Check for updates

Fast frequency modulation is encoded according to the listener expectations in the human subcortical auditory pathway

Alejandro Tabas[a,b,c], Stefan Kiebel[d], Michael Marxen[e,f], Katharina von Kriegstein[c,d]

[a]Basque Center on Cognition, Brain, and Language, San Sebastian, Spain
[b]Ikerbasque, Basque Foundation for Science, Bilbao, Spain
[c]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[d]Department of Psychology, Technische Universität Dresden, Dresden, Germany
[e]Department of Psychiatry, Technische Universität Dresden, Dresden, Germany
[f]Neuroimaging Center, Technische Universität Dresden, Dresden, Germany

Corresponding Author: Katharina von Kriegstein (katharina.von_kriegstein@tu-dresden.de)

ABSTRACT

Expectations aid and bias our perception. For instance, expected words are easier to recognise than unexpected words, particularly in noisy environments, and incorrect expectations can make us misunderstand our conversational partner. Expectations are combined with the output from the sensory pathways to form representations of auditory objects in the cerebral cortex. Previous literature has shown that expectations propagate further down to subcortical stations during the encoding of static pure tones. However, it is unclear whether expectations also drive the subcortical encoding of subtle dynamic elements of the acoustic signal that are not represented in the tonotopic axis. Here, we tested the hypothesis that subjective expectations drive the encoding of fast frequency modulation (FM) in the human subcortical auditory pathway. We used fMRI to measure neural responses in the human auditory midbrain (inferior colliculus) and thalamus (medial geniculate body). Participants listened to sequences of FM-sweeps for which they held different expectations based on the task instructions. We found robust evidence that the responses in auditory midbrain and thalamus encode the difference between the acoustic input and the subjective expectations of the listener. The results indicate that FM-sweeps are already encoded at the level of the human auditory midbrain and that encoding is mainly driven by subjective expectations. We conclude that the subcortical auditory pathway is integrated in the cortical network of predictive processing and that expectations are used to optimise the encoding of fast dynamic elements of the acoustic signal.

Keywords: predictive coding, auditory, sensory pathway, frequency modulation, fMRI, auditory midbrain, auditory thalamus, IC, MGB

1. INTRODUCTION

Expectations can have dramatic effects on sensory processing (de Lange et al., 2018). A prime example is speech perception, where word recognition is strongly affected by semantic context (Davis et al., 2011), word prevalence (Sereno et al., 2003), and prior knowledge (Sohoglu et al., 2012). Predictive coding is one of the leading frameworks explaining how expectations affect perceptual encoding (Friston, 2003, 2005; Rao & Ballard, 1999). A key hypothesis of the framework is that sensory neurons at lower levels do not encode the features of the stimuli but prediction error: the difference between the

sensory input and the predictions of an internal generative model of the sensory world.

The encoding of prediction error to fast dynamic stimuli has been robustly demonstrated in the auditory cortex (Blank & Davis, 2016; Blank et al., 2018; Hovsepyan et al., 2020; Signoret et al., 2020; Sohoglu & Davis, 2020; Stein et al., 2022; Vidal et al., 2019; Ylinen et al., 2016). However, anatomical and physiological properties make the subcortical auditory pathway very well suited to test hypotheses on fast dynamic sounds (Giraud et al., 2000; Osman et al., 2018; von Kriegstein et al., 2008): Neural populations in the auditory midbrain (inferior colliculus; IC) and thalamus (medial geniculate body; MGB) are endowed with much shorter time constants and faster access to acoustic information than neural populations in the cerebral cortex (Steadman & Sumner, 2018). Moreover, the nuclei are the target of massive cortico-thalamic and cortico-collicular efferent systems (Lee & Sherman, 2011; Schofield, 2011; Winer, 1984, 2005a) that are propitious to transmit complex predictions.

Stimulus-specific adaptation (SSA) has been used as a first attempt to test for predictive coding in the subcortical pathways. SSA is a phenomenon where individual neurons adapt to repetitions of a pure tone but show recovered responses to a frequency deviant (Ulanovsky et al., 2003). SSA is present in single neurons of the rodent's IC (Ayala et al., 2015; Gao et al., 2014; Parras et al., 2017; Pérez-González et al., 2012; Robinson et al., 2016; Zhao et al., 2011) and MGB (Anderson et al., 2009; Antunes et al., 2010; Bauerle et al., 2011; Parras et al., 2017), and in neural populations of the human IC and MGB (Cacciaglia et al., 2015; Cornella et al., 2015; Escera & Malmierca, 2014; Grimm et al., 2011; Tabas et al., 2020). SSA can, however, be explained by both neural habituation and predictive coding (see Tabas & von Kriegstein, 2021a for a review). In the case of pure tones, we have recently used a novel SSA paradigm which revealed that SSA in human IC and MGB is driven largely by an internal model of the sensory world informed by the subjective expectations of the listeners, as hypothesised by predictive coding but not by neural habituation (Tabas et al., 2020).

In contrast to pure tones, natural sounds comprise highly dynamic elements that cannot be fully characterised by mixtures of pure tones. An ubiquitous example of these dynamic elements are fast frequency-modulated (FM)-sweeps (Liberman & Studdert-Kennedy, 1978; Liberman et al., 1956). While pure tones are encoded according to their frequency along the tonotopic axis already at the basilar membrane (Hu, 2003), FM-sweeps are encoded in FM-direction and FM-rate selective neurons (Kuo & Wu, 2012). In humans, the lowest level in the auditory hierarchy with evidence for fast FM-direction

(Hsieh et al., 2012; Joanisse & DeSouza, 2014) and rate (Okamoto & Kakigi, 2015) selectivity is in auditory cortex; however, FM-sensitive neurons have been reported in the rodent IC and MGB (Issa et al., 2016; Kuo & Wu, 2012; Lui & Mendelson, 2003; Ye et al., 2010; Zhang et al., 2003). Here, we focus on the subcortical auditory pathway, where the encoding of FM has not been shown yet in humans, and where the encoding of FM as prediction error would be the most surprising.

We addressed two key questions. First, whether FM-rate and FM-direction are already encoded in neural populations of the human IC and MGB. Second, whether fast FM-sweeps are encoded in IC and MGB according to the principles of predictive coding; that is, as prediction error with respect to a generative model of the sensory world that incorporates the subjective expectations of the listener. We measured BOLD responses in the IC and MGB while participants listened to sequences of FM-sweeps. We used abstract rules to manipulate the subjective expectations of the participants on the incoming FM-sweeps independently of local stimulus statistics. We reasoned that, if FM-sweeps were encoded according to their objective properties, an FM-sweep embedded in a specific statistical context should elicit the same activation no matter the expectations that participants have on its occurrence. Reversely, if FM-sweeps were encoded according to the principles of predictive coding, BOLD responses should directly depend on how well the sensory input fits the expectations of the listeners.

2. METHODS

This study was approved by the Ethics committee of the Technische Universität Dresden, Germany (ethics approval number EK 315062019). All listeners provided written informed consent and received monetary compensation for their participation.

2.1. Participants

Eighteen German native speakers (12 female), aged 19 to 31 years (mean 24.6), participated in the study. None of them reported a history of psychiatric or neurological disorders, hearing difficulties, or current use of psychoactive medications. Normal hearing abilities were confirmed with pure tone audiometry (250 Hz to 8000 Hz); all participants had hearing threshold equal to or below 15 dB SPL in the frequency range of the stimuli used in the experiment (1000 Hz–3000 Hz). Participants were also screened for dyslexia (German SLRT-II test (Moll & Landerl, 2014), RST-ARR (Ibrahimović & Bulheller, 2013), and rapid automatised naming (RAN) test of letters, numbers, objects, and colours (Denckla & Rudel, 1974))

and autism (Autism Spectrum Quotient; AQ (Baron-Cohen et al., 2001)). All scores were within the neurotypical range (SLRT: $\min\left(\max\left(PR_{words}, PR_{pseudowords}\right)\right) = 21$, higher than the cut-off value of 16, following the same guidelines as Gutschmidt et al. (2021); RST-ARR: all $PR \geq 31$, higher than the cut-off value of 16; RAN: maximum of 3 errors and $RT < 36$ seconds in each of the four categories; AQ: all participants $AQ \leq 31$, under or equal to the cut-off value of 32).

Since we had no estimations of the possible sizes of the effects, we maximised our statistical power by recruiting as many participant as we could fit in the MRI measurement time allocated to the study. This number was fixed to 20 before we started data collection, but 2 participants dropped out of the study during data collection. We maximised the amount of data collected for participant to reduce random error to a minimum and maximise the likelihood of measuring effects at the single-subject level.

2.2. Stimuli

The stimuli were three fast FM-sweeps: One sweep with a fast negative FM-rate (frequency span $\Delta f = -200$ Hz), one with a fast positive FM-rate (frequency span $\Delta f = 200$ Hz), and one with a slow positive FM-rate (frequency span $\Delta f = 100$ Hz; Fig. 1A, B). We used 50 ms long sweeps in the frequency range of $f \sim 1100$ Hz so that they had the typical properties of formant transitions in speech (Liberman & Studdert-Kennedy, 1978). The sweep average frequencies
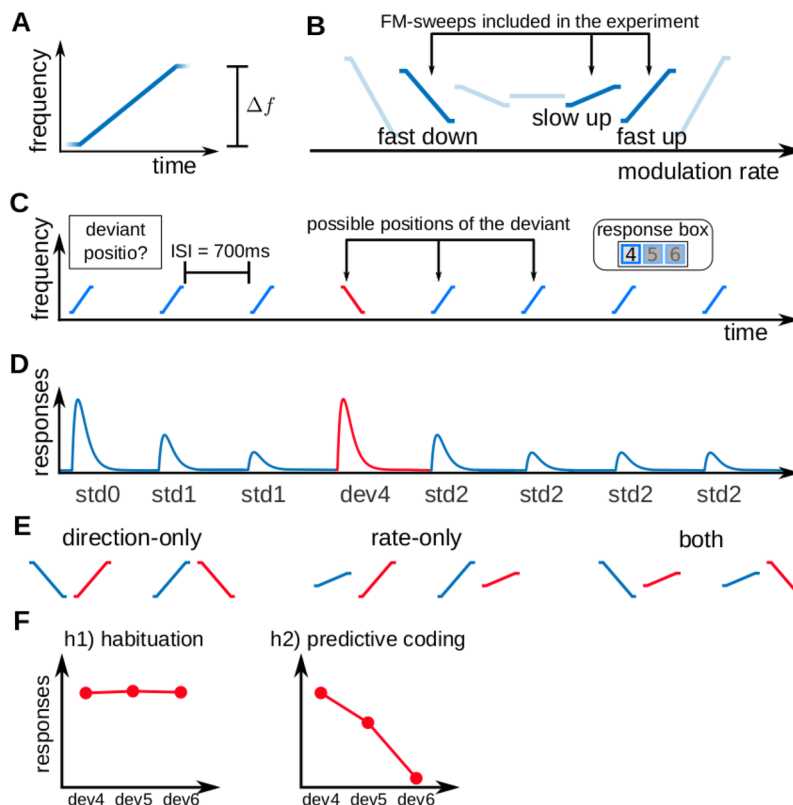

Fig. 1. Experimental design and hypotheses. (A) Example of an FM-sweep with positive FM-rate. (B) The three FM-sweeps used in the experiment (in dark blue) in comparison to an hypothetical family of seven sweeps with increasing modulation rate. All sweeps had the same duration of 50 ms. They were characterised by differences in the frequency span Δf. (C) Example trial. Each trial consisted of a sequence of seven repetitions of one FM-sweep (standards; blue) and one other FM-sweep (deviant; red). In each trial, a single deviant was located in positions 4, 5, or 6 of the sequence. Participants reported, in each trial, the position of the deviant right after they identified it. Each participant completed up to 540 trials in total, 60 per deviant position and FM-sweep combination $\Delta = \left|\Delta f_{deviant} - \Delta f_{standard}\right|$. Sweeps within a sequence were separated by 700 ms inter-stimulus-intervals (ISIs). (D) Schematic view of the expected underlying responses in the auditory pathway for the sequence shown in (C), together with the definition of the experimental variables (std0: first standard; std1: repeated standards preceding the deviant; std2: standards following the deviant; devx: deviant in position x). (E) Schematic view of the six standard (blue) and deviant (red) combinations. Combinations are characterised by whether deviant and standard differ in: modulation direction only, modulation rate only, or both. (F) Expected responses in the auditory pathway nuclei corresponding to the hypotheses: h1) responses reflect adaptation by habituation only; h2) responses reflect prediction error with respect to the participant's expectations.

were adjusted so that all FM-sweeps were perceived as having the same pitch (Nabelek et al., 1970; Tabas & von Kriegstein, 2021b). We used a previous computational model to confirm that the selected FM-sweeps would elicit the same average activity along the tonotopic axis (Tabas & von Kriegstein, 2021b) and thus the same representation (averaged across the 50 ms duration of the stimuli) in the receptive fields encoding pure tones; that is, the receptive fields that were putatively used to differentiate between the stimuli in Tabas et al. (2020). Thus, if the model predictions are correct, participants would needed to engage FM-direction and FM-rate selective circuits to differentiate between any two sweeps in the present paradigm.

All sounds were 50 ms long (including 5 ms in/out ramps) sinusoidal FM sweeps. The frequency sweeps lasted for 40 ms and were preceded and followed by 5 ms long segments of constant frequency that overlapped with the in/out ramps (Fig. 1A). The constant frequency and sweep segments were merged in the frequency space to avoid discontinuities in the stimulus waveforms. The constant frequency segments were used to guarantee that the entire sweep segments, which were controlled to elicit the same pitch percept, were played at the same loudness and audible in the loud noise generated by the MRI scanner. Participants could not have used these segments to characterise the sounds because they were only audible (having a loudness that was comparable or larger than the scanning noise) for about 2 ms, whereas the auditory system needs to integrate along four repetition cycles of the stimuli to characterise pitch (Pressnitzer et al., 2001); that is, for about 6 ms for a ~ 1.5 kHz tone. Moreover, the constant frequency segments would be strongly affected by forward and backward masking (Elliott, 1971) from the louder and longer sweep segment.

We used a total of three sweeps during the experiment: a *fast up* sweep with starting frequency $f_0 = 1000$ Hz and ending frequency $f_1 = 1200$ Hz ($\Delta f = 200$ Hz); a *slow up* sweep with $f_0 = 1070$ Hz and $f_1 = 1170$ Hz ($\Delta f = 100$ Hz), and a fast down sweep with $f_0 = 1280$ Hz and $f_1 = 1080$ Hz ($\Delta f = -200$ Hz). The sweep average frequencies were adjusted so that all FM-sweeps elicited the same average activity along the tonotopic axis and were perceived as having the same pitch (Nabelek et al., 1970; Tabas & von Kriegstein, 2021b); this design guaranteed that FM-direction and FM-rate selective neurons were necessary to differentiate between any two sweeps in the paradigm.

2.3. Experimental paradigm

We arranged the stimuli in sequences of 8 FM-sweeps with 7 repetitions of the same sweep (standard) and one deviating sweep (deviant) (Fig. 1C). Participants were instructed to report, with a button press, the position of the

deviant within the sequence as fast and accurately as possible after identifying the deviant. Each sequence was characterised by the position of the deviant and the combination of sweeps. There were three combinations (Fig. 1E): one where the sweeps differed only on modulation direction, one where the sweeps differed only on modulation rate, and one where the sweeps differed in both.

In each trial of the fMRI experiment, participants listened to one tone sequence and reported, *as fast and accurately as possible* using a button box with three buttons, the position of the deviant (4, 5 or 6). The inter-trial-interval (ITI) was jittered to maximise the efficiency of the response estimation of the deviants (Friston et al., 1999). To do this, we first sampled the time lapsed between deviants (inter-deviant-interval; IDI) from a truncated normal distribution with an average of 5 seconds and a standard deviation of 1 second, truncated between 3 and 11 seconds. We used the deviant position of the current and next trial to compute the corresponding ITI given the sampled IDI, and further constrained the ITI to be of a minimum of 1.5 seconds to ensure that participants were able to tell consecutive trials apart from each other. Participants were allowed to report the deviant position up to 2000 ms after the offset of the last tone; if participants had not responded after the minimum ITI of 1500 ms, this minimum was automatically extended to accommodate the additional waiting period. This construction resulted in the following summary statistics for the IDIs: mean of 7.2 seconds, minimum of 5.2 seconds, maximum of 13.0 seconds; and the following summary statistics for the ITIs: minimum of 1.5 seconds, mean of 4.8 seconds. These summary statistics ignored the periods of silence corresponding to the null trials.

We implicitly modulated the participant's subjective expectations on the incoming stimuli using two abstract rules that were disclosed to the participants: 1) all sequences have a deviant, and 2) the deviant is always located in position 4, 5, or 6. The three deviant positions were used the same number of times along the experiment, so that the three deviant positions were equally likely at the beginning of the sequence. Therefore, the likelihood of finding a deviant in position 4 after hearing 3 standards is 1/3. However, if the deviant is not located in position 4, it must be located in either position 5 or 6, so participant expect a deviant in position 5 after hearing 4 standards with a probability 1/2. The probability of finding a deviant in position 6 after hearing 5 standards is 1.

2.4. Experimental design

Participants completed the task while we measured BOLD responses in participants' IC and MGB with an

fMRI-sequence. All but one participant completed 9 runs of the main experiment across three sessions; participant 18 completed only 8 runs for technical reasons.

Each run contained 6 blocks of 10 trials. The 10 trials in each block used one of the 6 possible sweep combinations, so that all the sequences within each block had the same standard and deviant. Thus, within a block only the position of the deviant was unknown, while the deviant's FM-direction and FM-rate were known in nine of every 10 trials. The order of the blocks within the experiment was randomised. The position of the deviant was pseudorandomised across all trials in each run so that each deviant position happened 180 times per participant but an unknown amount of times per run. This constraint allowed us to keep the same a priori probability for all deviant positions in each block. In addition, there were 23 silent gaps of 5300 ms duration (i.e., null events of the same duration as the tone sequences) randomly located in each run (Friston et al., 1999), which did not necessarily fall at the end or beginning of a block. Each run lasted around 10 minutes, depending on the reaction times of the participant.

Due to an undetected bug in the presentation code, the standard/deviant combination of the trial was incorrectly recorded in some runs. The bug affected the first three runs of participants 1, 2, 4, and 5; and the first six runs of participant 3. This information was not relevant for the analyses that aggregated the data across sweep combinations, and affected only the analyses of Figure 5, where we excluded the affected runs of participants 1, 2, 4, and 5, and participant 3 altogether.

2.5. Functional localiser

We also run a functional localiser that was designed to activate the participant's IC and MGB. Each run of the functional localiser consisted of 20 blocks of 16 seconds and lasted for about 6.5 minutes. Ten of the blocks were silent; the remaining blocks consisted of presentations of 16 contiguous sounds of 1 second duration each. Sounds were taken from a collection of 85 natural sounds collected by Moerel et al. (2015). Participants were instructed to press a key when the same sound was repeated twice, which happened on 5% of the trials. The participants received this task to ensure that they attended the sounds: behavioural data from the functional localiser was not used in the analysis.

2.6. Experiment structure

Each session consisted of three runs of the main experiment, interspersed with two runs of the functional localiser. All runs were separated by breaks of a minimum of 1 min-

ute to allow the participants rest. Fieldmaps and a whole-head EPI were acquired between the third and fourth run. In the first session, we also measured a structural image before the fieldmaps. The first run of the first session was preceded by a *practice run* of four randomly chosen trials to ensure the participants had understood the task. We acquired fMRI during the practice run in order to allow the participants to undertake the training with MRI-noise.

2.7. Data acquisition

MRI data were acquired using a Siemens Trio 3 T scanner (Siemens Healthineers, Erlangen, Germany) with a 32-channel head coil. Functional MRI data were acquired using echo planar imaging (EPI) sequences. We used partial coverage with 24 slices. The volume was oriented in parallel to the superior temporal gyrus such that the slices encompassed the IC, the MGB, and the superior temporal gyrus. In addition, we acquired one volume of an additional whole-head EPI with the same parameters (including the FoV) and 84 slices during resting to aid the coregistration process (see Section 2.8).

The EPI sequence had the following acquisition parameters: TR = 1900 ms, TE = 42 ms, flip angle 66°, matrix size 88×88, FoV 154 mm × 154 mm, voxel size 1.75 mm isotropic, bandwidth per pixel 1386 Hz/px, and interleaved acquisition. During functional MRI data acquisition, cardiac signal was acquired using a scanner pulse oximeter (Siemens Healthineers, Erlangen, Germany).

Structural images were recorded using an MPRAGE (Brant-Zawadzki et al., 1992) T1 protocol with 1 mm isotropic resolution, TE = 1.95 ms, TR = 1000 ms, TI = 880 ms, flip angle 1 = 8°, and FoV = 256 mm × 256 mm.

Stimuli were presented using MATLAB (The Mathworks Inc., Natick, MA, USA) with the Psychophysics Toolbox extensions (Brainard, 1997) and delivered through an Optoacoustics (Optoacoustics Ltd, Or Yehuda, Israel) amplifier and headphones equipped with active noise-cancellation. Loudness was adjusted independently for each participant to a comfortable level before starting the data acquisition.

2.8. Data preprocessing

The preprocessing pipeline was coded in Nipype 1.5.0 (Gorgolewski et al., 2011), and carried out using tools of the Statistical Parametric Mapping toolbox, version 12; Freesurfer, version 6 (Fischl et al., 2002); the FMRIB Software Library, version 5 (FSL) (Jenkinson et al., 2012)); and the Advanced Normalization Tools, version 2.3 (ANTS) (Avants et al., 2011). All data were coregistered to the Montreal Neurological Institute (MNI) MNI152 1 mm isotropic symmetric template.

First, we realigned the functional runs. We used SPM's *FieldMap Toolbox* to calculate the geometric distortions caused in the EPI images due to field inhomogeneities. Next, we used SPM's *Realign and Unwarp* to perform motion and distortion correction on the functional data. Motion artefacts, recorded using SPM's ArtifactDetect, were later added to the design matrix (see Section 2.9).

Next, we used Freesurfer's recon-all routine to calculate the boundaries between grey and white matter (these are necessary to register the functional data to the structural images) and ANTs to compute the transformation between the structural images and the MNI152 symmetric template.

Last, we coregistered the functional data to the structural image with Freesurfer's *BBregister*, using the boundaries between grey and white matter of the structural data and the whole-brain EPI as an intermediate step. Data were analysed in the participant space, and then coregistred to the MNI152 template. Note that, since the resolution of the MNI space (1 mm isotropic) was higher than the resolution of the functional data (1.75 mm isotropic), the transformation resulted in a spatial oversampling.

All the preprocessing parameters, including the smoothing kernel size, were fixed before we started fitting the general linear model (GLM) and remained unchanged during the subsequent steps of the data analysis.

Physiological (heart rate) data were processed by the PhysIO Toolbox (Kasper et al., 2017), that computes the Fourier expansion of each component along time and adds the coefficients as covariates of no interests in the model's design matrix.

2.9.   Estimation of the BOLD responses

First level analyses were coded in Nipype and carried out using SPM. Second-level analyses were carried out using custom code in MATLAB. The coregistered data were first smoothed using a 2 mm FWHM Gaussian kernel with SPM's *Smooth*.

The first-level GLM's design matrix for the main experiment included 6 regressors: first standard (std0), standards before the deviant (std1), standards after the deviant (std2), and deviants in positions 4, 5, and 6 (dev4, dev5, and dev6, respectively; Fig. 1). Conditions std1 and std2 were modelled using linear parametric modulation (O'Doherty et al., 2007), whose linear factors were coded according to the position of the sound within the sequence to account for effects of habituation (Tabas et al., 2020; Supplementary Fig. S1). The first-level GLM's design matrix for the functional localiser included 2 conditions: sound and silence. On top of the main regressors, the design matrix also included the physiological PhysIO and artefact regressors of no-interest. Beta values were

z-scored per run and participant before running group statistics to ensure they all had zero mean and unit variance (Devore, 2008).

This design allowed us to maximally disentangle responses to stimuli that were close to each other in time at the cost of introducing the reasonable assumption that the responses to the repeated standards (*std*1 and *std*2) varied approximately linearly across successive repetitions. The resulting design matrix, convoluted by the hemodynamic response function (Glover, 1999), presents moderate correlations between most pairs of regressors (Supplementary Fig. S2). Although these correlations reduce the statistical power to detect differences in responses to correlated regressors, with over 360 minutes of measured data for the main task per participant, our study is well equipped to compensate for the resulting decrease on statistical power. Moreover, correlation between regressors can under no circumstance result in an increase of type I errors (i.e., false positives) (Mumford et al., 2015); therefore, the measured correlations do not challenge the interpretability of positive results.

Analyses geared towards testing whether responses to different deviants differed were carried out by fitting the regressors across all trials to maximise statistical power. Analyses geared towards testing specific sensitivity to FM-direction or FM-rate were carried out by defining a total of 18 regressors, 6 for each of the three standard/deviant combinations (Fig. 1E).

2.10.   Definition of the anatomical and SSA ROIs

We used a recent anatomical atlas of the subcortical auditory pathway (Sitek et al., 2019) to compute prior regions corresponding to the left IC, right IC, left MGB, and right MGB, respectively. The atlas comprises three different definitions of the ROIs calculated using 1) data from the big brain project, 2) postmortem data, and 3) fMRI in vivo-data. To compute the prior coarse region for each nuclei, we combined the three masks and inflated the resulting regions with a Gaussian kernel with FWHM = 1 mm isotropic. Next, we used SPM to compute the contrast *sound > silence* of the data from the functional localiser. We then masked this contrast with each of the prior coarse regions. Last, we iteratively thresholded the contrast to increasingly conservative higher values until the number of surviving voxels equal the volume of the region reported in Sitek et al. (2019); namely, 146 voxels for each of the ICs, and 152 for each of the MGBs.

The final IC and MGB regions were computed by combining the prior coarse regions with the results from the contrast *sound > silence* of the functional localiser. Within each region, we thresholded the contrast to increasingly higher values until the number of surviving voxels equalled

the volume of the region reported in Sitek et al. (2019); namely, 146 voxels for each of the ICs, and 152 for each of the MGBs.

To address our first research question, whether neural populations of human IC and MGB encode FM-rate and FM-direction, we tested whether these two nuclei show SSA to the FM-sweeps used in the experiment; namely, if neural responses in IC and MGB adapt to repeated FM-sweeps while preserving high responsiveness to FM-sweeps that deviate from the standards in FM-rate or FM-direction (Fig. 1D). Since all sweeps were designed to elicit the same average activation across the tonotopic axis and elicited the same pitch percept, neural populations showing SSA to these FM-sweeps necessarily comprise neurons that are sensitive to FM-rate and FM-direction.

We used the coefficients of the GLM or beta estimates from the first-level analysis to calculate the adaptation and deviant detection ROIs, defined as the sets of voxels within the IC and MGB ROIs that responded significantly to the contrasts $std0 > 0.5\, std1 + 0.5\, std2$ and $dev4 > 0.5\, std1 + 0.5\, std2$, respectively. Significance was defined as $p < 0.05$, family-wise-error (FDR)-corrected for the number of voxels within each of the IC/MGB ROIs. SSA voxels are defined as voxels that show both, adaptation and deviant detection; thus, we calculated an upper bound of the p-value maps for the SSA contrast as the maximum of the uncorrected p-values associated to the adaptation and deviant detection contrasts. The SSA ROIs were calculated by FDR-correcting and thresholding the resulting p-maps at $\alpha = 0.05$. All calculations were performed using custom-made scripts (see Data and Code Availability section).

2.11. Bayesian model comparison

To address our second research question, whether IC and MGB responses encode FM-sweeps as prediction error

with respect to the listener expectations, we used Bayesian model comparison. According to predictive coding, both the responses to deviants and standards should scale with the predictability of the stimuli. Due to the limited temporal resolution of fMRI, we cannot use a classical analysis to robustly estimate the responses to the standards and the deviants simultaneously in each single voxel. However, by introducing reasonable assumptions on the response patterns expected by habituation (Malmierca et al., 2009) and prediction error (Friston, 2003), Bayesian techniques can evaluate whether a voxel is significantly likely to encode prediction error to both, deviants and standards.

We considered two models. The first model assumed that adaptation to repeated fast FM-sweeps was driven by habituation to the stimulus sequence properties, independently of participant's expectations; namely, that neural populations habituate to repetitions of the standard, but show recovered responses to deviant irregardless of their position (habituation hypothesis; Fig. 1F, h1). The second model assumed that adaptation was driven by predictive coding; namely, that neural responses to the deviants encoded prediction error with respect to the expectations of the participants (predictive coding hypothesis; Fig. 1F, h2). Although we expect habituation to also contribute to the BOLD response in this last scenario, we conservatively decided not to include an additional habituation regressor in h2 to limit its explanatory power in voxels that are not driven by prediction error.

The Bayesian analysis of the data consisted as well of first- and second-level analyses. In the first level, we used SPM via nipype to compute the log-evidence in each voxel of each participant for each of the four models (see Fig. 2). The models were described using regressors with parametric modulation whose coefficients corresponded to a simplified view of the expected responses according to each model (Table 1). The expected responses of each
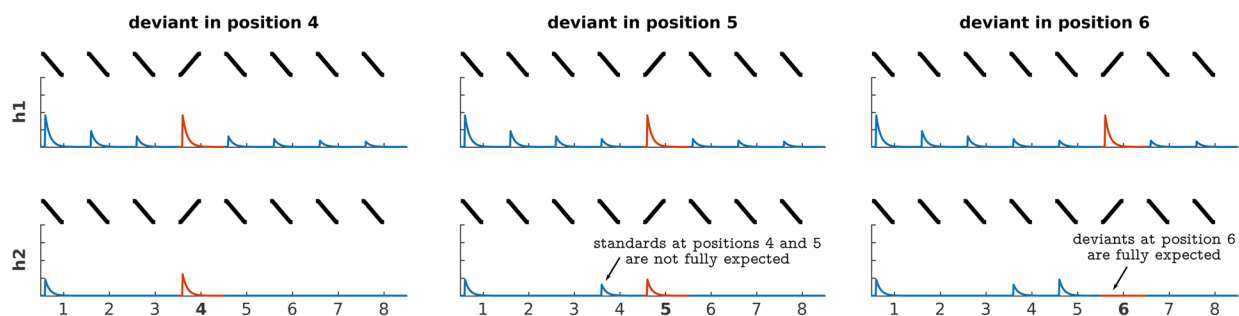

Fig. 2. Design of the Bayesian models. The table shows the parametrised expected response to each tone in the sequence (rows) for the two different models (h1/h2) and the three deviant positions. Each model was defined according to the relative amplitudes it predicted for the different sounds in the sequences. H1 assumed asymptotic habituation to consecutive standards and recovered responses to deviants. H2 assumed that responses to the stimuli depended on how predictable they were. Note that the models have free linear parameters: the displayed amplitudes are one of the many possible solutions of the linear fit. See Table 1 for an exact definition of each model.

Table 1. Amplitudes of the models used for Bayesian Model Comparison.

		1	2	3	4	5	6	7	8
h1	deviant in 4	a_0	a_1	$a_1/2$	a_2	a_3	$a_3/2$	$a_3/3$	$a_3/4$
	deviant in 5	a_0	a_1	$a_1/2$	$a_1/3$	a_2	a_3	$a_3/2$	$a_3/3$
	deviant in 6	a_0	a_1	$a_1/2$	$a_1/3$	$a_1/4$	a_2	a_3	$a_3/2$
h2	deviant in 4	a_0	0	0	$2a_1/3$	0	0	0	0
	deviant in 5	a_0	0	0	$a_1/3$	$a_1/2$	0	0	0
	deviant in 6	a_0	0	0	$a_1/3$	$a_1/2$	0	0	0

H1 assumes an asymptotic decay ($a_0 \propto 1/n$ where n is the position of the stimulus in the sequence) in the responses for all standards, a full response to deviants, and a recovery from the last standard before the deviant and the first standard after the deviant that is sufficient to make the responses to both standards comparable. The model was built as a simplification of the average dynamics reported in the animal literature on SSA (e.g., Malmierca et al., 2009). The free parameters encode the relative decay from the first to the second standard (a_0/a_1), the recovery between the last standard before the deviant and the first standard after the deviant (modulated by a_3), and the recovery of the responses to the deviant (a_2). H2 assumes that the responses scale with predictability ($a_0 = 1 - p$, where p is the likelihood of finding the heard stimuli in each position). The model was built following the precision-weighted formulation of predictive coding (Friston, 2003), which assumes that predictability is a multiplicative factor in the generation of prediction error. We used an additional free parameter to encode the amount of prediction error elicited by the first standard (a_0, which, unlike the rest of the stimuli in the trial, is additionally affected by uncertainty in the time onset). These models include a larger number of free parameters than the ones we used in our previous study (Tabas et al., 2020). The additional parameters allowed us to capture a variety of habituation and prediction error dynamics within the same model, rendering the definitions more general. However, using the more restricted models from Tabas et al. (2020) yields similar results (Supplementary Fig. S2).

model were the same in all trials that had the same standard-deviant combination and deviant position. Given the model amplitude(s) a_n and the timecourse of a voxel y, SPM calculates the log-evidence of the linear model $y = \sum \beta_n a_n + \xi$, where β_n are the linear coefficients of each regressor and ξ are noise terms.

Log-evidence maps for each participant were combined following the random-effects-equivalent procedure described in Rosa et al. (2010) and Stephan et al. (2009) to compute the posterior probability maps associated to each model at the group level. We combined the maps using custom scripts (see Data and Code Availability section). Histograms shown in Figures 7 and 8 are kernel-density estimates computed with the distribution of the posterior probabilities across voxels for each of the SSA ROIs.

2.12. Statistical analysis

All pairwise comparisons reported in the study were evaluated for significance using two-tailed Ranksum tests. Unless stated otherwise, p-values for all analyses that comprised multiple testing were corrected using the Holm-Bonferroni method. A result was deemed statistically significant when the corrected $p < 0.05$.

3. RESULTS

3.1. Behavioural responses

Behavioural results showed an average accuracy over 0.96 to all deviant positions (Fig. 3A). Accuracy was slightly higher for the two more expected deviant positions, but differences between conditions were not significant ($p > 0.1$, uncorrected). Reaction times (Fig. 3B)
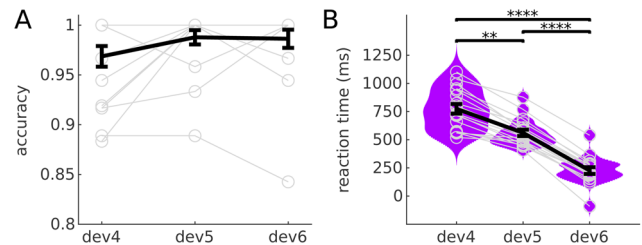

Fig. 3. Performance and reaction times. Mean accuracy (A) and reaction times (B) across deviant positions. Grey circles represent the average value per participant and deviant position. Violin plots are the kernel density estimations of the reaction times for each deviant position. $** \, p < 0.005$, $**** \, p < 0.00005$; all p-values corrected for 3 comparisons.

showed a behavioural benefit of expectations: Participants reacted faster to more expected deviants (average $RT = 770$ ms, 558 ms and 246 ms for deviants at positions 4, 5, and 6, respectively; all differences were significant with $p < 0.0001$, corrected for 3 comparisons).

3.2. Human IC and MGB show stimulus specific adaptation (SSA) to FM-sweeps

We first studied whether the IC and MGB show SSA to fast FM-sweeps to test if the two nuclei are sensitive to FM-rate and FM-direction in humans. To compute SSA, we determined which voxels within the ICs and MGBs adapted to the standard (i.e., adaptation) and recovered responsiveness to deviants (i.e., deviant detection). SSA regions were then defined as the intersection between adaptation and deviant detection regions. ICs and MGBs were identified based on structural MRI data and an independent functional localiser
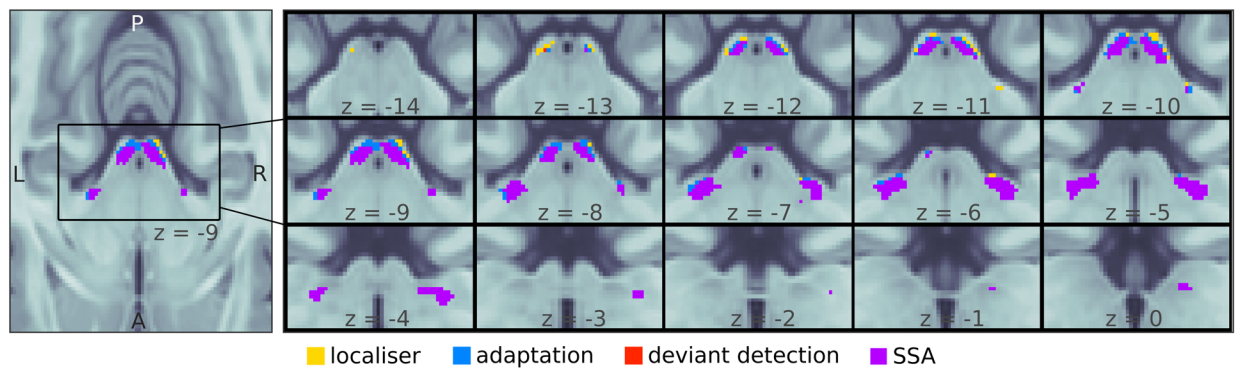
■ localiser   ■ adaptation   ■ deviant detection   ■ SSA

Fig. 4. Mesoscopic stimulus specific adaptation (SSA) in bilateral IC and MGB. Regions within the MGB and IC ROIs adapted to the repeated standards (adaptation; blue shows adaptation only, purple shows SSA, which includes adaptation) and recovered responses to deviants (deviant detection; red shows deviant detection only, purple shows SSA, which includes deviant detection). Stimulus-specific adaptation (i.e., recovered responses to a deviant in voxels showing adaptation; SSA) occurred in bilateral MGB and IC (purple). Maps were computed by thresholding the contrast p-maps at FDR < 0.05. Yellow patches show voxels included in the anatomical masks computed with a functional localiser that showed neither adaptation nor deviant detection.

(see Section 2; IC and MGB ROIs; coloured patches in Fig. 4). Within these ROIs, we used non-parametric rank-sum tests ($N = 18$; one sample per participant) to find which voxels showed significant adaptation to repeated standards (contrast $std0 > 0.5\ std1 + 0.5\ std2$). The associated p-maps were thresholded so that the false-discovery-rate FDR < 0.05. Surviving voxels constituted the *adaptation* ROIs (blue and purple patches in Fig. 4). The same procedure was used to delimit the *deviant detection* ROIs (red and purple patches in Fig. 4): the set of voxels within each anatomical ROI that responded significantly stronger to deviants than to repeated standards (contrast $dev4 > 0.5\ std1 + 0.5\ std2$; note that we compare the responses to the repeated standards with $dev4$ as this is the deviant position for which participants have the lowest expectation). The four anatomical ROIs showed significant adaptation (peak $p \leq 0.0001$) and deviant detection (peak $p < 0.0001$; cluster size, exact peak p-values, and MNI coordinates are shown in Table 2; all p-values corrected for four comparisons).

SSA regions were computed combining the unthresholded *adaptation* and *deviant detection* p-maps. The uncorrected p-value associated to SSA for a given voxel was $p_{SSA} = \max\left(p_{adaptation}, p_{deviant\ detection}\right)$. SSA p-maps where thresholded to $FDR < 0.05$ to compute the SSA ROIs (Fig. 4, purple). The four anatomical ROIs had extensive SSA regions (cluster sizes larger than 90 mm^3; peak $p \leq 0.0003$; exact peak p-values and MNI coordinates are shown in Table 2; all p-values corrected for four comparisons).

Significant SSA was also found in at least one of the nuclei of 15 of the 18 participants ($p \leq 0.048$ for each of the 15 participants, corrected for the 596 voxels included in a global subcortical auditory ROI that comprised bilat-eral IC and MGB), but not all participants showed significant SSA in all ROIs (IC-L: 8 participants, $p \leq 0.049$; IC-R, MGB-L, MGB-R: 6 participants each, with $p \leq 0.048$; all p-values corrected for the number of voxels in the ROI and further corrected for four ROIs).

These results confirmed that there are extensive regions of bilateral IC and MGB that selectively habituate, and therefore are sensitive, to FM direction and rate.

3.3. Human IC and MGB are sensitive to FM-direction and FM-rate

In the next step, we specifically tested whether the IC and MGB are similarly sensitive to FM-rate and FM-direction. To do that, we analysed the regressor fits corresponding to: 1) trials where the standard and deviant differed only in modulation direction but not in absolute modulation rate; and 2) trials where the standard and deviant differed only in modulation rate but not in direction. If IC and MGB encode direction and rate, we would expect similar results in both partitions of the data. Conversely, if human IC and MGB are only sensitive to one of the two properties, we would expect null effects in the partition of the data where the standard and deviants differ in the other property.

Results were similar in both partitions of the data (Fig. 5), demonstrating that the human IC and MGB encode both FM-direction and FM-rate.

We further corroborated that the levels of SSA were comparable for both types of FM changes at the single-subject level. In order to characterise FM-sensitivity with a number for each subject and FM-sweep combination, we used the SSA index (Ulanovsky et al., 2003) SI (Eq. (1); note that $SI > 0$ is equivalent to the deviant detection contrast used in Fig. 4).

Table 2. Statistics and MNI coordinates of the adaptation and deviant detection contrasts in the four regions of interest.

Contrast	ROI	Cluster size	MNI coordinates (mm)	Peak-level p-value
adaptation	left IC	130 voxels	$[-4, -35, -9]$	$p = 1 \times 10^{-4}$
	right IC	124 voxels	$[4, -35, -9]$	$p = 8 \times 10^{-5}$
	left MGB	152 voxels	$[-14, -25, -7]$	$p = 8 \times 10^{-5}$
	right MGB	146 voxels	$[14, -26, -6]$	$p = 1 \times 10^{-4}$
deviant detection	left IC	92 voxels	$[-6, -33, -10]$	$p = 9 \times 10^{-5}$
	right IC	91 voxels	$[6, -33, -8]$	$p = 7 \times 10^{-5}$
	left MGB	136 voxels	$[-14, -24, -7]$	$p = 5 \times 10^{-5}$
	right MGB	140 voxels	$[11, -27, -5]$	$p = 2 \times 10^{-5}$
SSA	left IC	91 voxels	$[-4, -35, -9]$	$p = 3 \times 10^{-4}$
	right IC	91 voxels	$[6, -33, -9]$	$p = 2 \times 10^{-4}$
	left MGB	136 voxels	$[-14, -25, -7]$	$p = 2 \times 10^{-4}$
	right MGB	140 voxels	$[12, -26, -5]$	$p = 1 \times 10^{-4}$

All p-values FDR-corrected for the number of voxels in each anatomical ROI and further corrected for 4 comparisons within each contrast.
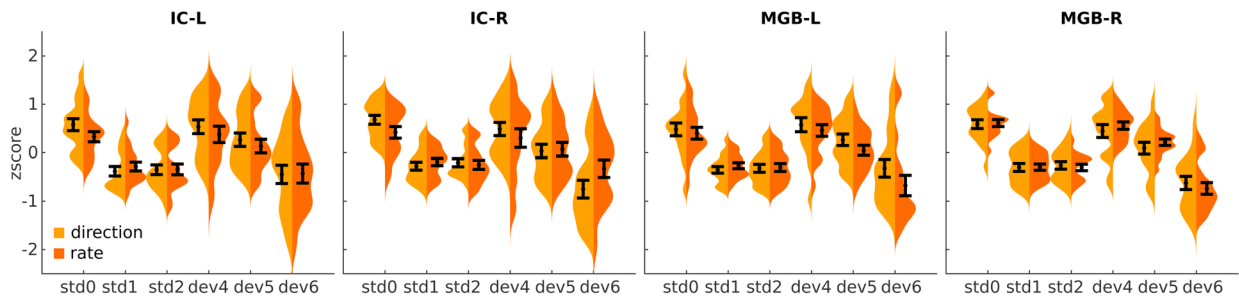

Fig. 5. Summary BOLD responses for partitions of the data where deviant and standard differed only in direction or rate. Average z-score in each of the four SSA ROIs to the different experimental conditions in trials where the standard and deviant differed only in direction (orange) or rate (yellow). Violin plots are kernel density estimations of the distribution of z-scores, averaged over voxels and runs of each ROI. Each distribution holds 17 samples, one per participant (one participant was excluded from this analysis because there were not enough trials available, see Section 2 for details). Black error bars show the mean and standard error of the distributions.

$$SI = \frac{dev4 - \frac{1}{2}(std1 + std2)}{dev4 + \frac{1}{2}(std1 + std2)} \qquad (1)$$

We measured the difference in SI to FM-direction (SI_{dir}) and FM-rate (SI_{rate}) in the voxels of the subject-specific SSA regions calculated in the previous section for each of the 15 subjects for which we obtained significant SSA. If FM-direction and FM-rate are both encoded in IC and MGB, we would expect no difference between these two partitions of the data. We measured the difference using Cohen's $d = (\langle SI_{dir} \rangle - \langle SI_{rate} \rangle) / \sigma$, where $\langle SI \rangle$ is the average of SI and σ is the pooled standard deviation. The difference ranged between $d \geq -0.33$ and $d \leq 0.475$ across participants. The expected value of the difference ($E[d] = 0.02 \pm 0.05$) overlapped with zero, indicating once again that both FM-direction and FM-rate are already encoded in the subcortical auditory pathway.

3.4. Expectations drive the encoding of FM-sweeps in IC and MGB

To address our second question, we evaluated whether the average pooled BOLD responses to deviants in the three different positions were affected by participant's subjective expectations within the SSA regions. In congruence with the predictive coding hypothesis (Fig. 1F, h2), the response profile showed reduced responses for more expected deviants (Fig. 6). This pattern was systematically reproduced in all subjects (Supplementary Fig. S4).

Formal statistical testing confirmed that responses to different deviant positions were different in all ROIs for all contrasts among deviant positions: $dev4 \neq dev5$ ($|d| \geq 0.99$ and $p < 0.006$), $dev4 \neq dev6$ ($|d| \geq 2.39$ and $p < 0.00005$), and $dev5 \neq dev6$ ($|d| \geq 1.74$ and $p < 0.0003$; all p-values corrected for 3×4 12 comparisons). Exact p-values and effect sizes are listed in Table 3. All statistical tests included one sample per participant, ROI, and deviant position.
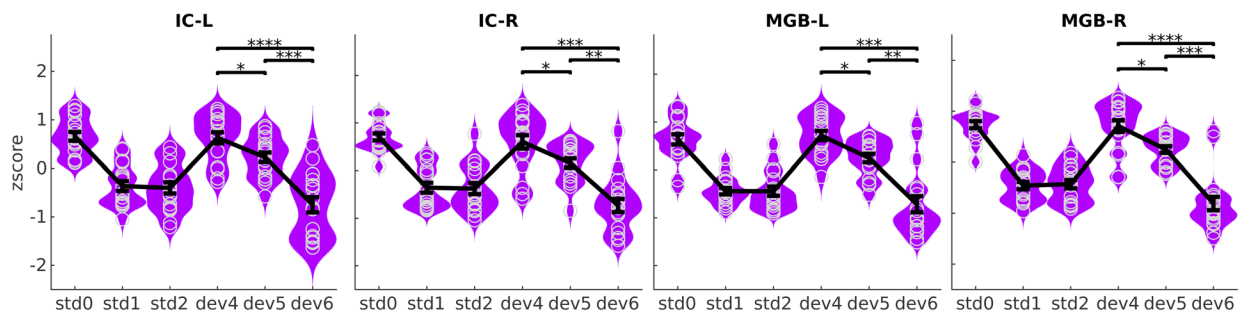
Fig. 6. Summary BOLD responses. Average z-score in each of the four SSA ROIs to the different regressors. Violin plots are kernel density estimations of the distribution of z-scores, averaged over voxels and runs of each ROI. Each distribution holds 18 samples, one per participant. Black error bars show the mean and standard error of the distributions. Significance bars were computed by pooling across standard-deviant combinations. Single-subject distributions are shown in Supplementary Figure S4. *Std*0, first standard; *std*1: standards preceding the deviant; *std*2: standards following the deviant; *dev*4, *dev*5, and *dev*6: deviants at positions 4, 5, and 6, respectively (Fig. 1D). *$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$, ****$p < 0.00005$; all p-values corrected for 12 comparisons.

Table 3. Statistics of the average BOLD response differences between deviant positions.

	dev5		dev6	
IC-L				
dev4	$d = 0.89$	$p = 0.017$	$d = 2.35$	$p = 4 \times 10^{-5}$
dev5			$d = 1.75$	$p = 4 \times 10^{-4}$
IC-R				
dev4	$d = 0.86$	$p = 0.022$	$d = 2.24$	$p = 10^{-4}$
dev5			$d = 1.74$	$p = 5 \times 10^{-4}$
MGB-L				
dev4	$d = 1.20$	$p = 0.0075$	$d = 2.46$	$p = 10^{-4}$
dev5			$d = 1.68$	$p = 0.0015$
MGB-R				
dev4	$d = 1.17$	$p = 0.0073$	$d = 2.91$	$p = 2 \times 10^{-5}$
dev5			$d = 2.40$	$p = 3 \times 10^{-4}$

Effect size is expressed as Cohen's d. Statistical significance was evaluated with two-tailed Ranksum tests between the distributions of the mean response in each ROI across participants ($N = 18$), pooling across standard-deviant combinations. All p-values in the table are corrected for $3 \times 4 = 12$ comparisons.

To corroborate that differences were present at the single-subject level, we run a correlation analysis for each of the 15 participants for which we obtained significant SSA. In each participant, we computed the Pearson's correlation between the BOLD responses elicited by each deviant position with its likelihood of occurrence (namely, 1/3 for deviant 4, 1/2 for deviant 5, and 1 for deviant 6). If BOLD responses reflect prediction error, we would expect a negative correlation between the likelihood and the responses. We found significantly negative correlations in all 15 participants ($\rho \in [-0.87, -0.42]$, all $p < 0.03$; all Pearson tests had $9 \times 3 = 27$ samples, 3 per run).

3.5. FM-sweeps are encoded as prediction error in the majority of IC and MGB voxels

We used Bayesian model comparison to formally evaluate whether the responses in each voxel of the IC and MGB ROIs were best explained as prediction error. This approach provides for a quantitative assessment of the likelihood that each of the two hypotheses (Fig. 1F) can explain the responses in each voxel. This analysis is sensitive to possible region-specific effects that could have been averaged out when aggregating the z-scores across voxels in each ROI.

Following the methodology described in Rosa et al. (2010) and Stephan et al. (2009), we first calculated the log-likelihood of each model in each voxel of the two ICs and MGBs in each participant. Each model yields different predictions on the relative amplitudes to different positions in the sequences (Fig. 1F). We tested h1 and h2 to adjudicate between the habituation and predictive coding explanations of the responses. H1 assumed an asymptotic decay of the standards and recovered responses to the deviants; h2 assumed that the responses to both deviants and standards would depend on the participant's expectations (Fig. 2; for exact values, see Section 2). Participant-specific log-likelihoods were used to compute the Bayes' factor K (i.e., the ratio of the posterior likelihoods) between h1 and h2.

H2 was the best explanation for the data in the majority of voxels of the four ROIs (Fig. 7): h2 was more likely than h1 in all voxels of the left and right IC, and in 85% and 61% of the voxels of the left and right MGB, respectively (see also results from an alternative BMC analysis based on Tabas et al. (2020) in Supplementary Fig. S3).
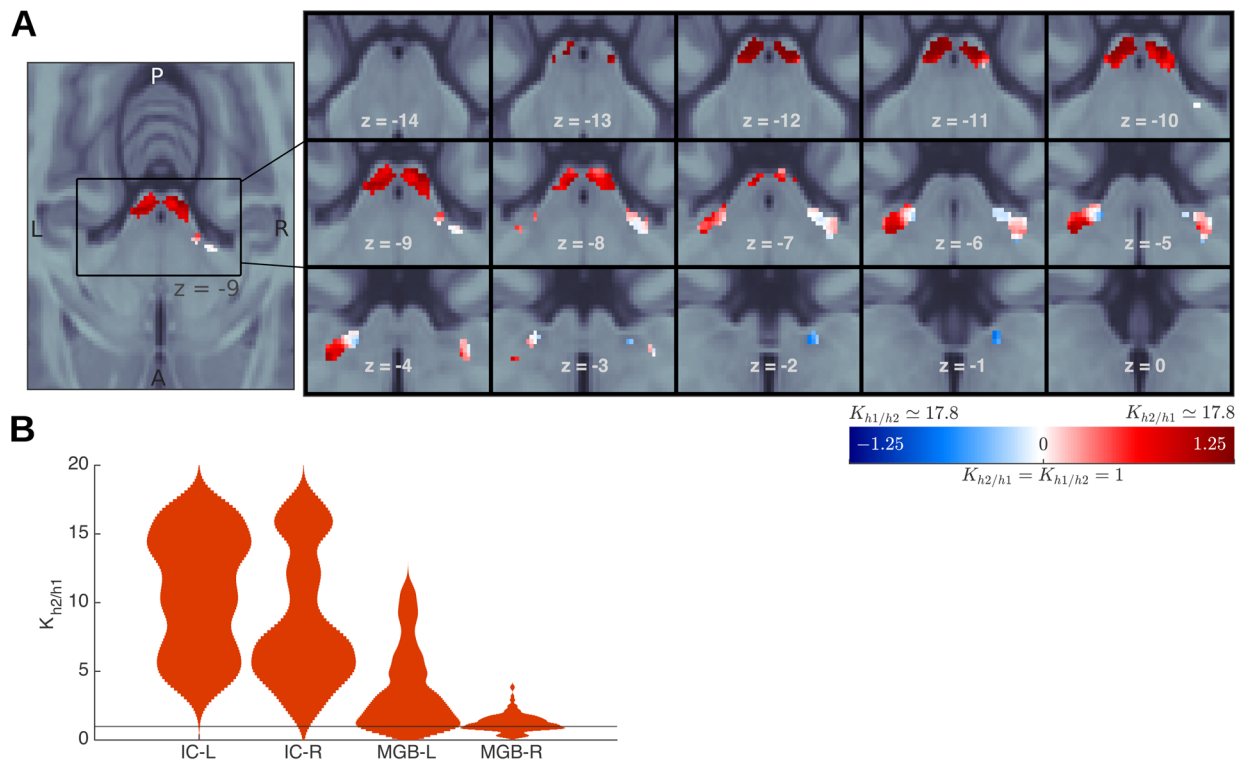
Fig. 7. Bayesian model comparison. (A) Bayes' factor K between h2 (predictive coding) and h1 (habituation) in each of the voxels of the subcortical ROIs in a logarithmic scale. Voxels with negative $\log K$ values ($K<1$; blue) are best explained by h1; voxels with positive $\log K$ values ($K>1$; red) are best explained by h2. Single-subject K factors are plotted in Supplementary Figure S5. (B) Kernel-density estimations of the distribution of K for the model comparison h2/h1 across voxels (i.e., one sample per voxel). See Supplementary Figure S3 for a replication of these results, obtained using the mathematical definitions of the models of Tabas et al. (2020).

To test whether the effect was present at the single-subject level, we computed K independently for each subject in the subject-specific bilateral IC and MGB. Since we performed the group analyses over the entire ROIs (and not only the SSA regions), here we used the full anatomical ROIs of each participant. We measured for how many voxels within each participant h2 was the better explanation of the data (see full results in Supplementary Fig. S5). In 15 of the 18 participants (all but subjects 3, 5, and 18), there were more voxels for which $K > \sqrt{10}$ than for which $K < 1/\sqrt{10}$; namely, more voxels for which there was substantial evidence in favour of h2 than for h1. These single-subject level results confirmed that responses do not simply habituate to successive repetitions of the tone but that, as hypothesised by predictive coding, they are also strongly affected by the subjective expectations of the listeners.

3.6. FM-sweeps are encoded as prediction error in primary and secondary MGB

The auditory pathway is anatomically subdivided into two sections: the primary (lemniscal) or secondary (non-lemniscal) pathways. The primary pathway is characterised by neurons that carry auditory information with high fidelity and it is generally regarded as responsible for the transmission of bottom-up sensory input (Hu, 2003). The secondary pathway has wider tuning curves and it is generally regarded as responsible for the integration of contextual and multisensory information (Hu, 2003).

Both IC and MGB comprise regions that participate in both, the primary and secondary pathways (Hu, 2003). The primary subdivision of the IC is its central nucleus, while the cortices constitute the secondary subdivisions. The primary subdivision of the MGB is its ventral section, while the medial and dorsal sections constitute the secondary subdivisions.

In rodents, SSA and prediction error to pure tones are significantly stronger in secondary subdivisions (e.g., Parras et al., 2017). In humans, prediction error is similarly strong in primary and secondary MGB for pure tones (Tabas et al., 2020). Here, we test for differential representations of prediction error to FM-sweeps in MGB.

Distinguishing between the primary and secondary subsection of the IC and MGB non-invasively is technically challenging (Moerel et al., 2015). A recent study (Mihai et al., 2019) distinguished two distinct tonotopic gradients of the MGB. The ventral tonotopic gradient was identified as the ventral or primary (vMGB) subsection of
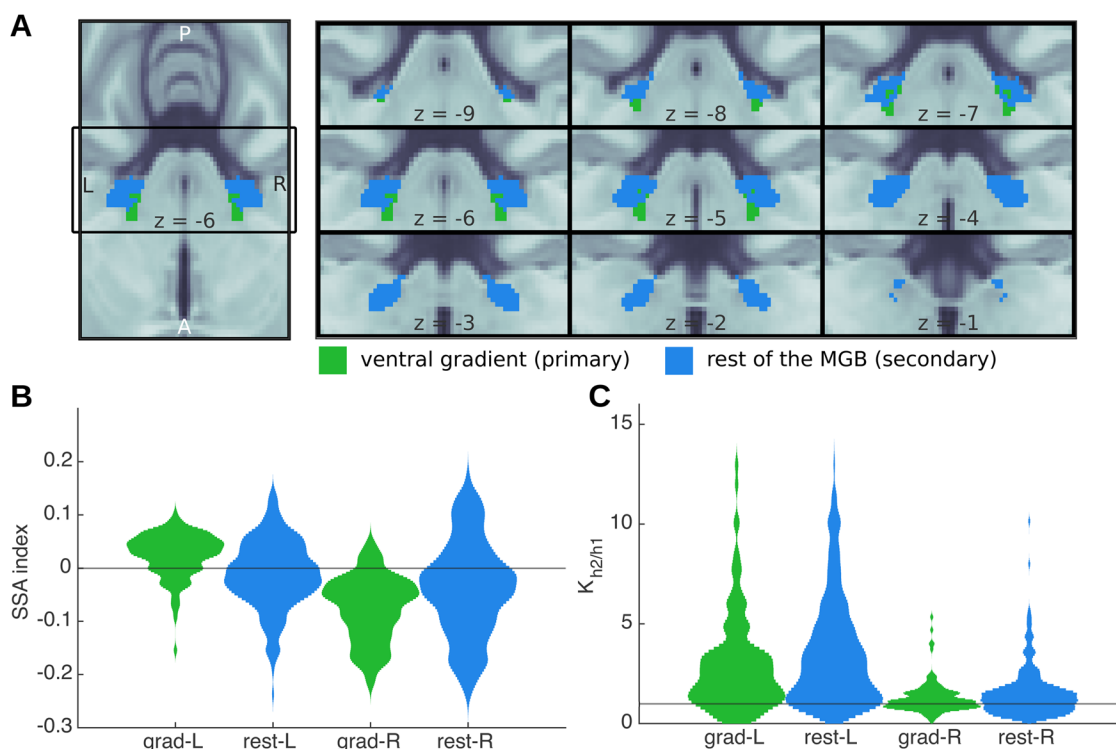
Fig. 8. Analyses of BOLD responses in ventral MGB. (A) Masks from Mihai et al. (2019) of the ventral MGBs (green); blue indicates the remainder of the anatomical MGB ROIs. (B) The distribution of the SSA index SI across each of the two subdivisions of the MGB ROIs; $SI > 0$ is usually interpreted as SSA in the animal literature (Ulanovsky et al., 2003). (C) Histograms showing Bayes' factor K for the comparison between h2 and h1 (Fig. 1F) in each of the subdivisions. No systematic functional differences are apparent between primary and secondary MGB.

the MGB (see Fig. 8A, green). Although the parcellation is based only on the topography of the tonotopic axes and their anatomical location, the region is the best approximation to-date of the vMGB in humans. No parcellation of the IC is available to-date.

Both primary and secondary subdivisions of bilateral MGB showed SSA. SSA strength was measured in each voxel using the SSA index (Eq. (1)). Distributions of the SI across the voxels of each of the subdivisions were comparable in both hemispheres (Fig. 8B), demonstrating that SSA is not confined to nor stronger in the secondary MGB.

Predictive coding (h2) was the best explanation for the responses in 84% and 87% in the two subdivisions of the left MGB, and in 58% and 64% of the primary and secondary subdivisions of the right MGB, demonstrating the encoding of prediction error to FM-sweeps in both, primary and secondary subdivisions of bilateral MGB. Moreover, the distributions of the Bayes' factor K between the predictive coding (h2) and adaptation (h1) hypotheses were comparable across subdivisions (Fig. 8C).

3.7. Prediction error to FM-sweeps and pure tones has similar topographic distributions in the IC

To study whether the same neural populations are in charge of encoding prediction error to FM-sweeps and

pure tones, we compared the topographic distribution of the Bayes' factor K between the h2 and h1 in our data with the topographic distribution of the Bayes' factor K we obtained in a previous experiment, where we measured BOLD responses to the same experimental paradigm as here but using pure tones (Tabas et al., 2020). We computed the correlation between both K across voxels of each of the four ROIs, as defined by the anatomical atlas from Sitek et al. (2019). To ensure that the analyses were comparable across studies, we ran a second Bayesian Model Comparison analysis on the current data using the same model definitions as in Tabas et al. (2020) (see Supplementary Fig. S3).

Distribution of K to both families of stimuli was strongly correlated across voxels of bilateral IC (left, $\rho = 0.47$, $p = 4 \times 10^{-9}$; right, $\rho = 0.34$, $p = 3 \times 10^{-5}$; p-values corrected for 4 comparisons), but not across voxels of the MGBs (left, $\rho = -0.11$, $p = 0.22$; right, $\rho = 0.15$, $p = 0.1$; uncorrected p-values). These results indicate that the topographic distributions of prediction error responses are similar across stimulus modalities.

4. DISCUSSION

The effects of expectations on sensory processing are readily evident in our daily lives. However, the neural

mechanisms underlying the integration of expectations at early stages of the acoustic processing pipeline are poorly understood. Here, we have investigated how fast FM-sweeps, an important dynamic component of natural sounds, are encoded in the human subcortical auditory pathway, and how the subjective expectations of the listener influence their processing. Our study provided four main findings: first, we showed that the human IC and MGB comprise FM-direction and FM-rate selective neuronal populations. Second, we showed that responses in IC and MGB were driven by subjective expectations of the participants, demonstrating that the IC and MGB are integrated in a global network of predictive coding. The findings were robust and present at the single-subject level, demonstrating the generalisation power of the result. Third, we showed that the expectations determined the responses to FM-sweeps in primary and secondary subdivisions of bilateral MGB. Last, we showed that the topographic distribution of neural populations encoding the FM-sweeps as prediction error was similar to that of pure tones in the IC.

Combined, our results provide first demonstration that the human IC and MGB are actively engaged in the predictive processing of dynamic stimuli that transcend the tonotopic static representation: fast FM-sweeps. This confirms the long-standing hypothesis that predictive coding combines high-level expectations with the exquisite temporal properties of the subcortical auditory pathway to promote the encoding of dynamic low-level features (von Kriegstein et al., 2008; Yildiz et al., 2013). This mechanism might be responsible for boosting encoding efficiency and aiding, for example, speech recognition.

Neurons that respond selectively to FM-direction and FM-rate have been located in rodents in the IC (Geis & Borst, 2013; Hage & Ehret, 2003; Li et al., 2010), MGB (Kuo & Wu, 2012; Lui & Mendelson, 2003), and auditory cortex (Issa et al., 2016; Trujillo et al., 2013; Ye et al., 2010; Zhang et al., 2003). In contrast, FM-selectivity has been reported in humans in auditory cortex (Okamoto & Kakigi, 2015) or higher-order areas of the cerebral cortex (Hsieh et al., 2012; Joanisse & DeSouza, 2014). One previous study (Chandrasekaran et al., 2012) showed that auditory training improves encoding of rising FM-sweeps in the IC as measured by the frequency-following-response (Coffey et al., 2019), supporting the active involvement of the IC in the processing of FM. Here, we have established that neural populations in the human IC and MGB show SSA to FM-direction and FM-rate; since our FM-sweeps were matched in duration, pitch, and expected elicited activity along the tonotopic axis, our results extend previous findings providing first evidence for FM-direction and FM-rate selectivity in the human subcortical auditory pathway.

Animal studies have extensively shown that the SSA index to pure tones in IC and MGB increases with increasing rarity and frequency difference of the deviant with respect to the standard (Anderson et al., 2009; Antunes & Malmierca, 2011; Antunes et al., 2010; Ayala et al., 2013, 2015; Duque & Malmierca, 2015; Duque et al., 2016; Malmierca et al., 2009; Zhao et al., 2011). These studies implicitly assume that sensory neurons form expectations based on the local statistics of the stimuli. This form of predictive coding, which we call *local* (Tabas & von Kriegstein, 2021a), is difficult to disambiguate from passive effects of neural habituation: Modelling studies have demonstrated that identical phenomenology can be produced by synaptic fatigue without the need of maintaining additional internal generative models (Eytan et al., 2003; Mill et al., 2011, 2012). Manipulating expectations orthogonally to stimulus regularities is the only way to assess if prediction error is computed with respect to a global model of the sensory world (Tabas & von Kriegstein, 2021a).

To date, the only evidence (see Tabas & von Kriegstein, 2021a for a review) that subcortical nuclei encode stimuli according to subjective expectations independently of stimulus regularities was provided by our previous study on pure tones in human IC and MGB (Tabas et al., 2020). Here, we used fast FM-sweeps that were explicitly designed to elicit the same activation across the tonotopic axis (Tabas & von Kriegstein, 2021b) to ensure that participants had to make use of FM-direction and FM-rate selective neurons to differentiate the deviant from the standards. The current findings demonstrate that the same principles apply to the encoding of dynamic FM-sweeps.

Our results also showed that the topographic distribution of voxels encoding pure tones and FM-sweeps according to the principles of predictive coding was highly correlated in the IC, but not in the MGB. This divergence might indicate a different functional role of the IC and the MGB with respect to both families of stimuli; however, it might also be caused by a greater variability in the anatomical location and orientation of the MGB across subjects (Moerel et al., 2015) and should be considered with caution.

The expectations induced by our paradigm are still far from the complexity of the predictive system putatively in charge of the processing of natural complex signals like speech. However, we speculate that an integrated inverted hierarchy could propagate linguistic predictions to the representational level of formant transitions (Friston et al., 2021; Tabas & von Kriegstein, 2021a; von Kriegstein et al., 2008), and use these predictions to compute prediction error in the IC and MGB.

The expectations induced by our paradigm are most likely generated in the cerebral cortex. However, since we

optimised our paradigm to study prediction error rather than the generation of expectations, we cannot test whether the subcortical responses we measured are driven or not by corticofugal projections. This possibility would be consistent with the massive corticofugal connections from cerebral cortex to MGB and IC (Winer, 1984, 2005b), and with results from animal studies where the deactivation of unilateral auditory cortex (Bauerle et al., 2011) or the thalamic reticular nucleus (Yu et al., 2009) led to reduction of SSA in the ventral MGB (but also see contradictory findings in non-lemniscal MGB (Antunes & Malmierca, 2011) and non-lemniscal IC (Anderson & Malmierca, 2013)).

The present and previous (Tabas et al., 2020) results demonstrate that the IC and the MGB encode auditory stimuli according to subjective expectations when the sensory signal is relevant for the listener's task. This encoding strategy might 1) be general to sensory processing or specific to the processing of task-relevant stimuli; and 2) be particular of processing under abstract expectations that are explicitly known by the listeners or general to any expectation that could be inferred from exposure to the sensory input. Previous studies showed that the IC and the MGB adapt to stimulus regularities even in the absence of a task (e.g., Cacciaglia et al., 2015). Whether abstract regularities also affect the encoding of task-irrelevant stimuli in the subcortical pathway is still an open question.

Despite the fact that the MGB is at a higher processing stage than the IC, we found similar prevalence of the predictive coding Bayesian model (h2) in both nuclei for FM-sweeps (Fig. 7) as well as for pure tones (Tabas et al., 2020). These results contrast with a study in rodents, concluding that the MGB encodes prediction error more strongly than the IC (Parras et al., 2017). We speculate that this fundamental difference is caused by the introduction of abstract rules in our paradigm. Since prediction error depends only on the local representation and the predictions (Friston, 2003), there is no reason for prediction error to vary across hierarchical stages that receive the same set of predictions and have comparable representation of the stimuli, as it is the case for FM-sweeps in IC and MGB (Kuo & Wu, 2012) (and also for pure tones (Hu, 2003)). Rodent studies use passive listening tasks where expectations are induced by repetition. Without an explicit high-level model, prediction error can only be computed with respect to local models that may vary in complexity across processing stages. A task involving stimuli that are represented differently in IC and MGB should shed light on the hierarchical role played by each of the two stages.

Previous studies on subcortical SSA rested almost exclusively on pure tones (Carbajal & Malmierca, 2018;

Malmierca et al., 2015; Tabas & von Kriegstein, 2021a). Only three studies considered whether SSA generalised to other acoustic properties. Thomas et al. (2012) reported SSA to FM-rate in the IC of the big brown bat; however, since the authors used stimuli in the rate range of echolocation signals, it was unclear whether this behaviour would generalise to auditory FM. Gao et al. (2014) measured SSA using ramped and damped broadband noises in the IC, demonstrating that neurons in the IC adapt to intensity modulation. Last, Duque et al. (2016) measured SSA to intensity, and showed that neurons in the IC do not adapt to nominal loudness. Our findings complement these results showing that the human IC and MGB adapt to fast FM without loudness or spectral changes, and provides first evidence for SSA to acoustic properties other than pitch and loudness in the subcortical pathways.

We have argued that the response pattern in MGB and IC (Fig. 6) can be interpreted as the encoding of prediction error with respect to the subjective expectations of the participants. There are two conceivable alternative interpretations for the results: that responses are driven by attention-driven gain modulation, and that responses are driven by general habituation to auditory stimuli. The former view interprets the higher responses to $dev4$ and $dev5$ as the result of a stronger attention of the participant to these positions, which are relevant to the task, and the lower responses to $dev6$, which is fully expected, as the results of a reduction of attention. Previous fMRI studies have indeed shown that attended stimuli elicited higher BOLD responses in auditory cortex (Lee et al., 2014; Paltoglou et al., 2011), and to a much weaker extent also in the IC (Riecke et al., 2018; Rinne et al., 2007, 2008; Varghese et al., 2015). However, this interpretation of our results is unsatisfactory because, first, we observe statistically significant differences in the BOLD responses to $dev4$ and $dev5$, although they are both equally relevant for the task. Second, we observed no systematic differences between responses to $dev6$ and $std2$, whereas in a previous fMRI study deviants always elicited statistically significantly higher responses than standards (Cacciaglia et al., 2015). This was the case although the study had lower statistical power in comparison to our study and used passive stimulation. Only by interpreting the BOLD responses in Figure 6 as prediction error with respect to the participant's expectations we can explain the similar responses to $dev6$ and $std2$ in our paradigm.

The other conceivable interpretation is that the response pattern found in MGB and IC (Fig. 6) is driven by a kind of habituation that partially generalises to tones of other frequencies. This kind of *general* habituation has been reported in the human auditory cortex (e.g., Rosburg

et al., 2002). Since *dev*6 is preceded by five standards, it is plausible that general habituation incurs into lower responses to *dev*6 than to *dev*5 or *dev*4, which are preceded by four and three standards, respectively. However, this interpretation of our results is also unsatisfactory. First, the effect size of the reduction of the responses to *dev*5 with respect to *dev*4 is $d \simeq 1$, and the effect of the response reduction from *dev*5 to *dev*6 is even stronger. If one more repetition of a standard was responsible for such a large reduction of the responses, we would expect the responses to *dev*4, which is preceded by three standards, to be much smaller than the responses to the first standard. However, we observe similar responses to *dev*4 and *std*0. Second, if there were strong general habituation effects capable of inducing a significant decrease in the BOLD responses in tones preceded by more than three standards, we would expect the responses to the standard preceding the deviant (*std*1) to elicit stronger responses than the standards following the deviant (*std*2). However, the results in Figure 6 show no systematic differences in the responses to *std*1 and *std*2.

It has been previously suggested that prediction error may be encoded exclusively in the non-lemniscal or secondary subdivisions of the IC and MGB (Ayala et al., 2015; Malmierca et al., 2015; Parras et al., 2017). In agreement with this hypothesis, SSA is stronger in secondary subdivisions of the rodent's IC (Ayala & Malmierca, 2018; Ayala et al., 2015; Duque et al., 2014; Gao et al., 2014; Pérez-González et al., 2012) and MGB (Antunes & Malmierca, 2011; Antunes et al., 2010; Duque et al., 2014).

In contrast, our results indicated an apparent lack of specialisation across subdivisions of the MGB during the encoding of FM-sweeps: both subdivisions were similarly responsive to FM, and they both encoded FM as prediction error. Similar results were apparent in our previous study when we investigated the encoding of pure tones (Tabas et al., 2020). This lack of specialisation would fit with the idea that expectations are used in the subcortical pathways to aid encoding: to optimise the resources of the subcortical stations requires to make use of the narrow receptive fields of the primary subdivisions (Hu, 2003).

The fundamental difference between our results and the findings in animals might stem from a number of reasons. First, our design involved an active task: lemniscal pathways might only be strongly modulated by predictions when they carry behaviourally relevant sensory information. Second, the modulation of the subcortical auditory pathway might be fundamentally different in humans compared to other mammals, as they have to accomplish processing of such complex and dynamic signals as speech. Last, given the strength of the SSA effects reported in this study, it is possible that regions with weak SSA might have been contaminated with signal stemming from areas with strong SSA due to smoothing and interpolation necessary for the analysis of fMRI data.

Given the paramount role of predictions on sensory processing (Blank et al., 2018; Davis et al., 2011; Davis & Johnsrude, 2007; de Lange et al., 2018; Sohoglu et al., 2012), atypical predictive coding in the subcortical sensory pathway could have profound repercussion at the cognitive level (Diaz et al., 2012; McFadyen et al., 2020; Tabas & von Kriegstein, 2021a). For instance, developmental dyslexia has been attributed to altered adaption dynamics to stimulus regularities (Ahissar et al., 2006; Chandrasekaran et al., 2009; Perrachione et al., 2016), altered responses in the left MGB (Chandrasekaran et al., 2009; Diaz et al., 2012), and atypical left hemispheric cortico-thalamic pathways (Müller-Axt et al., 2017; Tschentscher et al., 2019). Understanding the mechanisms underlying the predictive processing of dynamic acoustic features in subcortical sensory pathways is an essential prerequisite to understand dysfunction.

DATA AND CODE AVAILABILITY

Derivatives (beta maps and log-likelihood maps, computed with SPM) and all code used for data processing and analysis are publicly available in https://osf.io/f5tsy/.

AUTHOR CONTRIBUTIONS

A.T.: Conceptualization, methodology, investigation, and writing—original draft. S.K.: Conceptualization, writing—review & editing. M.M.: Methodology. K.v.K.: Conceptualization, writing—review and editing, supervision, and funding acquisition.

DECLARATION OF COMPETING INTEREST

Authors declare no competing interests.

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: https://doi.org/10.1162/imag_a_00292

REFERENCES

Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, *9*(12), 1558–1564. https://doi.org /10.1038/nn1800

Anderson, L. A., Christianson, G. B., & Linden, J. F. (2009). Stimulus-specific adaptation occurs in the auditory thalamus. *Journal of Neuroscience*, *29*(22), 7359–7363. https://doi.org/10.1523/jneurosci.0793-09.2009

Anderson, L. A., & Malmierca, M. S. (2013). The effect of auditory cortex deactivation on stimulus-specific adaptation in the inferior colliculus of the rat. *European Journal of Neuroscience*, *37*(1), 52–62. https://doi.org/10 .1111/ejn.12018

Antunes, F. M., & Malmierca, M. S. (2011). Effect of auditory cortex deactivation on stimulus-specific adaptation in the medial geniculate body. *Journal of Neuroscience*, *31*(47), 17306–17316. https://doi.org/10.1523/jneurosci .1915-11.2011

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

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044. https://doi.org/10.1016/j .neuroimage.2010.09.025

Ayala, Y. A., & Malmierca, M. S. (2018). The effect of inhibition on stimulus-specific adaptation in the inferior colliculus. *Brain Structure and Function*, *223*, 1391–1407. https://doi.org/10.1007/s00429-017-1546-4

Ayala, Y. A., Pérez-gonzález, D., Duque, D., Nelken, I., & Malmierca, M. S. (2013). Frequency discrimination and stimulus deviance in the inferior colliculus and cochlear nucleus. *Frontiers in Neural Circuits*, *6*, 119. https://doi .org/10.3389/fncir.2012.00119

Ayala, Y. A., Udeh, A., Dutta, K., Bishop, D., Malmierca, M. S., & Oliver, D. L. (2015). Differences in the strength of cortical and brainstem inputs to SSA and non-SSA neurons in the inferior colliculus. *Scientific Reports*, *5*, 10383. https://doi.org/10.1038/srep10383

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251. https://doi.org/10.1111/1469-7610.00715

Bauerle, P., von der Behrens, W., Kossl, M., & Gaese, B. H. (2011). 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. *Journal of Neuroscience*, *31*(26), 9708–9722. https://doi.org/10.1523/jneurosci.5814-10.2011

Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS Biology*, *14*(11), e1002577. https://doi.org/10.1371/journal.pbio.1002577

Blank, H., Spangenberg, M., & Davis, M. H. (2018). Neural prediction errors distinguish perception and misperception of speech. *The Journal of Neuroscience*, *38*(27), 6076–6089. https://doi.org/10.1523/jneurosci .3258-17.2018

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163 /156856897x00357

Brant-Zawadzki, M., Gillan, G. D., & Nitz, W. R. (1992). MP RAGE: A three-dimensional, T1-weighted, gradient-echo sequence—Initial experience in the brain. *Radiology*, *182*(3), 769–775. https://doi.org/10.1148/radiology.182.3 .1535892

Cacciaglia, R., Escera, C., Slabu, L., Grimm, S., Sanjuán, A., Ventura-Campos, N., & Ávila, C. (2015). Involvement of the human midbrain and thalamus in auditory deviance detection. *Neuropsychologia*, *68*, 51–58. https://doi.org/10.1016/j.neuropsychologia.2015.01.001

Carbajal, G. V., & Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: From the subcortical roots to cortical deviance detection. *Trends in Hearing*, *22*, 1–33. https://doi.org/10.1177 /2331216518784822

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron*, *64*(3), 311–319. https://doi.org/10.1016/j.neuron.2009 .10.006

Chandrasekaran, B., Kraus, N., & Wong, P. C. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology*, *107*(5), 1325–1336. https://doi.org/10 .1152/jn.00923.2011

Coffey, E. B., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R. J., & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*(1), 5036. https://doi.org/10.1038 /s41467-019-13003-w

Cornella, M., Bendixen, A., Grimm, S., Leung, S., Schröger, E., & Escera, C. (2015). Spatial auditory regularity encoding and prediction: Human middle-latency and long-latency auditory evoked potentials. *Brain Research*, *1626*, 21–30. https://doi.org/10.1016/j .brainres.2015.04.018

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top-down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*(12), 3914–3932. https://doi.org/10 .1162/jocn_a_00084

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1–2), 132–147. https://doi.org/10.1016/j.heares.2007 .01.014

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics .2018.06.002

Denckla, M. B., & Rudel, R. (1974). Rapid "automatized" naming of pictured objects, colors, letters and numbers by normal children. *Cortex*, *10*(2), 186–202. https://doi .org/10.1016/s0010-9452(74)80009-2

Devore, J. L. (2008). *Probability and statistics for engineering and the sciences*. Spinger.

Diaz, B., Hintz, F., Kiebel, S. J., & von Kriegstein, K. (2012). Dysfunction of the auditory thalamus in developmental dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(34), 13841–13846. https://doi.org/10.1073/pnas.1119828109

Duque, D., & Malmierca, M. S. (2015). Stimulus-specific adaptation in the inferior colliculus of the mouse: Anesthesia and spontaneous activity effects. *Brain Structure and Function*, *220*, 3385–3398. https://doi.org /10.1007/s00429-014-0862-1

Duque, D., Malmierca, M. S., & Caspary, D. M. (2014). Modulation of stimulus-specific adaptation by GABA(A)

receptor activation or blockade in the medial geniculate body of the anaesthetized rat. *The Journal of Physiology*, *592*(Pt. 4), 729–743. https://doi.org/10.1113/jphysiol.2013.261941

Duque, D., Wang, X., Nieto-Diego, J., Krumbholz, K., & Malmierca, M. S. (2016). Neurons in the inferior colliculus of the rat show stimulus-specific adaptation for frequency, but not for intensity. *Scientific Reports*, *6*, 1–15. https://doi.org/10.1038/srep24114

Elliott, L. L. (1971). Backward and forward masking. *Audiology*, *10*(2), 65–76. https://doi.org/10.3109/00206097109072544

Escera, C., & Malmierca, M. S. (2014). The auditory novelty system: An attempt to integrate human and animal research. *Psychophysiology*, *51*(2), 111–123. https://doi.org/10.1111/psyp.12156

Eytan, D., Brenner, N., & Marom, S. (2003). Selective adaptation in networks of cortical neurons. *The Journal of Neuroscience*, *23*(28), 9349–9356. https://doi.org/10.1523/jneurosci.23-28-09349.2003

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–355. https://doi.org/10.1016/s0896-6273(02)00569-x

Friston, K. (2003). Learning and inference in the brain. *Neural Networks: The Official Journal of the International Neural Network Society*, *16*(9), 1325–1352. https://doi.org/10.1016/j.neunet.2003.06.005

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K., Zarahn, E., Josephs, O., Henson, R., & Dale, A. (1999). Stochastic designs in event-related fMRI. *NeuroImage*, *10*(5), 607–619. https://doi.org/10.1006/nimg.1999.0498

Friston, K. J., Sajid, N., Quiroga-Martinez, D. R., Parr, T., Price, C. J., & Holmes, E. (2021). Active listening. *Hearing Research*, *399*, 107998. https://doi.org/10.1016/j.heares.2020.107998

Gao, P. P., Zhang, J. W., Cheng, J. S., Zhou, I. Y., & Wu, E. X. (2014). The inferior colliculus is involved in deviant sound detection as revealed by BOLD fMRI. *NeuroImage*, *91*, 220–227. https://doi.org/10.1016/j.neuroimage.2014.01.043

Geis, H.-R. A. P., & Borst, J. G. G. (2013). Intracellular responses to frequency modulated tones in the dorsal cortex of the mouse inferior colliculus. *Frontiers in Neural Circuits*, *7*(4), 2002–2016. https://doi.org/10.3389/fncir.2013.00007

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, *84*(3), 1588–1598. https://doi.org/10.1152/jn.2000.84.3.1588

Glover, G. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*, 416–429. https://doi.org/10.1006/nimg.1998.0419

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, *5*, 13. https://doi.org/10.3389/fninf.2011.00013

Grimm, S., Escera, C., Slabu, L., & Costa-Faidella, J. (2011). Electrophysiological evidence for the hierarchical organization of auditory change detection in the human brain. *Psychophysiology*, *48*(3), 377–384. https://doi.org/10.1111/j.1469-8986.2010.01073.x

Gutschmidt, K., Wenninger, S., Montagnese, F., & Schoser, B. (2021). Dyslexia and cognitive impairment in adult patients with myotonic dystrophy type 1: A clinical prospective analysis. *Journal of Neurology*, *268*(2), 484–492. https://doi.org/10.1007/s00415-020-10161-6

Hage, S. R. & Ehret, G. (2003). Mapping responses to frequency sweeps and tones in the inferior colliculus of house mice. *European Journal of Neuroscience*, *18*(8), 2301–2312. https://doi.org/10.1046/j.1460-9568.2003.02945.x

Hovsepyan, S., Olasagasti, I., & Giraud, A. L. (2020). Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-16956-5

Hsieh, I.-H., Fillmore, P., Rong, F., Hickok, G., & Saberi, K. (2012). FM-selective networks in human auditory cortex revealed using fMRI and multivariate pattern classification. *Journal of Cognitive Neuroscience*, *24*(9), 1896–1907. https://doi.org/10.1162/jocn_a_00254

Hu, B. (2003). Functional organization of lemniscal and nonlemniscal auditory thalamus. *Experimental Brain Research*, *153*(4), 543–549. https://doi.org/10.1007/s00221-003-1611-5

Ibrahimović, N., & Bulheller, S. (2013). *Rechtschreibtest RST-ARR: Aktuelle Rechtschreibregelung: Lückendiktate*. Pearson Assessment & Information. https://www.pearsonclinical.de/rst.html

Issa, J. B., Haeffele, B. D., Young, E. D., & Yue, D. T. (2016). Multiscale mapping of frequency sweep rate in mouse auditory cortex. *Hearing Research*, *344*, 207–222. https://doi.org/10.1016/j.heares.2016.11.018

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Joanisse, M. F., & DeSouza, D. D. (2014). Sensitivity of human auditory cortex to rapid frequency modulation revealed by multivariate representational similarity analysis. *Frontiers in Neuroscience*, *8*, 1–10. https://doi.org/10.3389/fnins.2014.00306

Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T. U., Sebold, M., Manjaly, Z. M., Pruessmann, K. P., & Stephan, K. E. (2017). The PhysIO toolbox for modeling physiological noise in fMRI data. *Journal of Neuroscience Methods*, *276*, 56–72. https://doi.org/10.1016/j.jneumeth.2016.10.019

Kuo, R. I., & Wu, G. K. (2012). The generation of direction selectivity in the auditory system. *Neuron*, *73*(5), 1016–1027. https://doi.org/10.1016/j.neuron.2011.11.035

Lee, A. K., Larson, E., Maddox, R. K., & Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hearing Research*, *307*, 111–120. https://doi.org/10.1016/j.heares.2013.06.010

Lee, C. C., & Sherman, S. M. (2011). On the classification of pathways in the auditory midbrain, thalamus, and cortex. *Hearing Research*, *276*(1–2), 79–87. https://doi.org/10.1016/j.heares.2010.12.012

Li, A.-A., Zhang, A.-Y., Chen, Q.-C., & Wu, F.-J. (2010). Effects of modulation range and presentation rate of FM stimulus on auditory response properties of mouse inferior collicular neurons. *Sheng li xue bao: [Acta physiologica Sinica]*, *62*(3), 210–218. https://actaps.sinh.ac.cn/article_en.php?id=8090

Liberman, A. M., Delattre, P. C., Gerstman, L. J., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology*, *52*(2), 127–137. https://doi.org/10.1037/h0041240

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz & H. L. Teuber (Eds.), *Perception* (pp. 143–178). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-46354-9_5

Lui, B., & Mendelson, J. R. (2003). Frequency modulated sweep responses in the medial geniculate nucleus. *Experimental Brain Research*, *153*(4), 550–553. https://doi.org/10.1007/s00221-003-1618-y

Malmierca, M. S., Anderson, L. A., & Antunes, F. M. (2015). The cortical modulation of stimulus-specific adaptation in the auditory midbrain and thalamus: A potential neuronal correlate for predictive coding. *Frontiers in Systems Neuroscience*, *9*, 19. https://doi.org/10.3389/fnsys.2015.00019

Malmierca, M. S., Cristaudo, S., Pérez-González, D., & Covey, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *The Journal of Neuroscience*, *29*(17), 5483–5493. https://doi.org/10.1523/jneurosci.4153-08.2009

McFadyen, J., Dolan, R. J., & Garrido, M. I. (2020). The influence of subcortical shortcuts on disordered sensory and cognitive processing. *Nature Reviews Neuroscience*, *21*(5), 264–276. https://doi.org/10.1038/s41583-020-0287-1

Mihai, G., Moerel, M., De Martino, F., Trampel, R., Kiebel, S., & von Kriegstein, K. (2019). Modulation of tonotopic ventral MGB is behaviorally relevant for speech recognition. *eLife*, *8*, e44837. https://elifesciences.org/articles/44837

Mill, R., Coath, M., Wennekers, T., & Denham, S. L. (2011). A neurocomputational model of stimulus-specific adaptation to oddball and Markov sequences. *PLoS Computational Biology*, *7*(8), e1002117. https://doi.org/10.1371/journal.pcbi.1002117

Mill, R., Coath, M., Wennekers, T., & Denham, S. L. (2012). Characterising stimulus-specific adaptation using a multi-layer field model. *Brain Research*, *1434*, 178–188. https://doi.org/10.1016/j.brainres.2011.08.063

Moerel, M., De Martino, F., Uğurbil, K., Yacoub, E., & Formisano, E. (2015). Processing of frequency and location in human subcortical auditory structures. *Scientific Reports*, *5*, 17048. https://doi.org/10.1038/srep17048

Moll, K., & Landerl, K. (2014). Lese-und rechtschreibtest (SLRT-II). Weiterentwicklung des salzburger lese-und rechtschreibtests (SLRT), 2., korrigierte auflage mit erweiterten normen. https://www.hogrefe.com/de/shop/lese-und-rechtschreibtest.html

Müller-Axt, C., Anwander, A., & von Kriegstein, K. (2017). Altered structural connectivity of the left visual thalamus in developmental dyslexia. *Current Biology*, *27*, 3692–3698. https://doi.org/10.1016/j.cub.2017.10.034

Mumford, J. A., Poline, J.-B., & Poldrack, R. A. (2015). Orthogonalization of regressors in fMRI models. *PLoS One*, *10*(4), e0126255. https://doi.org/10.1371/journal.pone.0126255

Nabelek, I., Nabelek, A., & Hirsh, I. J. (1970). Pitch of short tone bursts of changing frequency. *The Journal of the Acoustical Society of America*, *45*(1), 293–293. https://doi.org/10.1121/1.1970857

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, *1104*, 35–53. https://doi.org/10.1196/annals.1390.022

Okamoto, H., & Kakigi, R. (2015). Encoding of frequency-modulation (FM) rates in human auditory cortex. *Scientific Reports*, *5*, 1–9. https://doi.org/10.1038/srep18143

Osman, A. F., Lee, C. M., Escabí, M. A., & Read, H. L. (2018). A hierarchy of time scales for discriminating and classifying the temporal shape of sound in three auditory cortical fields. *Journal of Neuroscience*, *38*(31), 6967–6982. https://doi.org/10.1523/jneurosci.2871-17.2018

Paltoglou, A. E., Sumner, C. J., & Hall, D. A. (2011). Mapping feature-sensitivity and attentional modulation in human auditory cortex with functional magnetic resonance imaging. *European Journal of Neuroscience*, *33*(9), 1733–1741. https://doi.org/10.1111/j.1460-9568.2011.07656.x

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., & Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications*, *8*(1), 2148. https://doi.org/10.1038/s41467-017-02038-6

Pérez-González, D., Hernández, O., Covey, E., & Malmierca, M. S. (2012). GABA A-mediated inhibition modulates stimulus-specific adaptation in the inferior colliculus. *PLoS One*, *7*(3), e34297. https://doi.org/10.1371/journal.pone.0034297

Perrachione, T. K., Del Tufo, S. N., Winter, R., Murtagh, J., Cyr, A., Chang, P., Halverson, K., Ghosh, S. S., Christodoulou, J. A., & Gabrieli, J. D. (2016). Dysfunction of rapid neural adaptation in dyslexia. *Neuron*, *92*(6), 1383–1397. https://doi.org/10.1016/j.neuron.2016.11.020

Pressnitzer, D., Patterson, R. D., & Krumbholz, K. (2001). The lower limit of melodic pitch. *The Journal of the Acoustical Society of America*, *109*(5 Pt. 1), 2074–2084. https://doi.org/10.1121/1.1359797

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Riecke, L., Peters, J. C., Valente, G., Poser, B. A., Kemper, V. G., Formisano, E., & Sorger, B. (2018). Frequency-specific attentional modulation in human primary auditory cortex and midbrain. *NeuroImage*, *174*, 274–287. https://doi.org/10.1016/j.neuroimage.2018.03.038

Rinne, T., Balk, M. H., Koistinen, S., Autti, T., Alho, K., & Sams, M. (2008). Auditory selective attention modulates activation of human inferior colliculus. *Journal of Neurophysiology*, *100*(6), 3323–3327. https://doi.org/10.1152/jn.90607.2008

Rinne, T., Christopher Stecker, G., Kang, X., William Yund, E., Herron, T. J., & Woods, D. L. (2007). Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *NeuroReport*, *18*(13), 1311–1314. https://doi.org/10.1097/wnr.0b013e32826fb3bb

Robinson, B. L., Harper, N. S., & McAlpine, D. (2016). Meta-adaptation in the auditory midbrain under cortical influence. *Nature Communications*, *7*(1), 13442. https://doi.org/10.1038/ncomms13442

Rosa, M., Bestmann, S., Harrison, L., & Penny, W. (2010). Bayesian model selection maps for group studies. *NeuroImage*, *49*(1), 217–224. https://doi.org/10.1016/j.neuroimage.2009.08.051

Rosburg, T., Haueisen, J., & Sauer, H. (2002). Habituation of the auditory evoked field component N100m and its dependence on stimulus duration. *Clinical Neurophysiology*, *113*(3), 421–428. https://doi.org/10.1016/s1388-2457(01)00727-1

Schofield, B. R. (2011). Chapter 9. Central descending auditory pathways. In D. Ryugo & R. Fay (Eds.), *Auditory and vestibular efferents* (pp. 261–290). Springer

Handbook of Auditory Research. https://doi.org/10.1007/978-1-4419-7070-1_9

Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science*, *14*(4), 328–333. https://doi.org/10.1111/1467-9280.14471

Signoret, C., Andersen, L. M., Dahlström, Ö., Blomberg, R., Lundqvist, D., Rudner, M., & Rönnberg, J. (2020). The influence of form- and meaning-based predictions on cortical speech processing under challenging listening conditions: A MEG study. *Frontiers in Neuroscience*, *14*, 1–15. https://doi.org/10.3389/fnins.2020.573254

Sitek, K. R., Gulban, O. F., Calabrese, E., Johnson, G. A., Lage-castellanos, A., Moerel, M., Ghosh, S. S., & Martino, F. D. (2019). Mapping the human subcortical auditory system using histology, postmortem MRI and in vivo MRI at 7T. *eLife*, *8*, e48932. https://doi.org/10.7554/elife.48932

Sohoglu, E., & Davis, M. H. (2020). Rapid computations of spectrotemporal prediction error support perception of degraded speech. *eLife*, *9*, 1–25. https://doi.org/10.7554/elife.58077

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443–8453. https://doi.org/10.1523/jneurosci.5069-11.2012

Steadman, M. A., & Sumner, C. J. (2018). Changes in neuronal representations of consonants in the ascending auditory system and their role in speech recognition. *Frontiers in Neuroscience*, *12*, 1–16. https://doi.org/10.3389/fnins.2018.00671

Stein, J., von Kriegstein, K., & Tabas, A. (2022). Predictive encoding of pure tones and FM-sweeps in the human auditory cortex. *Cerebral Cortex Communications*, *3*(4), tgac047. https://doi.org/10.1093/texcom/tgac047

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Tabas, A., Mihai, G., Kiebel, S., Trampel, R., & von Kriegstein, K. (2020). Abstract rules drive adaptation in the subcortical sensory pathway. *eLife*, *9*, 1–19. https://doi.org/10.7554/elife.64501

Tabas, A., & von Kriegstein, K. (2021a). Adjudicating between local and global architectures of predictive processing in the subcortical auditory pathway. *Frontiers in Neural Circuits*, *15*, 1–14. https://doi.org/10.3389/fncir.2021.644743

Tabas, A., & von Kriegstein, K. (2021b). Neural modelling of the encoding of fast frequency modulation. *PLoS Computational Biology*, *17*, e1008787. https://doi.org/10.1371/journal.pcbi.1008787

Thomas, J. M., Morse, C., Kishline, L., O'Brien-Lambert, A., Simonton, A., Miller, K. E., & Covey, E. (2012). Stimulus-specific adaptation in specialized neurons in the inferior colliculus of the big brown bat, *Eptesicus fuscus*. *Hearing Research*, *291*(1–2), 34–40. https://doi.org/10.1016/j.heares.2012.06.004

Trujillo, M., Carrasco, M. M., & Razak, K. (2013). Response properties underlying selectivity for the rate of frequency modulated sweeps in the auditory cortex of the mouse.

Hearing Research, *298*, 80–92. https://doi.org/10.1016/j.heares.2012.12.013

Tschentscher, N., Ruisinger, A., Blank, H., Diaz, B., & von Kriegstein, K. (2019). Thalamus and the motion-sensitive planum temporale in developmental dyslexia. *The Journal of Neuroscience*, *39*(9), 1720–1732. https://doi.org/10.1523/jneurosci.1435-18.2018

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*(4), 391–398. https://doi.org/10.1038/nn1032

Varghese, L., Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Research*, *1626*, 146–164. https://doi.org/10.1016/j.brainres.2015.06.038

Vidal, Y., Brusini, P., Bonfieni, M., Mehler, J., & Bekinschtein, T. A. (2019). Neural signal to violations of abstract rules using speech-like stimuli. *eNeuro*, *6*(5), 1–14. https://doi.org/10.1523/eneuro.0128-19.2019

von Kriegstein, K., Patterson, R. D., & Griffiths, T. D. (2008). Task-dependent modulation of medial geniculate body is behaviorally relevant for speech recognition. *Current Biology*, *18*(23), 1855–1859. https://doi.org/10.1016/j.cub.2008.10.052

Winer, J. A. (1984). The human medial geniculate body. *Hearing Research*, *15*(3), 225–247. https://doi.org/10.1016/0378-5955(84)90031-5

Winer, J. A. (2005a). Decoding the auditory corticofugal systems. *Hearing Research*, *207*(1–2), 1–9. https://doi.org/10.1016/j.heares.2005.06.007

Winer, J. A. (2005b). Three systems of descending projections to the inferior colliculus. In J. A. Winer & C. E. Schreiner (Eds.), *The inferior colliculus* (pp. 231–247). Springer-Verlag. https://doi.org/10.1007/0-387-27083-3_8

Ye, C.-q., Poo, M.-m., Dan, Y., & Zhang, X.-h. (2010). Synaptic mechanisms of direction selectivity in primary auditory cortex. *Journal of Neuroscience*, *30*(5), 1861–1868. https://doi.org/10.1523/jneurosci.3088-09.2010

Yildiz, I. B., von Kriegstein, K., & Kiebel, S. J. (2013). From birdsong to human speech recognition: Bayesian inference on a hierarchy of nonlinear dynamical systems. *PLoS Computational Biology*, *9*(9), e1003219. https://doi.org/10.1371/journal.pcbi.1003219

Ylinen, S., Huuskonen, M., Mikkola, K., Saure, E., Sinkkonen, T., & Paavilainen, P. (2016). Predictive coding of phonological rules in auditory cortex: A mismatch negativity study. *Brain and Language*, *162*, 72–80. https://doi.org/10.1016/j.bandl.2016.08.007

Yu, X.-J., Xu, X.-X., He, S., & He, J. (2009). Change detection by thalamic reticular neurons. *Nature Neuroscience*, *12*(9), 1165–1170. https://doi.org/10.1038/nn.2373

Zhang, L. I., Tan, A. Y. Y., Schreiner, C. E., & Merzenich, M. M. (2003). Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature*, *424*(6945), 201–205. https://doi.org/10.1038/nature01796

Zhao, L., Liu, Y., Shen, L., Feng, L., & Hong, B.... (2011). Stimulus-specific adaptation and its dynamics in the inferior colliculus of rat. *Neuroscience*, *181*, 163–174. https://doi.org/10.1016/j.neuroscience.2011.01.060