Predictive encoding of pure tones and FM-sweeps in the human auditory cortex

Jasmin Stein1*, Katharina von Kriegstein2, Alejandro Tabas2

1Faculty of Psychology, Technical University Dresden, Dresden 01062, Germany,
2Chair of Cognitive and Clinical Neuroscience, Faculty of Psychology, Technical University Dresden, Bamberger Str. 7, Dresden 01187, Germany
*Corresponding author: Email: jasmin.stein@tu-dresden.de

Expectations substantially influence perception, but the neural mechanisms underlying this influence are not fully understood. A prominent view is that sensory neurons encode prediction error with respect to expectations on upcoming sensory input. Although the encoding of prediction error has been previously demonstrated in the human auditory cortex (AC), previous studies often induced expectations using stimulus repetition, potentially confounding prediction error with neural habituation. These studies also measured AC as a single population, failing to consider possible predictive specializations of different AC fields. Moreover, the few studies that considered prediction error to stimuli other than pure tones yielded conflicting results. Here, we used functional magnetic resonance imaging (fMRI) to systematically investigate prediction error to subjective expectations in auditory cortical fields Te1.0, Te1.1, Te1.2, and Te3, and two types of stimuli: pure tones and frequency modulated (FM) sweeps. Our results show that prediction error is elicited with respect to the participants’ expectations independently of stimulus repetition and similarly expressed across auditory fields. Moreover, despite the radically different strategies underlying the decoding of pure tones and FM-sweeps, both stimulus modalities were encoded as prediction error in most fields of AC. Altogether, our results provide unequivocal evidence that predictive coding is the general encoding mechanism in AC.

Key words: predictive coding; auditory cortex; FM-sweeps; fMRI.

Introduction

Subjective expectations influence our perception of the world (De Lange et al. 2018). They facilitate perceiving noisy (Chaik et al. 2010; Leonard et al. 2016) or ambiguous (Chambers et al. 2017; Sterzer et al. 2008) sensory input, and bias perception when we hold strong expectations about the sensory world (Lange 2009). Understanding the mechanisms integrating our expectations with the sensory input is an essential prerequisite to understand perception. The predictive coding framework is a theory of sensory processing aiming to describe these mechanisms. Its main tenet is that sensory neurons encode prediction error with respect to an internal generative model of the sensory world (Friston 2003; Mumford 1992; Rao and Ballard 1999; Spratling 2017).

Neurons in the rodent auditory cortex (AC) encode pure tones as prediction error (Nieto-Diego and Malmierca 2016; Parras et al. 2017; Pérez-González et al. 2021; Rubin et al. 2016). Prediction error is typically elicited using oddball paradigms, where predictable repetitions of a standard sound are rarely interrupted by a deviant. Individual neurons in the AC show reduced responses to repeated standards and recovered responses to deviants, a phenomenon that is called stimulus-specific adaptation (SSA; Ulansky et al. 2003). SSA is typically interpreted as prediction error (Malmierca 2015). However, it can also be explained in terms of neural habituation (Eytan et al. 2003; Mill et al. 2011; Wang et al. 2014; see Tabas et al. 2021a for review, and Carbajal and Malmierca 2018 for a different perspective). One way to disassociate whether SSA reflects habituation or prediction error is to manipulate participants’ subjective stimulus expectations on the sensory input orthogonally to stimulus repetition (Tabas et al. 2021a; Tabas et al. 2020).

Using abstract rules to induce subjective expectations, we have recently shown that SSA in neural populations of the subcortical auditory pathway reflects the encoding of prediction error (Tabas et al. 2021b; Tabas et al. 2020). Other studies have demonstrated similar effects in cortical areas measuring the mismatch negativity (MNN), a component of the event-related potentials assumed to encode prediction error (Cornell et al. 2012; Costa-Faidella et al. 2011; Dürschmid et al. 2016; Lecaignard et al. 2015; Phillips et al. 2016; Todorovic and De Lange 2012; Todorovic et al. 2011). However, since the MNN is at least partially generated in the frontal cortex (Deouell 2007; Shalgi and Deouell 2007), it is challenging to measure unambiguously what portion of the event-related potential stems from AC. Two studies using functional magnetic resonance imaging (fMRI) specifically investigated the encoding of prediction error in AC (Berlot et al. 2018; Cacciaglia et al. 2019). However, stimulus repetition and expectation were confounded in these studies. Thus, although the extended literature on prediction error in human AC seems to indicate that cortical SSA is also attributable to prediction error, this has not yet been formally tested: it is unclear whether prediction error is encoded by the same neural populations that exhibit SSA.

Additionally, previous studies did not consider in detail whether SSA or prediction error is similarly represented in distinct fields of the AC. The auditory system consists of a primary (lemniscal) and secondary (nonlemniscal) subdivision (Lee and Sherman 2011). Primary areas show narrow tuning curves;
secondary areas are tuned more widely and support multisensory integration (Hu 2003). In rodents, SSA is stronger in secondary subdivisions of AC (Nieto-Diego and Malmierca 2016; Parras et al. 2017).

Furthermore, most previous research on the encoding of prediction error in AC has focused on static pure tones. However, natural auditory scenes often involve dynamic acoustic components; for instance, frequency modulation (FM) is a basic element of animal (Suga 2012) and human (Divenyi 2009) vocalization. In human speech, FM-sweeps constitute formant transitions, the main components of consonants preceding a vowel (Liberman et al. 1967), which are critical for phoneme identification.

Although we have previously investigated prediction error to FM in subcortical areas (Tabas et al. 2021b), whether FM is encoded as prediction error in the human AC is still unclear. Previous studies considered prediction error to FM in the human cerebral cortex and yielded mixed results: Some reported an MMN to deviating FM-stimuli, suggesting that FM is also encoded as prediction error (Cornelia et al. 2013; Hsieh and Yeh 2021; Kung et al. 2020); others reported enhanced neural responses to repeated FM-stimuli, concluding that different predictive coding strategies underlie the encoding of pure tones and FM (Altmann et al. 2011; Heinemann et al. 2011; Heinemann et al. 2010; Okamoto and Kagiki 2017). This different encoding strategy might arise from the fact that, while pure tones are first encoded in the basilar membrane (Malmierca and Hackett 2010), FM selectivity is present only in the auditory midbrain, thalamus, and cortex (Altmann and Gaese 2014; Geis and Borst 2013; Hall et al. 2000; Hart et al. 2003; Issa et al. 2017; Lui and Mendelson 2003; Paltoglou et al. 2011). Studying whether FM is encoded as prediction error specifically in AC might shed light on these divergences.

Here, we systematically studied the encoding of prediction error to pure tones and FM-sweeps across fields of the human AC with high-resolution fMRI. We used a variation of the auditory oddball paradigm (Tabas et al. 2020) designed to manipulate participants’ expectations independently of local stimulus statistics. Our aim was to elucidate whether predictive coding is the general encoding mechanism in the human AC.

Materials and methods
The studies were approved by the Ethics committee of the Medical Faculty of the University of Leipzig, Germany (pure tone experiment) and the Ethics committee of the Technische Universität Dresden, Germany (FM-sweep experiment). All participants provided written informed consent and received monetary compensation for their participation.

Data acquisition and experimental design
We analyzed fMRI data from two experiments designed to study prediction error encoding of pure tones (Tabas et al. 2020) and FM-sweeps (Tabas et al. 2021b). Both experiments used the same variation of the auditory oddball paradigm, one with pure ones, one with FM-sweeps. They were both acquired at different scanning sessions and using different participant cohorts.

Participants
All participants were neurotypical normal-hearing German native speakers (see Tabas et al. 2020 and Tabas et al. 2021b for further details on inclusion criteria). Nineteen participants (12 female) between the ages of 24 and 34 (average 26.6) participated in the pure tone study; 18 participants (12 female) between the age of 19 and 31 (average 24.6) participated in the FM-sweep study.

Stimuli
In the pure tone experiment, there were three pure tone stimuli of 50 ms duration (including 5 ms onset/offset ramps) and frequencies of 1455, 1500, and 1600 Hz. The tones were combined into six pairings of standard and deviant tones. Each of the resulting oddball sequences was consequently characterized by one of three possible absolute frequency differences characterizing the distance between standards and deviants ($\Delta f = |f_{dev} - f_{std}|$) either 145, 100, or 45 Hz. See Fig. 1A for a visualization of sound stimuli and sequences.

In the FM-sweep experiment, there were three sinusoidal FM-sweeps with a duration of 50 ms (5 ms onset/offset ramps) and starting frequencies of 1000, 1070, or 1280 Hz. The FM-sweeps ended at either 1080, 1170, or 1200 Hz, respectively. The tones were also combined into six pairings of standard and deviant tones. An FM-sweep could deviate from the standard in its FM direction, FM rate, or both. Since all sweeps had the same duration, the defining property of the FM-sweeps was their frequency span $\Delta f$ — the difference between starting and ending frequencies. Each of the sequences was consequently characterized by one of three possible absolute frequency span differences characterizing the distance between standards and deviants ($\Delta f = |\Delta f_{dev} - \Delta f_{std}|$). See Fig. 1B for an exemplary illustration of the FM-sweep stimuli and sequences (Tabas et al. 2021b).

Experimental design
The design was a variation of the oddball paradigm where abstract rules were used to manipulate participants’ expectations on the upcoming stimuli independent of the local statistical regularity of the presented sound sequences. Specifically, participants listened to sequences of eight sounds: seven repeated standards and one deviant that could occur in positions four, five, or six. The stimuli were separated by a 700 ms interstimulus interval.

Participants were told explicitly that all sequences contained a deviant and that the deviant could occur in positions four, five, or six. Participants were instructed to report the position of the deviant via button press as fast and as accurately as possible. Deviants were equally likely to be placed in each of the three positions. Thus, the probability of a deviant in position four after hearing three standards was 1/3. However, if the deviant was not in position four, since deviants occur once in each sequence, the probability of hearing a deviant in position five after hearing four standards was 1/2. If the deviant was neither in position four nor five, the probability of a deviant in position six after hearing five standards was 1 (Tabas et al. 2021b; Tabas et al. 2020).

The intertrial interval was calculated so that the separation between deviants followed a Gaussian distribution $N(\mu = 5 s, \sigma = 1 s)$ to maximize the statistical power of the estimation of the responses to the deviants (Friston et al. 1999). The ITI was further truncated so that the minimum separation between trials was 1500 ms (i.e. the duration of two tones and two ITIs) to guarantee that participants could tell apart any two consecutive trials. This resulted in ITIs between 1500 and 11000 ms. All sound combinations were used the same number of times with all deviant locations across runs. Using three standard-deviant combinations reduced the effect of long-term adaptation effects across trials and rendered the task more engaging and less repetitive for the participants. The specific ordering of the trials was pseudorandomized independently for each participant to prevent potential effects of presentation ordering in the group results.

The pure tone experiment comprised four runs that were completed by all participants. The FM-sweep data were collected in...
three sessions with three runs each; most participants completed nine runs, one participant only completed eight due to technical reasons. We acquired a higher number of runs for the FM-sweep experiment to partially compensate for the lower SNR of the data: while the pure-tone experiment was performed using 7-Tesla fMRI at a tSNR of about 50, the FM-sweep experiment was performed using 3-Tesla fMRI at a tSNR of about 30. Fully compensating for the tSNR drop would have required to measure a 2.7-fold data; we measured as many runs as we could within our limited resources.

In both experiments, a run contained 6 blocks of 10 trials. Deviant positions were pseudo-randomized so that they all occurred 20 times in each run. The runs lasted for around 10 min and were separated by a 1 min break. Practice trials were presented at the beginning of the first run to ensure task understanding. Interspersed null events were used to optimize the fit of the GLMs (Friston et al. 1999). Further details can be found in Tabas et al. 2020 (pure tones) and Tabas et al. 2021b (FM-sweeps).

**FMRI data acquisition**

FMRI data were collected using EPI sequences and partial FoVs. Magnetic field strength and image resolution differed between data sets. Data from the pure tone experiment were collected using a Siemens Magnetom 7-Tesla scanner (Siemens Healthineers, Erlangen, Germany) with an 8-channel head coil and a voxel size of 1.5 mm isotropic; data from the FM-sweep experiment were collected using a Siemens Trio 3-Tesla scanner (Siemens Healthineers, Erlangen, Germany) with a 32-channel head coil and a voxel size of 1.75 mm isotropic. Interleaved slice acquisition was used in both data sets.

Pure tone data were collected using the following scanning parameters: TR = 1600 ms, TE = 19 ms, flip angle 65°, GRAPPA with acceleration factor 2, 33% phase oversampling, matrix size 88 × 88, phase partial Fourier 6/8, FoV = 132 mm × 132 mm (30 slices). We also acquired three whole-head volumes with 80 slices to aid coregistration. FM-sweep data were collected using the following scanning parameters: TR = 1900 ms, TE = 42.2 ms, flip angle 66°, matrix size 88 × 88, FoV = 154 mm × 154 mm (24 slices). We also acquired one whole head volumes with 84 slices to aid coregistration process.

Structural images for the pure tone experiment were measured using an MP2RAGE T1 protocol (700 mm isotropic resolution, TE = 2.45 ms, TR = 5000 ms, TI1 = 900 ms, TI2 = 2750 ms, flip angle 1 = 5°, flip angle 2 = 3°, FoV = 224 mm × 224 mm, GRAPPA acceleration factor 2). Structural images for the FM-sweep data were measured using an MPRAGE T1 protocol (1 mm isotropic resolution, TE = 1.95 ms, TR = 1000 ms, flip angle 1 = 8°, FoV = 256 mm × 256 mm).

Physiological data (heart rate and respiration in the pure tone experiment, heart rate in the FM-sweep experiment) were collected and processed for use as regressors of no-interest during model estimation for both stimulus types.
Data preprocessing

Anatomical data

Data preprocessing and analysis was implemented using Nipype 1.1.2 (Gorgolewski et al. 2011) and included functions from the FMRIB Software Library, version 5 (FSL, Jenkinson et al. 2012); FreeSurfer, version 7 (Fischl et al. 2002); the Advanced Normalization Tools, version 2.2.0 (ANTs, Avants et al. 2011); and the Statistical Parametric Mapping toolbox (SPM, Penny et al. 2011), version 12.

All anatomical data were resampled to a resolution of 1 mm isotropic. We computed the boundaries between gray and white matter using FreeSurfer’s recon-all. These boundaries were later used for coregistration of the functional data to the participants’ structural images. In the case of the pure tone experiment, we first computed a brain mask excluding voxels containing air, cerebrospinal fluid, scalp, and skull. This was necessary because MP2RAGE (but not MPRAGE) yields noisy signals outside the brain that interfere with the automatic processes of recon-all. The mask was calculated using FreeSurfer’s BET and SPM’s Segment and was applied using FSLmath. Then, FreeSurfer’s recon-all was used to obtain gray and white matter boundaries, and ANTs was used to calculate the coregistration matrix between the anatomical data and the MNI152 symmetric template.

FMRI data

Functional data were slice time corrected using SPM’s SliceTiming using the first slice as reference volume. We used SPM’s FieldMap Toolbox to calculate distortions due to magnetic field heterogeneities. Then, motion and distortion correction was performed on the functional data separately for each session (SPM Realign and Unwarp). Nipype module’s rapidant was used to detect artifacts from the realigned functional data to serve as regressors of no-interest in our design matrix during GLM estimation. The resulting functional data were smoothed (SPM Smooth) using a 2 mm FWHM Gaussian kernel.

In the case of the pure tone data, the derivatives (i.e. log-evidences and beta maps) were registered to the anatomical space after fitting (see GLM Estimation and Bayesian Model Comparison). For FM-sweeps, the realigned functional data were registered to the anatomical space using FreeSurfer’s ApplyVolTransform before model estimation to ensure all data were available in the same space during model fitting.

The transformation matrix between functional and structural data was computed using FreeSurfer’s BBRegister using the white and gray matter boundaries computed as described above and the whole-brain EPI as an intermediate stage.

Anatomical regions of interest

Anatomical regions of interest (ROIs) were taken from Morosan et al. (Morosan et al. 2001). The ROIs of interest were the bilateral auditory cortical fields Te1.0, Te1.1, Te1.2, and Te3 (Morosan et al. 2001). Areas Te1.0, Te1.1, and Te1.2 are located on Heschl’s gyrus (Te1.1 most medial, Te1.2 most lateral). Te1.0 and Te1.1 have been proposed to correspond mostly to BA 41. Te1.0 is generally considered representing the primary auditory cortex in humans (Moerel et al. 2014). Te1.2 overlaps with BA 42 (Moerel et al. 2014). Comparing human and primate auditory fields, it was assumed that Te1.0 corresponds to the auditory core and Te1.1 and Te1.2 represent medial junction and lateral belt (Moerel et al. 2014). Thus, Te1.1 is often considered an intermediate processing stage between primary and secondary auditory areas (Hackett et al. 2001). However, the correspondence of human and primate auditory fields is still unclear, e.g., Besle et al. 2019; Gulban et al. 2020; Moerel et al. 2014. Te3 is a secondary auditory association area lying on the lateral surface of the superior temporal gyrus as part of BA 22 and might correspond to parabelt areas in primates (Moerel et al. 2014).

Data analysis

GLM estimation

First level analyses were performed using SPM’s EstimateModel. Statistical analysis at the participant- and group-level was conducted in MATLAB (The MathWorks Inc., Version 2020b) using custom code.

We estimated one GLM per participant. The model included six task regressors: std0 (the first standard in a sequence), std1 (standards before the deviant), std2 (standards after the deviant), dev4, dev5, and dev6 (deviants in positions four, five, and six). The first standard was modeled as a separate regressor to test for adaptation by comparing the estimates corresponding to std0 and std1/std2. Modelling std1 and std2 separately was necessary because we expected a slight recovery of neural responses to standards after the deviant (see Fig 1D); without an informed hypothesis on how strong this recovery would be, we chose to set it as a free parameter of the model. In addition, we added physiological data, artifact regressors, and realignment parameters to the design matrix as regressors of no interest.

The regressors corresponding to std1 and std2 were linearly parametrically modulated according to the position of each sound within the sequence of each trial: values corresponding to each std1 sound were assigned amplitudes increasing from one to the total number of std1, and std2 sounds were assigned increasing amplitudes from one to the total number of std2 in each trial. For example, in a sequence with a deviant in position four, the two std1 sounds were assigned the amplitudes amps1 = [1, 2] and the four std2 sounds were assigned the amplitudes amps2 = [1, 2, 3, 4]. The parametric modulation indexes the standards, and it is meant to allow SPM a greater flexibility fitting the regressors under the assumption that the responses might vary linearly with successive repetition of the standard. We used this technique to circumvent the problem that responses to individual standards cannot be accurately fitted due to their temporal proximity. Since SPM has no biases for negative or positive slopes or finite or zeroed intercepts, using the indices was the most parsimonious way to declare the linear fit. All amplitudes were z-standardized before model fitting.

Model estimations for the pure tone data were done using the smoothed data in the native space of the functional data of the individual participants. For FM-sweeps, we estimated the models in the space of the participants’ anatomical scans. After model estimation, the spatial transformations calculated before were applied to the resulting statistical maps. The statistical maps of the pure tone data were first registered to the participants’ anatomical scans using FreeSurfer’s ApplyVolTransform and subsequently to the MNI152 symmetric template using ANTs’ ApplyTransforms. For FM-sweeps, statistical maps were registered directly to MNI space.

The resulting beta estimates were z-standardized according to participant, experimental run, and ROI before the second level analyses to reduce variance specific to participants, runs, and ROIs.

Identifying voxels showing SSA

To localize voxels showing SSA, we first identified voxels within the anatomical ROIs showing adaptation (reduced responses to
repeated standards) and deviant detection (stronger response to deviants compared with standards) as defined by commonly used indices from animal literature (Parras et al. 2017): the repetition suppression and neuronal mismatch indices, respectively. Adapting voxels were identified using the repetition suppression index, in our study defined by the contrast std0 > 0.5 std1 + 0.5 std2. This assumes that the responses to the first standard, std0, were unaffected by habituation. This was a possible in our design because, unlike in the animal literature where oddball sequences last for several minutes, we used numerous short trials that allowed us to accurately estimate the responses to the first, unadapted standard.

Deviant detecting voxels were identified using the neuronal mismatch index, in our study defined by the contrast dev4 > 0.5 std1 + 0.5 std2 (cf., with the animal literature index dev > std). Since we had two estimations of the adapted standards, std1 and std2, we calculated their average. We only included dev4 (excluding dev5 and dev6) because the effects of abstract expectations, absent in animal studies, were expected to be the lowest in dev4. We tested both contrasts using right-tailed rank-sum tests. Before conducting the rank-sum tests, we averaged single-voxel beta estimates for each experimental condition across all experimental runs.

We defined SSA regions for each stimulus type as the set of voxels showing significant adaptation and deviant detection; namely, voxels that adapted to frequently occurring sounds while showing preserved responses to deviants (Ulanovsky et al. 2003). We computed voxel-wise P-values for SSA as the maximum of the uncorrected P-values for adaptation and deviant detection in each voxel; 

$$\text{SSA} = \max(P_{\text{adaptation}}, P_{\text{deviant detection}}).$$

All voxels’ P-values were subsequently controlled for the false discovery rate (FDR) using the Benjamini–Hochberg method (Benjamini and Hochberg 1995) and thresholded at \(\alpha = 0.05\). Peak-level P-values were corrected for the FWE rate: we corrected for the number of voxels per ROI using Bonferroni-correction and for the total number of comparisons using the Holm–Bonferroni method (Holm 1979).

### Quantifying SSA magnitude

To quantify SSA magnitude in each voxel of the anatomical ROIs, we computed the standardized voxel-wise index of SSA (SSAi) (Ulanovsky et al. 2003) following the procedure described in previous research (Tabas et al. 2021b; Tabas et al. 2020). We normalized the beta estimates for dev4, std1, and std2 to a range from zero to one, averaged these values in each voxel across participants and runs, and computed the index of SSA as 

$$\text{SSAi} = (\text{dev}4 - 0.5 \text{std}1 - 0.5 \text{std}2)/((\text{dev}4 + 0.5 \text{std}1 + 0.5 \text{std}2)).$$

### Classical analysis

For both stimulus types, we conducted a classical statistical analysis to test for differences between responses to deviants in positions four, five, and six. To specifically investigate mechanisms driving SSA, we restricted this analysis to SSA clusters with significant \((P < 0.05, \text{FWE-corrected})\) peak-level P-values—the SSA ROIs. We tested the pairwise differences between responses to deviants in different positions (dev4 > dev5, dev4 > dev6, and dev5 > dev6) in each SSA ROI using one-sided Wilcoxon sign rank tests at the group-level. Before testing the contrasts, we averaged data corresponding to each experimental condition across runs and voxels within each participant and SSA ROI. In line with the idea of prediction error encoding, we expected deviant responses to be stronger when deviants are less expected.

We also measured the effect size of adaptation and deviant detection by testing the contrasts std0 > std2 and dev4 > std2. Note that these contrasts are not independent of the contrasts used for SSA voxel selection. However, we included them here to be able to quantify the size of both effects. Additionally, we included the comparison of dev6 and std2 using two-tailed Wilcoxon sign rank tests. We included this analysis to test whether responses to fully predictable deviants were comparable to standard responses. In line with predictive coding, we expected no statistically significant difference between responses to dev6 and std2. That is because both types of sound are fully predictable given our experimental design. All P-values were corrected for multiple comparisons using the Holm–Bonferroni method (Holm 1979).

### Correlational analysis and linear mixed-effects model

To investigate the hypothesized negative relationship between deviant predictability and deviant responses further, we estimated a linear mixed-effects model (LMM) in each SSA ROI for each data set at the group-level. For pure tones, the model included deviant predictability as a fixed effect and random intercepts and slopes for experimental runs and participants

$$\text{beta} \sim 1 + \text{predictability} + (1 + \text{predictability}|\text{run}) + (1 + \text{predictability}|\text{participant}).$$

For FM-sweeps, we used the same model but added experimental session as an additional random effect. All P-values were Bonferroni-corrected for the total number of SSA ROIs. To test if the group-level results were replicated at the participant-level, we computed Spearman’s rank correlation between deviant predictability \((1/3 \text{ for dev4}, 1/2 \text{ for dev5}, \text{ and } 1 \text{ for dev6})\) and standardized beta estimates for dev4, dev5, and dev6 in each participant. Before computing the correlation coefficient \(\rho\), beta estimates for the different deviant conditions in each voxel were averaged across experimental runs.

### Bayesian model comparison

We constructed two models, each representing one potential encoding mechanism driving SSA (see Fig. 1E). The models were defined using parametric amplitude modulation vectors that specified the predicted responses to all tones in each trial.

\textbf{h1) Habituation:} SSA is based on stimulus repetition. Responses to standards undergo habituation over time and recover slightly after the deviant. Deviant responses are fully recovered and do not differ between deviants of differential predictability.

The h1 model (see Fig. 1E, left) was specified by assigning the amplitude 1 to std0 and the deviant of a sequence. Standards before the deviant were assigned the amplitudes \(1/n\) and standards after the deviant were assigned the amplitudes \(1/(n – 1)\), where \(n\) is the position of the standard within the sequence (see Table 1 for the exact amplitudes). Amplitudes were chosen based on a large body of literature demonstrating that adaptation in the AC is best described as exponential decay (e.g., Bäuerle et al. 2011; Ulanovsky et al. 2004).

\textbf{h2) Prediction error:} Neural responses to sounds are scaled by stimulus predictability and thus represent prediction error responses, which are stronger when stimuli are less expected.

The h2 model (see Fig. 1E, right) used an amplitude of 0.5 for std0 and an amplitude of \(P = \text{probability of stimulus occurrence}\) for the rest of the tones. For example, a sequence with a deviant in position five was assigned the amplitude vector \(\text{amp}_5 = [1/2, 1, 1, 2, 1/2, 1, 1, 1]\). Thus, std4 was assigned a value of 2/3 since a standard in position four is expected with a probability of 2/3, and dev5 was assigned a value of 1/2 because deviants in
position five are expected with a probability of 1/2 after hearing four standards. This definition corresponds to the established definition of precision-weighted prediction error (Friston 2003). See Table 1 for all amplitudes of h2. Since model estimation using parametric modulation is symmetric with respect to linear transformations of the amplitudes, the above-described amplitudes are equivalent to assuming a decreasing response with increasing sound predictability. We disregarded the trial information on the deviant-standard combination because the predictability of the deviant was the regressor that best separated the habituation and prediction error hypotheses. Adding additional regressors might have helped fine-tuning the definitions of the models, but we preferred to tailor our analysis to solve our specific research question.

For each subject, we computed the log-evidence of the two models in each voxel of all anatomical ROIs using SPM’s Bayesian Estimation functions in Nipype. Before model fitting, the amplitudes of all models were z-standardized according to experimental runs.

For pure tones, all models were estimated using the smoothed functional data in their native space. The log-evidence maps were registered to the individual T1 scans and then to the MNI152 symmetric template using Freesurfer’s ApplyVolTransform and ANTs’ ApplyTransforms, respectively. We combined the log-evidence-maps of all participants and calculated posterior probability maps for each model using previously written custom code (Tabas et al. 2020) following the methodology described elsewhere (Rosa et al. 2010; Stephan et al. 2009).

For FM-sweeps, all models were estimated using the smoothed functional data across sessions in the space of the participants’ anatomical scans. The resulting log-evidence-map of each participant was registered to the MNI152 symmetric template using ANTs’ ApplyTransforms and posterior probability maps were calculated as described before.

To compare the posterior likelihood of the two models (h1 and h2), we computed the Bayes factor K for each stimulus type in each voxel of each anatomical ROI. We plotted Bayes factor maps (i.e., $K_{h2/h1}$) as an indicator of the strength of evidence in favor of h2, our model of interest.

### Results

#### Topography of SSA to pure tones and FM-sweeps in AC

There was a significant SSA to pure tones in bilateral Te1.0, bilateral Te1.1, and right Te3 ($P < 0.008$, FWE-corrected, see Table 2). Significant SSA clusters were present in all anatomical ROIs for FM-sweeps ($P < 0.03$, FWE-corrected, see Table 2). SSA clusters formed coherent fields for both stimulus types, indicating a systematic spatial encoding pattern (see Fig. 2 for a detailed map of adapting, deviant detecting, and SSA voxels).

For pure tones, SSA voxels were located lateral within bilateral Te1.0, superior within bilateral Te1.1, and predominantly posterior in right Te3 (Fig. 3). For FM-sweeps, the majority of voxels in bilateral Te1.0 and Te1.1 showed significant SSA. SSA voxels in bilateral Te1.2 were localized posterolaterally. In Te3, SSA voxels were mostly found in posterior areas, mirroring the findings from the pure tone experiment (Fig. 3C).

The SSA magnitude SSAi was similarly distributed across all anatomical ROIs and both stimulus types (Fig. 2, lower panels). This indicates that, in opposite to animal literature (Nieto-Diego and Malmierca 2016; Parras et al. 2017), SSA magnitude did not differ between areas of the primary and secondary AC.

### SSA to pure tone and FM-sweep deviants is driven by the encoding of prediction error

To test whether SSA regions encoded prediction error, we examined whether subjective expectations modulated the responses to deviants in different positions by testing the linear relationship between deviant predictability and deviant responses. Generally, beta estimates qualitatively decreased with increasing deviant predictability for both stimulus types in accordance with the predictive coding hypothesis. We corroborated the effect quantitatively by conducting pair-wise statistical comparisons between the responses to each deviant position at the group-level (Supplementary Tables S1 and S2).

In most SSA ROIs, deviants in position four elicited stronger responses than deviants in positions five and six ($P < 0.03$); deviants in position five elicited stronger responses than deviants in position six ($P < 0.04$; all P-values of the pure tone data corrected for 30 comparisons and all P-values of the FM-sweep data corrected for 48 comparisons).

Responses to deviants in position six were not statistically different from responses to standards after the deviants in most SSA ROIs and both stimulus types, again in consistency with the predictive coding framework. The only exception was Te1.1 R in the pure tone experiment, where responses to deviants in position six were significantly higher than responses to standards after the deviant. In the remaining regions, full predictability seemed to eliminate deviant detection responses completely. It might be tempting to alternatively interpret these results as an effect of selective attention: both deviants in position six and standards after the deviant are irrelevant for the task because the correct response in each trial is already known; while tones in positions four and five are the most relevant for the task because their identity is uncertain. We find this interpretation unconvincing because of two reasons: (i) unattended but unexpected deviants have been shown to elicit stronger responses than unattended

| Table 1. Amplitudes of the models used for Bayesian model comparison. h1 (habituation): asymptotically decreasing responses to repeating standards, full responses to deviants, and a slight recovery after the deviant; h2 (prediction error): neural responses scaled by sound predictability, amplitude defined as the probability P of finding the heard sound in each position of the sequence. |
|---------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| **h1 deviant position** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| a_1 | 4 | 1 | 1/2 | 1/3 | 1 | 1/3 | 1/4 | 1/5 | 1/6 |
| a_2 | 5 | 1 | 1/2 | 1/3 | 1/4 | 1 | 1/4 | 1/5 | 1/6 |
| a_3 | 6 | 1 | 1/2 | 1/3 | 1/4 | 1/5 | 1 | 1/5 | 1/6 |

| **h2 deviant position** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| a_1 | 4 | 1/2 | 1 | 1 | 1/3 | 1 | 1 | 1 |
| a_2 | 5 | 1/2 | 1 | 1 | 2/3 | 1/2 | 1 | 1 |
| a_3 | 6 | 1/2 | 1 | 1 | 2/3 | 1/2 | 1 | 1 |

| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
Table 2. Cluster sizes, MNI peak coordinates (mm), and peak-level P-values for adaptation, deviant detection, and SSA to pure tones and FM-sweeps. Voxel-wise P-values were FDR-corrected in each ROI and thresholded at $\alpha = 0.05$. Peak-level P-values were corrected for the number of voxels per ROI and the total number of comparisons using Bonferroni– and Holm–Bonferroni-correction, respectively. Maps showing voxels in which the respective contrasts were significant are provided in Fig. 2.

| Contrast      | Pure Tones | FM-Sweeps |
|---------------|------------|-----------|
|               | Size       | Coordinates | Peak $P$   | Size       | Coordinates | Peak $P$   |
| Adaptation    |            |            |            |            |            |            |
| Te1.0 L       | 539        | [$-48, -25, 9$] | 0.007      | 950        | [$-42, -20, 5$] | 0.008      |
| Te1.0 R       | 759        | [$54, -16, 7$] | 0.003      | 1085       | [$54, -15, 4$] | 0.009      |
| Te1.1 L       | 920        | [$-37, -33, 14$] | 0.003      | 1145       | [$39, -23, 5$] | 0.008      |
| Te1.1 R       | 1374       | [$35, -29, 18$] | 0.003      | 1528       | [$50, -11, -1$] | 0.009      |
| Te1.2 L       | 118        | [$-48, -13, 2$] | 0.008      | 715        | [$51, -4, -4$] | 0.02       |
| Te1.2 R       | 119        | [$57, -3, -6$] | 0.02       | 730        | [$51, -4, -4$] | 0.01       |
| Te3 L         | 4002       | [$-65, -19, 9$] | 0.02       | 4747       | [$-58, -14, -4$] | 0.02       |
| Te3 R         | 3688       | [$63, -22, 0$] | 0.008      | 1420       | [$-64, -28, 4$] | 0.02       |
| Deviant Detection |        |            |            |            |            |            |
| Te1.0 L       | 788        | [$-48, -25, 10$] | 0.003      | 908        | [$-48, -22, 7$] | 0.008      |
| Te1.0 R       | 1042       | [$54, -16, 7$] | 0.002      | 1015       | [$51, -18, 5$] | 0.008      |
| Te1.1 L       | 1000       | [$-38, -32, 13$] | 0.002      | 1134       | [$-38, -24, 8$] | 0.008      |
| Te1.1 R       | 1532       | [$38, -22, 6$] | 0.003      | 1424       | [$48, -23, 7$] | 0.008      |
| Te1.2 L       | 619        | [$-53, 2, -3$] | 0.003      | 631        | [$-51, -11, -1$] | 0.008      |
| Te1.2 R       | 462        | [$54, -1, -3$] | 0.02       | 547        | [$57, -4, -4$] | 0.008      |
| Te3 L         | 3128       | [$-62, -21, 9$] | 0.01       | 3954       | [$-66, -19, 1$] | 0.03       |
| Te3 R         | 3239       | [$65, -22, 1$] | 0.007      | 3901       | [$63, -28, 4$] | 0.02       |
| SSA           |            |            |            |            |            |            |
| Te1.0 L       | 499        | [$-48, -25, 9$] | 0.006      | 906        | [$-48, -22, 8$] | 0.008      |
| Te1.0 R       | 748        | [$54, -16, 7$] | 0.002      | 995        | [$51, -18, 6$] | 0.007      |
| Te1.1 L       | 88         | [$-37, -33, 14$] | 0.003      | 1092       | [$-38, -24, 8$] | 0.007      |
| Te1.1 R       | 1372       | [$35, -30, 18$] | 0.005      | 1413       | [$41, -28, 14$] | 0.007      |
| Te1.2 L       | 82         | [$-48, -13, 3$] | 0.06       | 570        | [$-51, -9, -1$] | 0.02       |
| Te1.2 R       | 0          | -         | -          | 531        | [$51, -4, -4$] | 0.01       |
| Te3 L         | 2435       | [$-63, -20, 10$] | 0.05       | 3586       | [$-63, -15, -3$] | 0.02       |
| Te3 R         | 2439       | [$65, -22, 1$] | 0.007      | 3351       | [$66, -20, 6$] | 0.02       |

Neural responses to pure tones and FM-sweeps are best explained by predictive coding

The prediction error model (h2) outperformed the habituation model (h1) in most voxels of all ROIs for both stimulus types, indicating that the effects of predictability are stronger than the effects of habituation in the investigated ROIs. Voxel-wise maps and the distributions of the Bayes factor $\text{K}_{h2/h1}$ comparing both models are shown in Fig. 4.

For the pure tone data, the prediction error model (h2) was the best explanation for the data (posterior density of $h2 \mid P \gg 0.5$) in 96% of voxels in Te1.0 L, 94% in Te1.0 R, 97% in Te1.1 L, 99% in Te1.1 R, 99% in Te1.2 L, 90% in Te1.2 R, 82% in Te3 L, and 72% in Te3 R. H2 was also the best explanation for the FM-sweep data in a majority of voxels of all anatomical ROIs (98% of voxels in Te1.0 L, 89% in Te1.0 R, 90% in Te1.1 L, 94% in Te1.1 R, 97% in Te1.2 L, 83% in Te1.2 R, 91% in Te3 L, and 82% in Te3 R).

Discussion

The main tenet of the predictive coding framework is that sensory neurons encode prediction error with respect to an internal generative model of the sensory world (Friston 2003; Mumford 1992; Rao and Ballard 1999; Spratling 2017). Despite a growing body of evidence demonstrating prediction error encoding of pure tones in the human AC, many previous studies have potentially confounded habituation with prediction error; disregarded the possibility that prediction error encoding differs across fields within the AC, and focused on the encoding of pure tones, neglecting the rich diversity of dynamic sounds that populates the auditory landscape.

The present study yielded three key findings: first, we found that the prediction error model explained our data better than the habituation model. This suggests that, at least at the coarse temporal resolution of fMRI, BOLD responses to sequences of sounds in the AC were mainly driven by the participants’ expectations that the prediction error model explained our data better than the habituation model. Second, we showed that the magnitude of SSA was similar across different auditory fields, indicating that there is no systematic difference between adaptation in the primary and secondary auditory pathways in humans. Third, we demonstrated that SSA and prediction error encoding are pervasive in pure tones and...
frequency-modulated sweeps, suggesting that both stimulus types share a common mechanism for the computation of prediction error. Together, our results indicate that predictive coding is the general encoding mechanism of acoustic information in the human AC.

We found significant SSA to pure tones in bilateral Te1.0 and Te1.1 and in right Te3; SSA in Te1.2 did not reach statistical significance due to a lack of adaptation to the standards. Previous studies have shown that Te1.0 (Bailey et al. 2007) and Te1.2 (Patterson et al. 2002; von Kriegstein et al. 2010) selectively responded to pitch in auditory stimuli, whereas Te1.1 responded selectively to broadband matched noises (Bailey et al. 2007; von Kriegstein et al. 2010). Consequently, Te1.2 has been suggested as the locus of pitch processing in the human AC (De Angelis et al. 2018). Since pitch perception requires integrating auditory information along time windows of only a few milliseconds (Balaguer-Ballester et al. 2008; Hall III and Peters 1981; Meddis and O’Mard 1997), pitch sensitive areas might be insensitive to regularities unfolding at longer time spans (e.g., the ITI of 700 ms that we used for our experiments).

Further work will need to examine how regions that are particularly sensitive to one specific auditory feature show adaptation to other features. For instance, a recent study (An et al. 2021) has shown that evoked responses and mismatch responses to different stimulus features (including pitch) in the rat’s AC show different spatial distributions, with broader distributed mismatch responses.

Our results are the first robust evidence for prediction error encoding of FM in human AC. In line with our results, SSA to FM direction was reported in A1 of rats (Klein et al. 2014). Results from the human literature are more difficult to reconcile with our findings. Three previous studies reported a significant MMN to deviating FM-sweeps (Cornella et al. 2013; Hsieh and Yeh 2021; Kung et al. 2020), but since they did not localize the source of the potentials, it remained unclear to what extent these responses
were generated in the AC. One further study investigated sources in the AC but reported no significant results (Altmann et al. 2011). Three other studies reported increasing neuromagnetic responses to repeated FM-sweeps (Heinemann et al. 2011; Heinemann et al. 2010; Okamoto and Kakigi 2017), in direct contradiction with predictive coding. One of these studies (Okamoto and Kakigi 2017) reported the effect specifically in the AC. The short ISIs used in some of these latter studies (e.g. 200 and 100 ms in Heinemann et al. 2010 and Heinemann et al. 2011, respectively) might have contributed to the contradictory results: different temporal integration mechanisms might apply to stimulus sequences spanning shorter or longer time scales.

Our study had a much lower temporal resolution than previous E/MEG studies, with a TR that was longer than twice the

![Fig. 3.](image-url) (A and B) 3D view of voxels showing significant (P < 0.05, FDR-corrected) SSA (purple) within Te1.0 (dark blue), Te1.1 (green), and Te1.2 (light blue). (C) Voxels of Te3 (yellow) showing significant (P < 0.05, FDR-corrected) SSA to both, pure tones and FM-sweeps (purple), voxels showing SSA only to pure tones (blue) or FM-sweeps (green).

Table 3. Fixed effect coefficients of the group-level LMM. Pure tone model: beta ∼ 1 + predictability + (1 + predictability|run) + (1 + predictability|participant); FM-sweep model: beta ∼ 1 + predictability + (1 + predictability|session) + (1 + predictability|run) + (1 + predictability|participant). Int.: Intercept; Pred.: fixed effect regressor of the model, predictability of the deviants; DF: degrees of freedom; P: P-value of the t-test testing for the equality of the coefficient to zero; CI: limits of the confidence interval for the coefficients. All P-values were Bonferroni-corrected for N = 13 SSA ROIs. Empty cells indicate that no SSA-cluster with a statistically significant P-value was found in the respective ROIs.

| ROI   | Name  | Pure Tones | FM-Sweeps |
|-------|-------|------------|-----------|
|       | β     | DF         | P         | CI         | β         | DF         | P         | CI         |
| Te1.0 L | Int.  | 0.89       | 113770    | 2.1e-02   | [0.33, 1.45] | 0.87       | 437596    | 8.2e-05   | [0.49, 1.25] |
|       | Pred. | -1.13      | 113770    | 4.3e-02   | [-1.89, -0.38] | -1.27      | 437596    | 3.8e-05   | [-1.86, -0.67] |
| Te1.0 R | Int.  | 0.10       | 170542    | 8.4e-02   | [-2.10, 1.14] | 0.80       | 480583    | 3.0e-05   | [0.47, 1.13] |
|       | Pred. | -1.24      | 170542    | 5.2e-02   | [-2.33, -0.16] | -1.11      | 480583    | 1.1e-03   | [-1.67, -0.55] |
| Te1.1 L | Int.  | 1.22       | 202234    | 7.5e-02   | [0.69, 1.04] | 1.04       | 527434    | 9.2e-11   | [0.74, 1.34] |
|       | Pred. | -1.52      | 202234    | 1.1e-02   | [-2.24, -0.80] | -1.47      | 527434    | 8.5e-11   | [-1.89, -1.95] |
| Te1.1 R | Int.  | 1.77       | 312814    | 1.6e-02   | [0.70, 1.64] | 0.90       | 682477    | 1.3e-11   | [0.65, 1.15] |
|       | Pred. | -1.41      | 312814    | 2.9e-02   | [-2.08, -0.74] | -1.36      | 682477    | 4.2e-10   | [-1.77, -0.96] |
| Te1.2 L | Int.  | -          | -         | -         | -         | 0.69       | 275308    | 3.9e-02   | [0.23, 1.14] |
|       | Pred. | -          | -         | -         | -         | -0.96      | 275308    | 8.2e-02   | [-1.65, -0.27] |
| Te1.2 R | Int.  | -          | -         | -         | -         | 0.49       | 256471    | 6.7e-02   | [0.14, 0.84] |
|       | Pred. | -          | -         | -         | -         | -0.60      | 256471    | 3.9e-01   | [-1.14, -0.06] |
| Te3 L  | Int.  | -          | -         | -         | -         | 0.61       | 1732036   | 1.4e-05   | [0.36, 0.86] |
|       | Pred. | -          | -         | -         | -         | -0.75      | 1732036   | 4.4e-04   | [-1.11, -0.40] |
| Te3 R  | Int.  | 0.70       | 556090    | 3.8e-02   | [0.23, 1.16] | 0.70       | 1618531   | 3.1e-05   | [0.41, 1.00] |
|       | Pred. | -0.8       | 556090    | 3.0e-01   | [-1.46, -0.19] | -0.83      | 1618531   | 2.6e-03   | [-1.26, -0.39] |
Fig. 4. Bayes’ $K_{02/01}$ for pure tones and FM-sweeps. The upper panels show $K_{02/01}$ in each voxel of the anatomical ROIs; the lower panels show the distribution of $K_{02/01}$ across all voxels of the anatomical ROIs. (A) Pure tones; (B) FM-sweeps. The results indicate evidence in favor of the prediction error model ($h2$) in most voxels of all anatomical ROIs for both stimulus types.

We adopted several strategies to circumvent this problem. First, we ensured that deviants were separated by an average of 5 s, which allowed us to robustly estimate their responses; since inspecting the responses to the deviants is sufficient to differentiate between prediction error and habituation, the temporal resolution did not detriment the potential of the paradigm to answer our main research question. Second, we used eight-tone sequences to ensure that the first standard within each sequence was always sufficiently separated from the previous and the next deviant, also allowing for an accurate estimation of its responses. Estimating the responses to the repeated standards was substantially more challenging; we circumvented this problem by introducing the reasonable assumption that responses to subsequent repeated standards would only vary linearly, allowing us to replace all repetitions with four single parameters. Since the repeated standards are the most common regressor in the design, the GLMs had also more data to accurately estimate their responses. Nevertheless, the comparatively lower temporal resolution of our study should be kept in mind when interpreting our findings.

With respect to SSA magnitude, we did not find differences between primary and secondary auditory areas: magnitudes were comparable across all ROIs. However, the proportion of SSA-voxels was descriptively lower in Te3 compared with all other ROIs for FM-sweeps and compared with Te1.0 and Te1.1 for pure tones. This result suggests that not all auditory fields might exhibit SSA to the same extent. Furthermore, we found deviant detection and repetition suppression for pure tones to differ across fields: Te1.2 was dominated by deviant detection, while all other ROIs additionally showed significant repetition suppression and thus SSA.

In rodents, on the contrary, SSA magnitude clearly differed between fields of the primary and secondary auditory pathway in previous studies: secondary cortical areas showed stronger SSA than primary cortical areas (Nieto-Diego and Malmierca 2016; Parras et al. 2017) and the same pattern was reported in the subcortical auditory pathway (Antunes and Malmierca 2011; Antunes et al. 2010; Ayala et al. 2015; Duque et al. 2014; Duque et al. 2012; Parras et al. 2017).

The discrepancies between studies in rodents and our results may stem from a poor correspondence between auditory fields in these two species: while AC in rodents is typically subdivided in three lemniscal and two nonlemniscal fields (Parras et al. 2017), in primates, AC is subdivided in three distinct fields: core, belt, and parabelt (Besle et al. 2019). We used a previous segmentation...
of human AC with four different fields (Morosan et al. 2001) that cannot be trivially mapped to the animal subdivisions and whose classification as primary or secondary is still of debate (Beslie et al. 2019, Moerel et al. 2014). It is possible that using a different anatomical segmentation (e.g. using Brodman areas instead) would yield different results. However, our previous results showed that the distributions of SSA in the subcortical auditory pathway, which is closely replicated across mammals (Glendenning and Masterton 1998), also differed between humans and rodents: SSA magnitude in our human subjects was similar across different primary and secondary subcortical auditory areas (Tabas et al. 2021b; Tabas et al. 2020). Moreover, the spatial and temporal resolution of fMRI might obscure differences between auditory fields that may only be apparent in single-unit recordings.

Until recently, it was unclear whether predictions from generative model units inform prediction errors only at the immediate lower stage of the processing hierarchy or also at subsequently lower stages (see Tabas et al. 2021a for a review of the empirical evidence on both standpoints). MMN studies showed that prediction error is elicited with respect to high-level expectations; namely by the violation of complex statistical regularities (see Paavilainen 2013 for review), the omission of expected sounds (Bendixen et al. 2009; Chennu et al. 2016; Wacongne et al. 2011), and abstract expectations about the occurrence of deviating sounds (Wacongne et al. 2011). However, since the generators of the MMN are partly located in the frontal cortex (Paavilainen 2013), MMN research cannot clarify whether subjective expectations are used to compute prediction errors at lower levels of the auditory processing hierarchy.

We found prediction error encoding to be a dominant encoding principle for both stimulus types in all anatomical ROIs. This result suggests that high-level predictions informed by the task instructions, putatively computed in regions at higher processing stages than the sensory cortices, are used to compute prediction errors in the primary AC. We had previously shown that these same predictions are also used to compute prediction error in the human auditory midbrain and thalamus (Tabas et al. 2021b; Tabas et al. 2020). Previous studies also showed that prediction error in the AC was computed with respect to language-specific expectations (e.g. Blank and Davis 2016; Blank et al. 2018; Heilbron et al. 2021; Ylinen et al. 2016). Together, the empirical evidence supports the hypothesis that high-level predictions are used to compute prediction errors along the entire predictive processing hierarchy.

Our study did not tackle the question of where SSA and prediction errors are first generated. Since auditory cortical areas receive direct bottom-up input from the auditory thalamus (Schrofield 2011), SSA and prediction error signals in AC could reflect ascending input from prediction error units in subcortical structures (Tabas et al. 2021b; Tabas et al. 2020). Conversely, subcortical SSA and prediction error signals might as well be inherited from cerebral cortex areas via corticofugal modulation (Malmierca 2015). Animal studies have shown that SSA in secondary auditory midbrain and thalamus persists under deactivation of the AC (Anderson and Malmierca 2013; Antunes and Malmierca 2011), but not in the primary auditory thalamus, where SSA is the weakest (Bäuerle et al. 2011). This suggests that subcortical SSA cannot be entirely inherited from the AC. Further work is needed to clarify the interplay of bottom-up and top-down signaling in the computation of prediction error.

Our results suggest that predictive coding is the general mechanism underlying the encoding of acoustic features in AC. Impaired predictive processes in AC have been linked to speech processing disorders and clinical conditions such as developmental dyslexia (e.g. Gu and Bi 2020; Neuhoff et al. 2012; Plakas et al. 2013), stuttering (Daliri and Max 2015), autism spectrum disorder (van Schalkwyk et al. 2017), psychosis (Fryer et al. 2020; Sterzer et al. 2018), or schizophrinia (Perez et al. 2014). Investigating how predictive coding is implemented in the human AC is essential for a mechanistic understanding of perception and its dysfunction.

### Funding

H2020 European Research Council Consolidator Grant SENSO-COM (647051 to KvK and AT).

### References

Altman CF, Gaese BH. Representation of frequency-modulated sounds in the human brain. Hear Res. 2014:307:74–85.

Altman CF, Klein C, Heinemann LV, Wibral M, Gaese BH, Kaiser J. Repetition of complex frequency-modulated sweeps enhances neuromagnetic responses in the human auditory cortex. Hear Res. 2011:282(1–2):216–224.

An H, Aukstzulewicz R, Kang H, Schnupp JW. Cortical mapping of mismatch responses to independent acoustic features. Hear Res. 2021:399:107894. https://doi.org/10.1016/j.heares.2020.107894.

Anderson L, Malmierca M. The effect of auditory cortex deactivation on stimulus-specific adaptation in the inferior colliculus of the rat. Eur J Neurosci. 2013:37(1):52–62.

Antunes FM, Malmierca MS. Effect of auditory cortex deactivation on stimulus-specific adaptation in the medial geniculate body. J Neurosci. 2011:31(47):17306–17316.

Antunes FM, Nelken I, Covey E, Malmierca MS. Stimulus-specific adaptation in the auditory thalamus of the anesthetized rat. PLoS One. 2010:5(11):e14071. https://doi.org/10.1371/journal.pone.0014071.

Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage. 2011:54(3):2033–2044.

Ayala YA, Udeh A, Dutta K, Bishop D, Malmierca MS, Oliver DL. Differences in the strength of cortical and brainstem inputs to ssa and non-ssa neurons in the inferior colliculus. Sci Rep. 2015:5(1):1–17.

Bailey L, Abolmaesumi P, Tam J, Morosan P, Cusack R, Arnunts K, Johnsrude I. Customised cytoarchitectonic probability maps using deformable registration: primary auditory cortex. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer. 2007: pp. 760–768.

Balaguer-Ballester E, Denham SL, Meddis R. A cascade autocorrelation model of pitch perception. J Acoust Soc Am. 2008:124(4):2186–2195.

Bäuerle P, von der Behrens W, Kössl M, Gaese BH. Stimulus-specific adaptation in the gerbil primary auditory thalamus is the result of a fast frequency-specific habituation and is regulated by the corticofugal system. J Neurosci. 2011:31(26):9708–9722.

Bendixen A, Schröger E, Winkler J. I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. J Neurosci. 2009:29(26):8447–8451.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995:57(1):289–300.

Berlot E, Formisano E, De Martino F. Mapping frequency-specific tone predictions in the human auditory cortex at high spatial resolution. J Neurosci. 2018:38(21):4934–4942.
Nieto-Diego J, Malmierca MS. Topographic distribution of stimulus-specific adaptation across auditory cortical fields in the anesthetized rat. PloS Biol. 2016;14(5):e1002397. https://doi.org/10.1371/journal.pbio.1002397.

Okamoto H, Kakigi R. Modulation of auditory evoked magnetic fields elicited by successive frequency-modulated (fm) sweeps. Front Hum Neurosci. 2017;11:36. https://doi.org/10.3389/fnhum.2017.00036.

Paavilainen P. The mismatch-negativity (mmn) component of the auditory event-related potential to violations of abstract regularities: a review. Int J Psychophysiol. 2013;88(2):109–123.

Paltoglou AE, Sumner CJ, Hall DA. Mapping feature-sensitivity and attentional modulation in human auditory cortex with functional magnetic resonance imaging. Eur J Neurosci. 2011;33(9):1733–1741.

Parras GG, Nieto-Diego J, Carabajal GV, Valdés-Baizabal C, Escera C, Malmierca MS. Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. Nat Commun. 2017;8(1):1–17.

Patterson RD, Upenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. Neuron. 2002;36(4):767–776.

Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. Statistical parametric mapping: the analysis of functional brain images. London: Elsevier; 2011.

Perez VB, Woods SW, Roach BJ, Ford JM, McGlasshan TH, Srinhari VH, Mathalon DH. Automatic auditory processing deficits in schizophrenia and clinical high-risk patients: forecasting psychosis risk with mismatch negativity. Biol Psychiatry. 2014;75(6):459–469.

Pérez-González D, Parras GG, Morado-Díaz CJ, Aedo-Sánchez C, Carabajal GV, Malmierca MS. Deviance detection in physiologically identified cell types in the rat auditory cortex. Hear Res. 2021;399:107997. https://doi.org/10.1016/j.heares.2020.107997.

Phillips HN, Blenkmann A, Hughes LE, Kochen S, Bekinschtein TA, Rowe JB, Cam-CAN. Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. Cortex. 2016;82:192–205.

Plakas A, van Zuijen T, van Leeuwen T, Thomson JM, van der Leij A. Impaired non-speech auditory processing at a pre-reading age is a risk-factor for dyslexia but not a predictor: an erp study. Cortex. 2013;49(4):1034–1045.

Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci. 1999;2(1):79–87.

Rosa MJ, Bestmann S, Harrison L, Penny W. Bayesian model selection maps for group studies. NeuroImage. 2010;49(1):217–224.

Rubin J, Ulavonsky N, Neiken I, Tishby N. The representation of prediction error in auditory cortex. PloS Comput Biol. 2016;12(8):e1005058. https://doi.org/10.1371/journal.pcbi.1005058.

Schofield BR. Central descending auditory pathways. In: Auditory and vestibular efferents. New York, NY: Springer; 2011. pp. 261–290.

Shalgi S, Deouell LY. Direct evidence for differential roles of temporal and frontal components of auditory change detection. Neuropsychologia. 2007;45(8):1878–1888.

Spratling MW. A review of predictive coding algorithms. Brain Cogn. 2017;112:92–97.
Suga N. Basic acoustic patterns and neural mechanisms shared by humans and animals for auditory perception. In: Listening to speech. New York, NY: Psychology Press; 2012.

Tabas A, Mihai G, Kiebel S, Trampel R, von Kriegstein K. Abstract rules drive adaptation in the subcortical sensory pathway. *elife*. 2020;9:e64501. https://doi.org/10.7554/eLife.64501.

Tabas A, Kiebel S, Marxen M, von Kriegstein K. Fast frequency modulation is encoded according to the listener expectations in the human subcortical auditory pathway. *arXiv* preprint arXiv:2108.02066. 2021a.

Tabas A, von Kriegstein K. Adjudicating between local and global architectures of predictive processing in the subcortical auditory pathway. *Front Neural Circuits*. 2021b;15:644743. https://doi.org/10.3389/fncir.2021.644743.

Todorovic A, de Lange FP. Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J Neurosci*. 2012;32(39):13389–13395.

Todorovic A, van Ede F, Maris E, de Lange FP. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an meg study. *J Neurosci*. 2011;31(25):9118–9123.

Ulanovsky N, Las L, Nelken I. Processing of low-probability sounds by cortical neurons. *Nat Neurosci*. 2003;6(4):391–398.

Ulanovsky N, Las L, Farkas D, Nelken I. Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci*. 2004;24(46):10440–10453.

van Schalkwyk GI, Volkmar FR, Corlett PR. A predictive coding account of psychotic symptoms in autism spectrum disorder. *J Autism Dev Disord*. 2017;47(5):1323–1340.

Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S. Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc Natl Acad Sci*. 2011;108(51):20754–20759.

Wang H, Han Y-F, Chan Y-S, He J. Stimulus-specific adaptation at the synapse level in vitro. *PLoS One*. 2014;9(12):e114537. https://doi.org/10.1371/journal.pone.0114537.

Ylinen S, Huuskonen M, Mikola K, Saure E, Sinkkonen T, Paavilainen P. Predictive coding of phonological rules in auditory cortex: A mismatch negativity study. *Brain Lang*. 2016;162:72–80.