term context was smaller the larger the range was (Table 3). The interaction between range and the long-term context effect (Interval-Mean) was significant for N1 but not for P2 (Table 3). Notably, the interaction terms interval_mean ∗ interval_previous both for N1 and P2 did not contribute significantly to this model, so we omitted them (Methods), but Supplementary Table S3 presents a comparison to a model including these terms, with similar results.

In summary, the ERP results demonstrated that N1 and P2 were affected differently by context: N1 was highly affected by long-term context (Interval-Mean) and P2 was not. Additionally, both were affected by short-term context (Interval-Previous) but P2 more robustly so. Furthermore, the spectral context (frequency range in the sequence) had a distinct effect on the N1 and P2 amplitudes: Smaller sequence range reduced N1 more than P2 amplitudes, suggesting that adaptation affects the N1 more than it affects the P2. These results imply that neural activity at the latencies of the N1 and P2 has distinct timescales of contextual influences. Additionally, the effects of long- and short-term context were generally reduced for larger frequency ranges, suggesting that some limited frequency bandwidth plays a role in these effects.

The specific values of the timescales and frequency bandwidth cannot be directly computed using the ERP analysis presented until here. Importantly, the 2 context predictors we used in the LME models, Interval-Mean and Interval-Previous made it possible to consider only very short- (1 previous tone) or very long-term (the sequence mean) contextual effects. In order to estimate the relevant temporal and spectral scales of these contextual effects, we employed a computational model to explain single trial variability.

Adaptation Model

We hypothesized that both the N1 and P2 results could be generated by a single underlying neural mechanism—adaptation of frequency-selective neural populations with relatively wide bandwidths that have 2 different time constants.

In the auditory system, frequency-selective neurons respond not only to their best frequency but also to nearby frequencies. Therefore, a tone presentation would adapt not only neural populations tuned exactly to that tone's frequency but also populations tuned to nearby frequencies. Further, if the interval to the next tone is short enough relative to the timescale of recovery from adaptation, this adaptation would not recover fully before the next stimulus occurs. Thus, if the effective frequency response profiles of neuronal populations are wider than the frequency intervals between tones in a stimulus sequence, cross-frequency adaptation (Taaseh et al. 2011; also termed co-adaptation, Herrmann et al. 2014, 2015) would render the adapting populations sensitive to frequency intervals. Moreover, the time it takes for neurons to recover from adaptation determines the duration of this effect. If recovery rates are slow relative to the inter-stimulus interval, neurons would accumulate adaptation due to their responses to more than one previous tone in the sequence.

With these premises, we used computational modeling (see Methods for equations) to test the feasibility of adaptation as the neural mechanism accounting for the ERP results presented above and to estimate quantitatively the effective time and frequency scales underlying the context-sensitivity of the N1 and P2 potentials. A similar modeling approach was applied in the past for neural responses in rats (Taaseh et al. 2011). Further, this model was applied for EEG by Herrmann et al. (2013, 2014, 2015), and we used a similar formulation to the latter studies

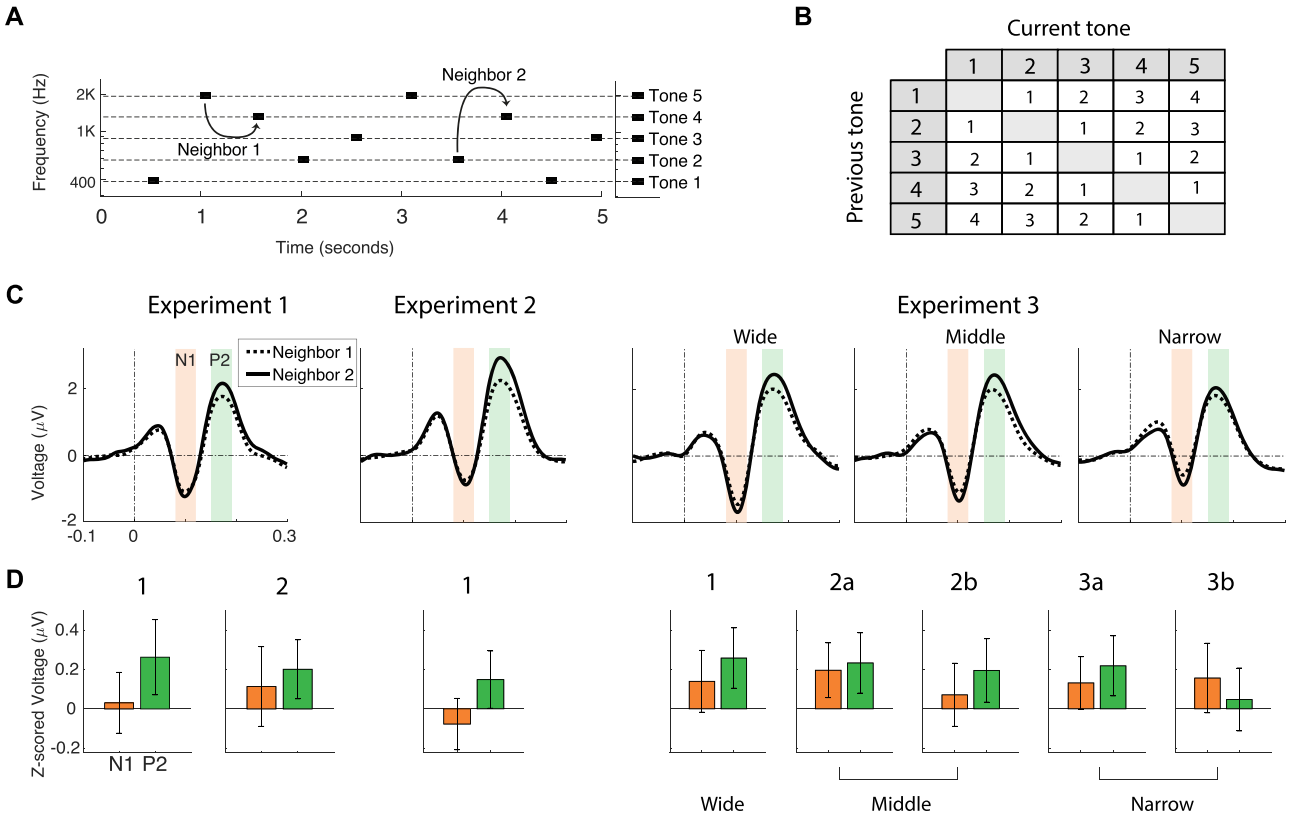

Figure 3. P2 is more sensitive than N1 to short-term context. (*A*) Example stimulus sequence denoting one frequency interval between 2 consecutive tones that are first-order neighbors on the sequence frequency axis ("Neighbor 1") and the second-order neighbors ("Neighbor 2"). (*B*) All possible combinations of current and previous tones in a sequence, and their grouping into the "degree of neighborhood." Column and row headers of the table denote ordinal tone numbers (1 to 5 from low to high frequencies, see right axis in *A*). Numbers inside the table denote the "degree of neighborhood." (*C*) ERPs of "current tones" when the previous tone was "Neighbor 1" and "Neighbor 2" for each experiment and condition separately. Shaded orange and green areas illustrate the time windows across which an extremum was defined as the N1 or P2 peak in individual subjects. (*D*) Bar graphs denote the mean and 95% CIs (across participants) of the difference between peak amplitudes in the "Neighbor 2" and "Neighbor 1" conditions. Peak amplitudes were *z*-scored for N1 and P2 separately (after reversing the sign of the N1 data points).

Table 3 Linear mixed-effects (LME) results including interactions with frequency range

		Estimate	SE	F(1,6188)	P-value	*d*
Predictors						
Intercept	N1	−0.76	0.1	55.3	1.2E−13	−1.4
	P2	−0.36	0.1	9.7	1.9E−03	−0.57
Interval-Mean	N1	0.054	0.01	17.3	3.2E−05	0.76
	P2	−0.028	0.01	4.2	4.1E−02	−0.37
Interval-Previous	N1	0.027	0.009	8.8	3.0E−03	0.54
	P2	0.043	0.009	21.2	4.1E−06	0.84
Range ∗ Intercept	N1	0.018	0.003	36.2	1.9E−09	1.1
	P2	0.0041	0.003	1.8	1.8E−01	0.24
Range ∗ Interval-Mean	N1	−0.0011	0.0004	7.2	7.4E−03	−0.49
	P2	0.0006	0.0004	2.5	1.1E−01	0.29
Range ∗ Interval-Previous	N1	−0.0006	0.0003	4.0	4.7E−02	−0.36
	P2	−0.0007	0.0003	6.2	1.2E-02	−0.46
Comparisons between pairs of predictors						
N1 versus P2 (Range ∗ Intercept)				10.9	9.5E−04	0.6
N1 versus P2 (Range ∗ Interval-Mean)				9.08	2.5E−03	−0.55
N1 versus P2 (Range ∗ Interval-Previous)				0.13	0.71	0.06

Note: Entries are similar to Table 2. Here only data from Experiment 3 (31 participants) were used to train the model, resulting in 6200 observations for N1 and P2 altogether (Methods).

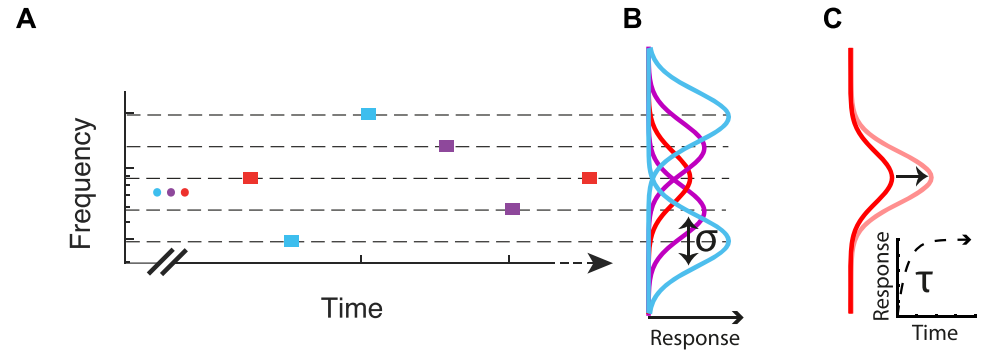

Figure 4. Illustration of the adaptation model. (*A*) An example segment of a tone sequence in the experiment serving as stimulus. Color code as in Figure 1. (*B*) Schematic Gaussian frequency RA curves of neural populations assumed by the model. All curves have the same bandwidth σ and each curve is centered around 1 of the 5 tone frequencies in the sequence. The color of the curves matches the color of the stimulus at its best frequency. The amplitudes of the curves represent the average adaptation of the neural populations throughout the sequence (excluding 5 initial tone presentations). Populations with RA curves centered around the middle frequency (red), respond most frequently throughout the sequence and therefore have the most adapted (attenuated) RA profile. (*C*) Exponential recovery from adaptation. The red curves represent the frequency RA curves of a population at 2 time points during a period with no tone presentations. Inset below is an exponential recovery function for a given time constant τ. The curve increases along the arrow connecting the dark to the light red curves, as specified by the exponential function in the inset.

for comparability. However, while Herrmann et al.'s studies fixed a priori the time constants for recovery from adaptation, we went a step further and used the model to estimate them from the data. We fitted model predictions to single-trial N1 and P2 amplitudes and estimated σ, the bandwidth of frequency RA profiles, and τ, the time constant of recovery from adaptation (Fig. 4), for N1 and P2 separately.

The Adaptation Model Accounts for both N1 and P2 Data

The model was fitted to single-trial N1 and P2 data separately, and the values of the time and frequency scale parameters, $\vec{\theta} = (\sigma, \tau)$, were estimated by selecting the $\vec{\theta}^{max} = (\sigma^{max}, \tau^{max})$ values maximizing the log-likelihood (LL^{max}), using a search over a predetermined grid of parameter values. To test the significance of the fit, we fitted the model to surrogate data consisting of random permutations of the measured responses across time (Methods). We did that for Experiments 1, 2, and 3 separately, as well as when using all data together. The adaptation model fitted the data better than chance in all cases. In almost all cases, for both N1 and P2, the LL^{max} values obtained using the actual data were much larger than all null LL^{max} values obtained from the surrogate data. The only exception was the case of N1 in Experiment 2, for which the LL^{max} value obtained using the actual data was smaller than the null LL^{max} in 1/100 cases. Since 100 repetitions were used to estimate the null distribution, $P < 0.01$ in all cases but N1 in Experiment 2 for which $P < 0.02$ (Supplementary Fig. S4).

N1 Has a Longer Adaptation Recovery Time than P2

The estimated time constant for recovery from adaptation, τ^{max}, was consistently longer for N1 relative to P2. This result was found when fitting the model using data from Experiments 1, 2, and 3 separately, as well as when using the data from all experiments together. The values of τ^{max}were 5, 4.6, 5, or 5 s for N1 and 0.4, 0.8, 1.6, or 0.8 s for P2 in Experiments 1, 2, 3, or when using all data together, respectively (Fig. 5A,B). In Experiments 1, 3, and when using all data together, τ^{max} of N1 was on the upper boundary of the allowed parameter range (5 s, equaling 2 repetitions of a 5-stimulus sequence). We limited the τ scale to 5 s since the predicted values become almost constant for larger τ values, due to the fact that the stimulus sequence was composed of successive permutations of the 5 frequencies. Therefore, a time constant of 5 s should be interpreted as 5 or longer.

The difference between τ^{max} of N1 and of P2 was significant in Experiments 1, 3, and when using all data together, while Experiment 2 showed the expected trend but did not reach statistical significance. We used 3 methods for comparing the time constants estimated for N1 versus for P2 (Methods): First, the values of τ^{max} of each potential type fell outside the 95% confidence region of τ^{max} of the other potential type (Fig. 5A,B, confidence regions based on the $\chi^2(2)$distribution for *D*, Methods, Supplementary Fig. S5) in all cases except for Experiment 2 for which the value of τ^{max} of P2 fell close to the lower edge of the confidence region of τ^{max} of N1 and vice versa. Second, the distribution of differences between τ^{max} of N1 and τ^{max} of P2 in 100 bootstrap repetitions we performed (sampling with replacement over the participants and repeating the parameter estimation procedure) had a mean significantly larger than 0 in all experiments (Supplementary Fig. S6). Third, the difference between the log-likelihoods calculated at τ^{max} and at τ^{max} of the other potential type was significantly larger than expected from the null distribution, estimated by fitting the model to permuted data, in Experiments 1, 3 and using all data together, although in Experiment 2 this effect did not reach significance ($P < 0.05$ in all cases but Experiment 2, Supplementary Fig. S7). Note that Experiment 2 contained less data than the other experiments (see Methods) and was therefore likely underpowered, explaining the lack of significance found in some of these statistical tests for Experiment 2.

N1 and P2 Have Similar Frequency Bandwidths of Adaptation

In contrast to the time constant, the estimated frequency bandwidth, σ^{max}, was largely similar for N1 and for P2. σ^{max} was 10, 7, 9, or 9 semitones for N1 and 4, 9, 9, or 7 semitones for P2 in Experiments 1, 2, 3, or all together, respectively (Fig. 5A). The only exception was for Experiment 1, in which the value of σ^{max}

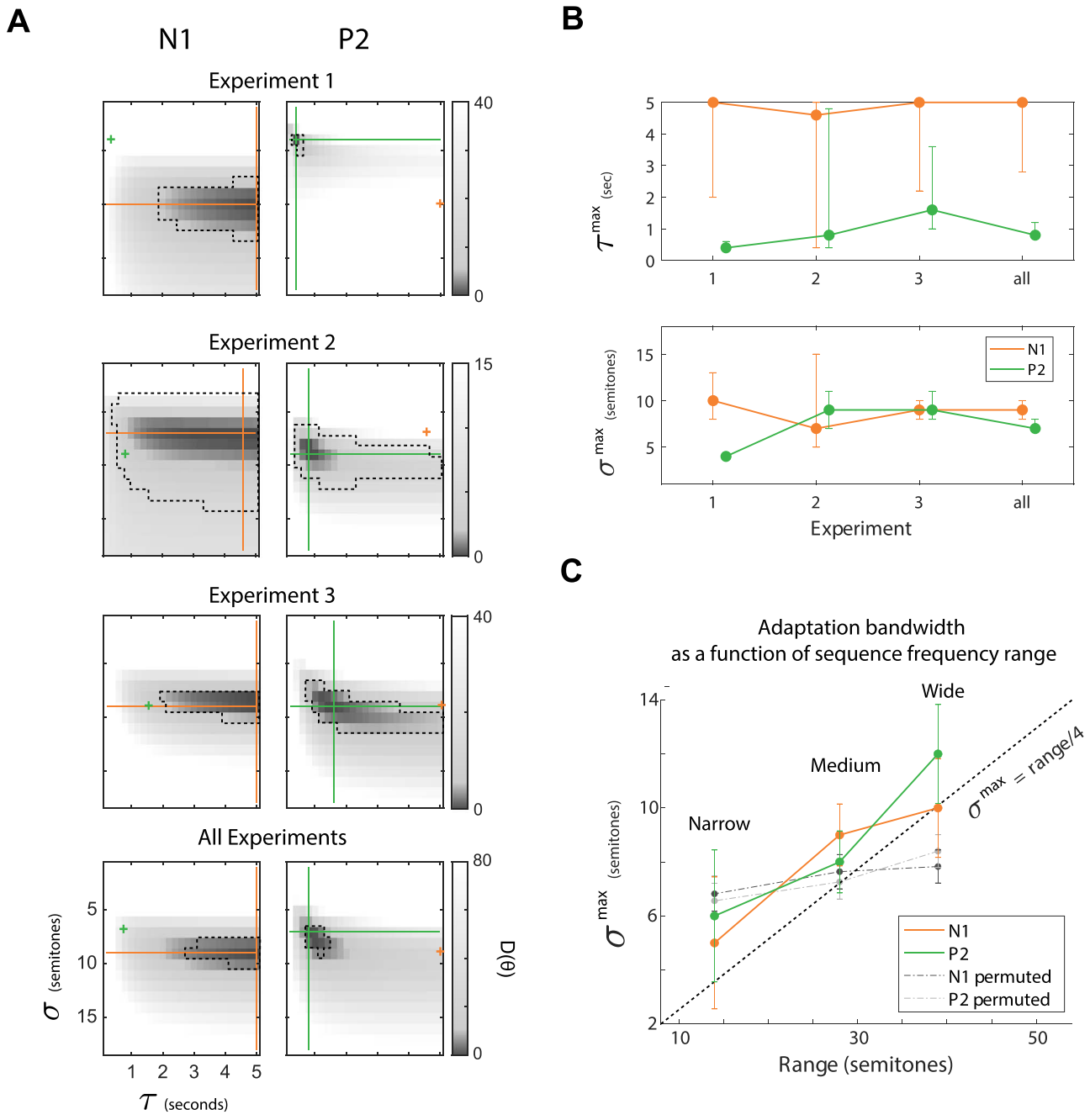

Figure 5. Adaptation model reveals the time and frequency scales of N1 and P2 context sensitivity. (*A*) *D* values (−2* log-likelihood ratio relative to $\vec{\theta}^{max}$ which maximizes the likelihood, Methods), for each possible value of $\vec{\theta} = (\tau, \sigma)$, for each experiment, for N1 (left) and P2 (right). Cross hairs (N1: orange, P2: green) are located at the maximum-likelihood estimated parameter values (i.e., at $\vec{\theta}^{max}$). Small cross-signs are located at $\vec{\theta}^{max}$ of the other potential type (exactly the crossing point of the neighboring plot, same color code), for visual comparison between $\vec{\theta}^{max}$ of the 2 potential types. Dashed lines surround 95% confidence regions for $\vec{\theta}^{max}$ calculated according to the null distribution of *D* (χ^2 with 2 DF, Methods). (*B*) All values of θ_{max} estimated across the different experiments presented together to facilitate visual assessment of the stability of results. The dots represent the values of the estimated parameters (identical to the cross-hairs in A). Error bars represent 95% confidence intervals (identical to horizontal or vertical cuts of the dashed confidence regions in A). (*C*) Frequency bandwidth depends on spectral context. σ^{max} as a function of stimulus frequency range for N1 (orange), P2 (green) and for N1 and P2 after permuting the order of trials within each range condition 100 times (dashed, see legend). Data from Experiment 3 only. Error bars of the N1 and P2 plots are 95% CIs estimated from 100 bootstrap repetitions sampling with replacement the 31 participants and repeating parameter estimation procedure. Error bars for the permuted data plots are 95% CIs calculated from the 100 permutations.

Table 4 LME results—effect of frequency range on adaptation bandwidth

Predictors		Estimate	SE	F(1,602)	P-value	*d*
Intercept	N1	2.7	0.37	52	1.7E−12	0.72
	P2	5.7	0.39	218	2.7E−42	1.5
Range	N1	0.22	0.013	283	2E−52	1.7
	P2	0.13	0.013	95	5.8E−21	0.97

Note: The data consisted of 101 bootstrap estimates of adaptation bandwidth (sampling with replacement from the 31 participants of Experiment 3 and repeating parameter estimation procedure) for each range condition and potential type (606 data points). The model consisted of fixed and random factors (grouped by bootstrap number) for each of the listed predictors (Methods). Column headers are similar to Tables 2 and 3.

for P2 was smaller than for N1. This was also reflected in the confidence regions we calculated (Fig. 5B) and the bootstrap test (Supplementary Fig. S6).

Adaptation Bandwidth Rescales to the Sequence Frequency Range

Next, we asked whether the parameters of the model (σ^{max} and τ^{max}) are constant when the spectral context changes. We fitted the model and estimated parameter values separately for each of the frequency range conditions in Experiment 3 (see Methods). We found that σ^{max} scaled with the stimulus frequency range such that it was larger for wider stimulus ranges both for N1 and P2: σ^{max} was 10, 9, and 5 semitones for N1 and 12, 8, and 6 semitones for P2, for the Wide, Medium, and Narrow conditions, respectively (Fig. 5C, Supplementary Fig. S8). Interestingly, the values of σ^{max} were consistently close to range/4 (Fig. 5C, diagonal dashed line), which is the mean interval between neighboring tones on the sequence-specific frequency axis (since there are 5 possible tone frequencies and 4 frequency intervals between them).

Notably, τ^{max}, the time constants estimated for the different frequency range conditions were: 5, 5, and 5 s for N1 and 4, 0.6, and 0.8 s for P2, for the Wide, Medium, and Narrow conditions, respectively (Supplementary Fig. S8). These values are in line with our result of longer timescales for N1: The timescale of N1 was larger than the timescale of P2 for all range conditions and the values of the timescales were consistent with the values we estimated in the other experiments. The only exception was the longer timescale of 4 s for P2 estimated for the Wide range condition. However, the Wide condition had half as many points as the other conditions (Methods and Fig. 1B) and therefore this estimate is less reliable.

To test the robustness of the modulation of σ by range, we performed 100 bootstrap repetitions (sampling with replacement from the 31 participants) and analyzed the resulting σ^{max} using an LME model (fixed factors: Intercept and range, random factors: Intercept and range grouped by bootstrap number, Methods). Table 4 shows that the dependence of the bandwidth on the range was highly significant with large effect sizes, for both N1 and P2. Thus, the bandwidths estimated separately for each range were highly stable both for N1 and P2.

The dependence of σ^{max} on the stimulus frequency range could potentially be caused merely by an overall reduction of peak amplitudes in sequences with smaller ranges, without any relation to the specific order of the tones in the sequence. To test if this is the case, we permuted the order of data trials within each range condition separately and repeated the parameter estimation procedure 100 times. The values of σ^{max} obtained using the permuted data showed a slight modulation by frequency range (Fig. 5C, pale dashed lines), but it was not significant (slope of σ^{max} as a function of frequency range estimated for permuted data, N1: 0.04, 95% CI = (−0.03 0.11), not different from 0: $t(298) = 0.99$, $P = 0.32$; P2: 0.05, 95% CI = (−0.02, 0.12), $t(298) = 1.37$, $P = 0.17$). Furthermore, the slopes of σ^{max} as a function of frequency range were significantly larger when using the real and bootstrapped data compared to using the permuted data (Fig. 5C) both for N1 (0.21, 95% CI = (0.19 0.24); note that this CI did not overlap with the CI of the slope for permuted data presented above) and for P2 (0.13, 95% CI = (0.1 0.15)). Thus, the overall reduction of amplitudes in sequences with smaller ranges was not sufficient to explain the spectral context effect.

Discussion

Our results suggest that neural sensitivity to past stimulation, integrated over 2 distinct timescales, manifested itself at 2 separate latencies following stimulus onset. Specifically, we show that a longer integration timescale (~5 s) is evident in the ERPs early after stimulus onset, at a latency of about 100 ms (N1), and is reflected in apparent sensitivity to the mean tone frequency. A shorter integration timescale (~1 s) is evident at a longer latency after stimulus presentation, about 200 ms (P2) and is reflected in sensitivity to the immediately preceding stimulus, in line with recent MEG reports (Quiroga-Martinez et al. 2020; Andermann et al. 2021). Further, we show quantitatively that both of these responses can be accounted for by frequency-specific adaptation of neural populations having distinct effective time constants of recovery. Finally, by varying the frequency range of the stimulus sequence, we show that spectral context affects the frequency adaptation bandwidth of these neural populations, independent of integration times.

Context Integration across Several Timescales

The N1 and P2 auditory-evoked potentials (AEPs) were among the first recorded human EEG responses (Davis 1939). Their neural generators and the computations underlying these responses are still not fully understood (e.g., Picton 2011; Lanting et al. 2013), although their generators are thought to reside mainly in the posterior superior temporal plane, especially when evoked using a relatively short SOA of about 0.5 s as in our case (Knight et al. 1980; Hari et al. 1982). Our results contribute to characterizing the functional distinction between the responses at the N1 and P2 latencies and thus support the claim that they reflect distinct neural computations (Knight et al. 1980; Hari et al. 1982; Lanting et al. 2013). Each of these AEPs probably does not reflect a single neural generator, but rather sums up several processes occurring at the time of measurement. For example,

the N1 was suggested to reflect at least 3 distinct components, although these components are hard to disentangle (Näätänen and Picton 1987). Thus, our results may not characterize any of the individual generating processes for either N1 or P2, but rather an effective average of those processes that are active at a specific latency post-stimulus. We found that these processes have an effective longer integration timescale at the N1 latency and a shorter integration timescale at the P2 latency.

The electric scalp response at the N1 latency is well known for being strongly attenuated by stimulus repetition, an effect termed adaptation, habituation, or refractoriness (e.g., Crowley and Colrain 2004; Picton 2011). Previous studies have established that the timescale of N1 adaptation is longer than about 1 s (Zacharias et al. 2012; Okamoto and Kakigi 2014; Herrmann et al. 2016). N1 adaptation in a sequence of pure tones with varying frequencies was well accounted for by the frequency-specific adaptation model we used here (Herrmann et al. 2013, 2014). However, the adaptation time constant in the latter studies was preset to 1.8 s, based on Sams et al. (1993). We developed a rigorous methodology to quantitatively estimate the time constant from the data in such adaptation models and found the N1 adaptation time constant to be longer than previously assumed by more than a factor of 2, using very similar stimulation parameters as Herrmann et al. (2013, 2014).

In contradistinction to N1, the effect of stimulus repetition on the electric scalp response at the P2 latency is more controversial. P2 was sometimes suggested to be less affected by adaptation than N1 (Crowley and Colrain 2004; Herrmann et al. 2013, 2016) and sometimes more (Lanting et al. 2013). There is some suggestive but inconclusive evidence for a shorter adaptation recovery rate of P2 compared to N1 (Hari et al. 1982; Lanting et al. 2013). Importantly, previous studies failed to account for P2 responses using adaptation models (Herrmann et al. 2013, 2016), including exactly the same frequency-specific adaptation model we used here (Herrmann et al. 2013). We suggest that the failure to account for P2 adaptation could have occurred because Herrmann et al. (2013) used the same time constant for both N1 and P2 when fitting the model to the data (see above). Here, we show instead that P2 fits the frequency-specific adaptation model very well with a shorter time constant than previously assumed. In sum, our study sheds new light on the AEP adaptation debate by quantitatively showing that frequency-specific adaptation is present at both the N1 and P2 latencies, but with a shorter timescale of recovery at the P2 relative to the N1 latency.

The time constants we report may depend on specific details of the paradigm we used. For example, the sequences we used did not contain long-term information over timescales longer than about 5 s. For this reason, we only tested recovery time constants up to 5 s. In Experiments 1 and 3, the time constant estimated for N1 reached the maximal value of 5 s (Fig. 5A). This may indicate that the timescale of N1 sensitivity may be even longer than what we report here.

Indeed, other human electrophysiology studies used sequences that contained longer-term structure and reported longer timescales of integration for N1, for example, ~10 s (Herrmann et al. 2016). Furthermore, Costa-Faidella et al. (2011) found that in pure tone sequences, P2 and the MMN were sensitive both to regularities established across short (<1 s) and long (~10 s) timescales. It could be that sequences embedding longer-term structure can reveal longer timescale sensitivity of P2 as well. However, the long-term structure in Costa-Faidella et al. (2011) was the violation of a globally established regularity. Thus, the reported long-term sensitivity of P2 in the latter study may reflect deviance-detection and prediction mechanisms, whereas the effects we report here may be due to adaptation mechanisms that are not specifically related to regularity extraction. The equiprobable structure of the tone sequences as well as the unattended paradigm we used suggests that both response timescales we characterized reflect automatic processes that do not directly depend on attention mechanisms.

Importantly, future studies of context-dependent auditory processing should consider the complex effects that can be caused by adaptation mechanisms prior to assuming deviance-detection or other predictive processes. One way to do so is to check whether such results can be explained by a flexible adaptation model as we presented here. More complex mechanisms may be implicated only if they improve the data fit beyond such adaptation models (e.g., Taaseh et al. 2011; Garrido et al. 2013).

Early Processing of Long and Later Processing of Short Timescales

Interestingly, we found that the earlier responses, at the N1 latency, had a longer time constant compared to the later P2 latency. Similarly, early auditory processing of long-scale sequence properties and later processing of short-scale properties were recently reported in a MEG study (Maheu et al. 2019). These results are surprising because longer timescales of integration are frequently associated with a higher-level of processing, which is expected to take place higher in the processing hierarchy. For example, some studies employing "global–local" paradigms found that early sensory responses were sensitive to short-term, local regularities whereas later activity was sensitive to long-term regularities (Bekinschtein et al. 2009; Recasens et al. 2014; Zhang et al. 2018). However, the distinction between short- and long-term regularities in the latter studies, established in the context of deviance detection paradigms, is likely to reflect neural mechanisms that may be distinct from the ones underlying our results, obtained in a passive paradigm with no regularities or deviance.

One can interpret our results in terms of resolution, associating a longer scale with coarser temporal resolution and a shorter scale with a higher resolution for processing finer details. This view corresponds with theories such as the frame-and-fill model developed for vision (Bar 2006; Snyder et al. 2012) or reverse hierarchy theory initially also developed for vision (Ahissar and Hochstein 2004) but later extended to audition (Nahum et al. 2008). By these theories, we quickly obtain a general gist of the sensory scene at a coarse resolution. Later on, the perceptual system may gradually resolve the details of the sensory scene, expressed in the shorter temporal scales of processing at later stages, consistent with our results.

Sensitivity to the Mean

Temporal averaging plays an important role in perception (Hollingworth 1910; Haberman et al. 2009; Albrecht and Scholl 2010; McDermott and Simoncelli 2011; McDermott et al. 2013). Our results manifest modulation of neural responses due to the mean of previous stimuli: N1 responses were most adapted for stimuli closest to the mean, consistent with previous reports (Ulanovsky et al. 2004; Herrmann et al. 2013, 2014). Although we did not measure perception in this study, the neural machinery described here may be related to the neural computation of mean and its influence on perception. For instance, judgments in a frequency discrimination task were biased toward the

mean of past stimuli (Raviv et al. 2012; Lieder et al. 2019). This perceptual bias to the mean was correlated with ERP adaptation (Jaffe-Dax et al. 2017), potentially linking it with the neural effects we study here. We suggest that the mean is represented by the adaptation level of a frequency-specific, yet wideband, neuronal population that adapts momentarily to stimuli as they come, then recover with a long time constant. The frequency corresponding to the most adapted population is the mean frequency of the sequence. This results in a time-dependent estimate of the mean frequency, which effectively averages past stimuli with a sliding window whose duration depends on the time constant of recovery. Our results demonstrate that information about the mean can co-exist in the auditory cortex with more local information about fast changes in stimulus properties.

Frequency Bandwidth Rescales to Spectral Range

We found that adaptation bandwidths rescale to the range of frequency distributions in the sequence. Our results are consistent with previous reports for N1 (Herrmann et al. 2013, 2014, 2015) and generalize them for P2 as well. This suggests that sensitivity of adaptation bandwidths to the spectral context is a general feature of auditory cortex, independent of the integration timescales. Somewhat similar effects, although in the context of deviance detection, have been shown by Garrido et al. (2013).

What could be the mechanism of this sensitivity to the spectral context? Dynamic adjustment of neural input–output functions due to changes in statistical properties of preceding stimulation has been directly shown using neural measurements in nonhuman animals for various perceptual dimensions such as light intensity (Dunn and Rieke 2006), visual motion (Brenner et al. 2000), and whisker motion in the rat barrel cortex (Maravall et al. 2007). In the auditory modality, rescaling of neuronal tuning to varying statistical properties of sounds was reported in the bird midbrain (Nagel and Doupe 2006), mammalian inferior colliculus (Kvale and Schreiner 2004; Dean et al. 2008; Dahmen et al. 2010), and primary auditory cortex (Blake and Merzenich 2002; Gourévitch et al. 2009; Rabinowitz et al. 2011). Thus, the change of neural response patterns due to stimulation statistics could reflect activity from the same neural population that undergoes rapid changes in network connectivity (Arnsten et al. 2010; Rabinowitz et al. 2011). Alternatively, the changes of neuronal response characteristics we measured using different frequency ranges may also reflect recruitment of distinct neuronal populations, each with its own fixed, distinct spectral and temporal response properties (Lee et al. 2016; Osman et al. 2018).

In conclusion, we find distinct responses in high-order auditory cortex, which show shorter and longer timescales of adaptation and which quickly and automatically adjust to the spectral context. Adaptation with distinct timescales can lead to heterogenous response characteristics enabling a complex analysis of incoming stimuli. We show that the temporal effects may be due to different rates of recovery from adaptation. The mechanisms underlying the spectral effects remain to be determined.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

Authors' Contributions

T.I.R. formulated the study, conducted experiments, developed and conducted analysis methods, and wrote the first draft of the paper. G.M. conducted experiments and performed data preprocessing. I.N. and L.Y.D. jointly supervised the study and participated in the design of the experiments, formulation, and application of analysis methods and writing the paper.

Data and Code Availability

Data and code are available via Open Science Framework (OSF): http://osf.io/mswhv.

Funding

T.I.R. was supported by the Hoffman Leadership and Responsibility Program at the Hebrew University of Jerusalem. L.Y.D. was supported by the Israel Science Foundation and Jack H. Skirball research fund. I.N. was supported by a grant from the Israel Academy of Sciences (390/13) and by AdERC grant 340063 (project RATLAND) from the European Research Council and is the Milton and Brindell Gottlieb Chair in Brain Science. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Notes

We thank research assistants who aided with EEG data acquisition and preprocessing: Michal Rabinovits, Eden Krispin, Anael Benistri. TIR would like to thank the members of Leon Deouell's lab through the years 2013–2021 for insightful discussions, support and facilitation of data collection and analysis while using the lab's computers and equipment. *Conflict of Interest*: The authors declare no competing interests. L.Y.D. declares he is a co-founder and advisor of InnerEye Ltd.

References

Ahissar M, Hochstein S. 2004. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci*. 8:457–464.

Albrecht AR, Scholl BJ. 2010. Perceptually averaging in a continuous visual world. *Psychol Sci*. 21:560–567.

Albrecht AR, Scholl BJ, Chun MM. 2012. Perceptual averaging by eye and ear: computing summary statistics from multimodal stimuli. *Atten Percept Psychophys*. 74:810–815.

Andermann M, Günther M, Patterson RD, Rupp A. 2021. Early cortical processing of pitch height and the role of adaptation and musicality. *Neuroimage*. 225:117501.

Arnsten AFT, Paspalas CD, Gamo NJ, Yang Y, Wang M. 2010. Dynamic network connectivity: a new form of neuroplasticity. *Trends Cogn Sci*. 14:365–375.

Bar M. 2006. Top down facilitation of visual recognition. *Proc Natl Acad Sci U S A*. 103:449–454.

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L. 2009. Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci U S A*. 106:1672–1677.

Blake DT, Merzenich MM. 2002. Changes of AI receptive fields with sound density. *J Neurophysiol*. 88:3409–3420.

Brainard, DH. 1997. The psychophysics toolbox. *Spatial vision*. 10(4):433–436.

Brenner N, Bialek W, De Ruyter Van Steveninck R. 2000. Adaptive rescaling maximizes information transmission. *Neuron*. 26:695–702.

Costa-Faidella J, Grimm S, Slabu L, Díaz-Santaella F, Escera C. 2011. Multiple time scales of adaptation in the auditory system as revealed by human evoked potentials. *Psychophysiology*. 48:774–783.

Crowley KE, Colrain IM. 2004. A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin Neurophysiol*. 115:732–744.

Dahmen JC, Keating P, Nodal FR, Schulz AL, King AJ. 2010. Adaptation to stimulus statistics in the perception and neural representation of auditory space. *Neuron*. 66:937–948.

Davis PA. 1939. Effects of acoustic stimuli on the waking human brain. *J Neurophysiol*. 2:494–499.

Dean I, Robinson BL, Harper NS, McAlpine D. 2008. Rapid neural adaptation to sound level statistics. *J Neurosci*. 28:6430–6438.

Dunn FA, Rieke F. 2006. The impact of photoreceptor noise on retinal gain controls. *Curr Opin Neurobiol*. 16:363–370.

Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR. 2001. Efficiency and ambiguity in an adaptive neural code. *Nature*. 412:787–792.

Garrido MI, Sahani M, Dolan RJ. 2013. Outlier responses reflect sensitivity to statistical structure in the human brain. *PLoS Comput Biol*. 9(3):e1002999.

Gourévitch B, Noreña A, Shaw G, Eggermont JJ. 2009. Spectrotemporal receptive fields in anesthetized cat primary auditory cortex are context dependent. *Cereb Cortex*. 19:1448–1461.

Haberman J, Harp T, Whitney D. 2009. Averaging facial expression over time. *J Vis*. 9:1–1.

Hari R, Kaila K, Katila T, Tuomisto T, Varpula T. 1982. Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: implications for their neural generation. *Electroencephalogr Clin Neurophysiol*. 54: 561–569.

Herrmann B, Augereau T, Johnsrude IS. 2020. Neural responses and perceptual sensitivity to sound depend on sound-level statistics. *Sci Rep*. 10:1–12.

Herrmann B, Henry MJ, Fromboluti EK, McAuley JD, Obleser J. 2015. Statistical context shapes stimulus-specific adaptation in human auditory cortex. *J Neurophysiol*. 113: 2582–2591.

Herrmann B, Henry MJ, Johnsrude IS, Obleser J. 2016. Altered temporal dynamics of neural adaptation in the aging human auditory cortex. *Neurobiol Aging*. 45:10–22.

Herrmann B, Henry MJ, Obleser J. 2013. Frequency-specific adaptation in human auditory cortex depends on the spectral variance in the acoustic stimulation. *J Neurophysiol*. 109:2086–2096.

Herrmann B, Maess B, Johnsrude IS. 2018. Aging affects adaptation to sound-level statistics in human auditory cortex. *J Neurosci*. 38:1989–1999.

Herrmann B, Schlichting N, Obleser J. 2014. Dynamic range adaptation to spectral stimulus statistics in human auditory cortex. *J Neurosci*. 34:327–331.

Hollingworth HL. 1910. The central tendency of judgment. *J Philos Psychol Sci Methods*. 7:461.

Jaffe-Dax S, Frenkel O, Ahissar M. 2017. Dyslexics' faster decay of implicit memory for sounds and words is manifested in their shorter neural adaptation. *Elife*. 6:e20557.

Jung T-P, Makeig S, Humphries C, Lee T-W, McKeown MJ, Iragui V, Sejnowski TJ. 2000. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*. 37:163–178.

Khouri L, Nelken I. 2015. Detecting the unexpected. *Curr Opin Neurobiol*. 35:142–147.

Knight RT, Hillyard SA, Woods DL, Neville HJ. 1980. The effects of frontal and temporal-parietal lesions on the auditory evoked potential in man. *Electroencephalogr Clin Neurophysiol*. 50:112–124.

Koelsch S. 2009. Music-syntactic processing and auditory memory: similarities and differences between ERAN and MMN. *Psychophysiology*. 46:179–190.

Kvale MN, Schreiner CE. 2004. Short-term adaptation of auditory receptive fields to dynamic stimuli. *J Neurophysiol*. 91:604–612.

Lanting CP, Briley PM, Sumner CJ, Krumbholz K. 2013. Mechanisms of adaptation in human auditory cortex. *J Neurophysiol*. 110:973–983.

Lee CM, Osman AF, Volgushev M, Escabí MA, Read HL. 2016. Neural spike-timing patterns vary with sound shape and periodicity in three auditory cortical fields. *J Neurophysiol*. 115:1886–1904.

Lieder I, Adam V, Frenkel O, Jaffe-Dax S, Sahani M, Ahissar M. 2019. Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nat Neurosci*. 22:256–264.

Maheu M, Dehaene S, Meyniel F. 2019. Brain signatures of a multiscale process of sequence learning in humans. *Elife*. 8:e41541.

Maravall M, Petersen RS, Fairhall AL, Arabzadeh E, Diamond ME. 2007. Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. *PLoS Biol*. 5:0323–0334.

McDermott JH, Schemitsch M, Simoncelli EP. 2013. Summary statistics in auditory perception. *Nat Neurosci*. 16:493–U169.

McDermott JH, Simoncelli EP. 2011. Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron*. 71:926–940.

Näätänen R, Picton T. 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*. 24:375–425.

Nagel KI, Doupe AJ. 2006. Temporal processing and adaptation in the songbird auditory forebrain. *Neuron*. 51:845–859.

Nahum M, Nelken I, Ahissar M. 2008. Low-level information and high-level perception: the case of speech in noise. *PLoS Biol*. 6:e126.

Okamoto H, Kakigi R. 2014. History of silence affects auditory evoked fields regardless of intervening sounds: a magnetoencephalographic study. *Eur J Neurosci*. 40:3380–3386.

Osman AF, Lee CM, Escabí MA, Read HL. 2018. A hierarchy of time scales for discriminating and classifying the temporal shape of sound in three auditory cortical fields. *J Neurosci*. 38:6967–6982.

Piazza EA, Sweeny TD, Wessel D, Silver MA, Whitney D. 2013. Humans use summary statistics to perceive auditory sequences. *Psychol Sci*. 24:1389–1397.

Picton TW. 2011. *Human auditory evoked potentials*. San Diego, CA: Plural Publishing.

Poulin-Charronnat B, Bigand E, Koelsch S. 2006. Processing of musical syntax tonic versus subdominant: an event-related potential study. *J Cogn Neurosci*. 18:1545–1554.

Quiroga-Martinez DR, Hansen NC, Højlund A, Pearce M, Brattico E, Vuust P. 2020. Decomposing neural responses to melodic surprise in musicians and non-musicians: evidence for a hierarchy of predictions in the auditory system. *Neuroimage*. 215:116816.

Rabinowitz NC, Willmore BDB, Schnupp JWH, King AJ. 2011. Contrast gain control in auditory cortex. *Neuron*. 70: 1178–1191.

Raviv O, Ahissar M, Loewenstein Y. 2012. How recent history affects perception: the normative approach and its heuristic approximation. *PLoS Comput Biol*. 8:e1002731.

Recasens M, Grimm S, Wollbrink A, Pantev C, Escera C. 2014. Encoding of nested levels of acoustic regularity in hierarchically organized areas of the human auditory cortex. *Hum Brain Mapp*. 35:5701–5716.

Regev TI, Nelken I, Deouell LY. 2019. Evidence for linear but not helical automatic representation of pitch in the human auditory system. *J Cogn Neurosci*. 31:669–685.

Sams M, Hari R, Rif J, Knuutila J. 1993. The human auditory sensory memory trace persists about 10 sec: neuromagnetic evidence. *J Cogn Neurosci*. 5:363–370.

Snyder AC, Shpaner M, Molholm S, Foxe JJ. 2012. Visual object processing as a function of stimulus energy, retinal eccentricity and Gestalt configuration: a high-density electrical mapping study. *Neuroscience*. 221:1–11.

Sussman E. 2007. A new view on the MMN and attention debate: the role of context in processing auditory events. *J Psychophysiol*. 21:164–175.

Taaseh N, Yaron A, Nelken I. 2011. Stimulus-specific adaptation and deviance detection in the rat auditory cortex. *PLoS One*. 6:e23369.

Teng X, Poeppel D. 2020. Theta and gamma bands encode acoustic dynamics over wide-ranging timescales. *Cereb Cortex*. 30:2600–2614.

Teng X, Tian X, Rowland J, Poeppel D. 2017. Concurrent temporal channels for auditory processing: oscillatory neural entrainment reveals segregation of function at different scales. *PLoS Biol*. 15:1–29.

Tremblay K, Kraus N, McGee T, Ponton C, Otis B. 2001. Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear*. 22:79–90.

Ulanovsky N, Las L, Farkas D, Nelken I. 2004. Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci*. 24:10440–10453.

Wilkinson GN, Rogers CE. 1973. Symbolic description of factorial models for analysis of variance. *Appl Stat*. 22:392.

Zacharias N, König R, Heil P. 2012. Stimulation-history effects on the M100 revealed by its differential dependence on the stimulus onset interval. *Psychophysiology*. 49: 909–919.

Zhang J, Zhou X, Chang R, Yang Y. 2018. Effects of global and local contexts on chord processing: an ERP study. *Neuropsychologia*. 109:149–154.